University of Bremen

Department of Mathematics and Computer Science
University of Bremen, Germany

Multisensory Guidance Under Sensory Constraints in Augmented Reality

Alexander Marquardt

This dissertation is submitted for the degree of
Doctor of Philosophy

Date of Colloquium: March 24, 2023

1st Supervisor: Prof. Dr. Johannes Schöning
2nd Supervisor: Prof. Dr. Ernst Kruijff

Abstract

Visual search is a goal-oriented task that involves actively scanning the visual environment for a specific target. Augmented Reality technologies can be used to facilitate visual search by superimposing virtual cues into the user's field of view. While researchers have been primarily concerned with developing visual guidance techniques in Augmented Reality, we have focused on the provision of multisensory guidance cues to support the search task. This approach appears promising, as spatially-informative multisensory cues have been shown to be useful in directing user attention and improving target search performance.

This thesis introduces novel multisensory guidance cues for head-mounted Augmented Reality displays. We have investigated different sensory cues to draw the user's attention to spatially distributed object locations in the environment. Our approaches address typical sensory constraints associated with the use of Augmented Reality systems. In this context, the use of visual guidance methods is limited because the perception of augmentations can be severely affected by internal or external influencing factors. Our method uses a novel type of audio-tactile feedback to provide spatial information to support search even when perception is affected by sensory constraints. Through this thesis, we clarify to what extent search under sensory constraints in Augmented Reality can benefit from multisensory guidance methods. For this purpose, we highlight the relevance of multisensory guidance under sensory constraints, describe the contributions of five papers, and discuss implications regarding search guidance in Augmented Reality.

We first describe how non-visual guidance cues can contribute to search performance in scenes with high information density. In such scenarios, users may not be able to localize target locations because they are visually occluded by other objects. Our results show how audio-tactile proximity feedback and the presentation of vibrotactile cue patterns can enhance search performance for 3D interaction tasks in dense information scenes. Subsequently, we examine the effectiveness of multisensory guidance in supporting search under sensory constraints. We have found that users can locate spatially distributed objects effectively using head-based audio-tactile directional cues. This approach is particularly useful for head-mounted Augmented Reality systems with a limited field of view, in which information is often cluttered or located out-of-view. Our audio-tactile guidance approach, albeit generally slower than current visual guidance techniques, can achieve comparable results in terms of hit-rate and accuracy when searching for information. Lastly, we have investigated how to improve situation awareness during search under sensory constraints using multisensory guidance. Our results show that the provision of multisensory proximity and transition cues can improve the perception of moving out-of-view objects. Thus, we can effectively increase situation awareness by actively directing the user's attention to

previously unrecognized information. In this context, multisensory cues involving tactile stimuli were found to be particularly salient in the presence of external noise.

In summary, this thesis provides important insights into the key values of multisensory guidance under sensory constraints. We have demonstrated that our multisensory guidance approach has the potential to mitigate the effects caused by sensory constraints to effectively support guided search in augmented reality. We have also highlighted limitations of our methods and have discussed how these can be addressed in future work. The main findings of this work will remain relevant even as visual guidance methods evolve and display technologies continue to improve in the future. The results indicate the general applicability of our methods for various search-related tasks in augmented environments, making them potentially useful in other domains as well.

Zusammenfassung

Die visuelle Suche ist eine zielgerichtete Aufgabe, bei der die visuelle Umgebung aktiv nach einem bestimmten Ziel abgesucht wird. Augmented Reality-Technologien können die visuelle Suche erleichtern, indem sie virtuelle Hinweise in das Sichtfeld des Nutzers einblenden. Während sich Forscher in erster Linie mit der Entwicklung visueller Führungstechniken in Augmented Reality befasst haben, haben wir uns auf die Bereitstellung multisensorischer Führungshinweise zur Unterstützung der Suchaufgabe konzentriert. Dieser Ansatz erscheint vielversprechend, da sich räumlich-informative multisensorische Hinweise als nützlich erwiesen haben, um die Aufmerksamkeit des Nutzers zu lenken und die Leistung bei der Zielsuche zu verbessern.

In dieser Arbeit werden neuartige multisensorische Führungshinweise für kopfgetragene Augmented Reality-Displays vorgestellt. Wir haben verschiedene sensorische Hinweise untersucht, um die Aufmerksamkeit des Benutzers auf räumlich verteilte Objektpositionen in der Umgebung zu lenken. Unsere Ansätze gehen auf typische sensorischen Einschränkungen bei der Nutzung von Augmented Reality-Systemen ein. In diesem Zusammenhang ist der Einsatz von visuellen Führungsmethoden begrenzt, da die Wahrnehmung von Augmentierungen durch interne oder externe Einflussfaktoren stark beeinträchtigt sein kann. Unsere Methode verwendet eine neue Art von audio-taktilem Feedback, um räumliche Informationen zur Unterstützung der Suche bereitzustellen, selbst wenn die Wahrnehmung durch sensorische Einschränkungen beeinträchtigt wird. In dieser Arbeit wird geklärt, inwieweit die Suche unter sensorischen Einschränkungen in Augmented Reality von multisensorischen Führungsmethoden profitieren kann. Zu diesem Zweck heben wir die Relevanz der multisensorischen Führung unter sensorischen Einschränkungen hervor, beschreiben die Beiträge von fünf Publikationen und diskutieren die Auswirkungen auf die Suchführung in Augmented Reality.

Wir beschreiben zunächst, wie nicht-visuelle Führungshinweise zur Suchleistung in Szenen mit hoher Informationsdichte beitragen können. In solchen Szenarien kann es vorkommen, dass Benutzer die Zielorte nicht lokalisieren können, weil sie visuell von anderen Objekten verdeckt werden. Unsere Ergebnisse zeigen, wie audio-taktiles Proximity-Feedback und die Präsentation von vibrotaktilen Hinweismustern die Suchleistung bei 3D-Interaktionsaufgaben in Szenen mit hoher Informationsdichte verbessern können. Anschließend untersuchen wir die Effektivität der multisensorischen Führung bei der Unterstützung der Suche unter sensorischen Einschränkungen. Wir haben herausgefunden, dass Benutzer räumlich verteilte Objekte mit Hilfe von kopfbasierten audio-taktilen Richtungshinweisen effektiv lokalisieren können. Dieser Ansatz ist besonders nützlich für kopfgetragene Augmented Reality-Systeme mit einem eingeschränktem Sichtfeld, bei denen Informationen oft unübersichtlich dargestellt sind oder sich außerhalb des Sichtfeldes befinden. Unser audio-taktiler Führungsansatz ist zwar im Allgemeinen langsamer

als aktuelle visuelle Führungstechniken, kann aber vergleichbare Ergebnisse in Bezug auf die Trefferquote und Genauigkeit bei der Suche nach Informationen erzielen. Abschließend haben wir untersucht, wie das Situationsbewusstsein während der Suche unter sensorischen Einschränkungen durch multisensorische Führung verbessert werden kann. Unsere Ergebnisse zeigen, dass die Bereitstellung von multisensorischen Proximity- und Transition-Hinweisen die Wahrnehmung von sich bewegenden Objekten außerhalb des Sichtfeldes verbessern kann. Auf diese Weise können wir das Situationsbewusstsein effektiv erhöhen, indem wir die Aufmerksamkeit des Benutzers aktiv auf zuvor unerkannte Informationen lenken. In diesem Zusammenhang wurde festgestellt, dass multisensorische Hinweise, die taktile Reize beinhalten, in Gegenwart von externen Störfaktoren besonders ausgeprägt sind.

Zusammenfassend lässt sich sagen, dass diese Arbeit wichtige Einblicke in die wesentlichen Eigenschaften der multisensorischen Führung liefert. Wir haben gezeigt, dass unser multisensorischer Führungsansatz das Potenzial hat, die durch sensorische Einschränkungen verursachten Effekte zu mildern und die geführte Suche in Augmented Reality effektiv zu unterstützen. Wir haben auch die Grenzen unserer Methoden aufgezeigt und erörtert, wie diese in zukünftigen Arbeiten berücksichtigt werden können. Die wichtigsten Ergebnisse dieser Arbeit werden auch dann relevant bleiben, wenn sich die Methoden der visuellen Führung weiterentwickeln und die Displaytechnologien in Zukunft weiter verbessert werden. Die Ergebnisse weisen auf die allgemeine Anwendbarkeit unserer Methoden für verschiedene suchbezogene Aufgaben in Augmented Reality-Umgebungen hin, was sie auch in anderen Bereichen potenziell nützlich macht.

So, as much as you can . . . choose whatever you'll regret the least.

- Levi Ackermann in Hajime Isayama's *Attack on Titan*

Acknowledgements

This thesis would not have been possible without many people who supported me during my time as a doctoral student. First and foremost, I would like to thank Ernst Kruijff for the great support, useful discussions, and comments and for giving me the opportunity to explore an interesting field of research. I would also like to thank Johannes Schöning for his helpful advice and patient guidance throughout my entire research journey.

In addition, I would especially like to thank Jens Maiero for his constant support. He often helped us to focus on the essentials when research ideas bubbled over again in our countless brainstorming sessions. Many thanks also go to my long-time office colleague David Eibich, who has supported me with the technical implementation for so many years. Special thanks also go to Christina Trepkowski, with whom I have been doing research together practically since my time as an undergraduate student. It has always been a great pleasure to gather and discuss new research ideas together. I am also very grateful for her valuable advice and support in the statistical analysis and processing of the data. Special thanks also to my dear colleagues and collaborators André Hinkenjann, Kiyoshi Kiyokawa, Martin Weier, Saugata Biswas, Andrea Schwandt, Wolfgang Stürzlinger, Bernhard Rieke, and everyone who has supported me in any way. I would also like to thank the Graduate Institute of the Bonn-Rhein-Sieg University of Applied Sciences, in cooperation with the Institute for Visual Computing, which supported me with a scholarship. This helped me to fully concentrate on my research over a long period of time, which had a positive impact on the progress and quality of the research presented in this PhD thesis.

Last but not least, I would like to thank Ikuyo Shishido for her unconditional support, for always listening to me and encouraging me.

1000 Dank.

Publications

The thesis builds upon five publications that I (co-)authored during my PhD studies. In the following, I describe my contributions to these publications. Publication I, II and III were published at peer-reviewed conferences, publication IV and V were published in peer-reviewed journals. Publications I-V are listed in the order they appear in the contributions section (see Section 1.3). Lastly, I list other publications that I (co-)authored during my PhD studies, but which are not included in the thesis.

Included in Thesis

I. Marquardt, A., Kruijff, E., Trepkowski, C., Maiero, J., Schwandt, A., Hinkenjann, A., Stürzlinger, W., & Schöning, J. (2018). Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation. *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 1–10. DOI: 10.1145/3281505.3281525
Contribution: I was responsible for designing the theoretical framework with Kruijff and Stürzlinger, assembled the hardware supported by Maiero and Schwandt, implemented the software, conducted the user study, analyzed the data with Trepkowski and contributed to all parts of the manuscript.

II. Marquardt, A., Maiero, J., Kruijff, E., Trepkowski, C., Schwandt, A., Hinkenjann, A., Schöning, J., & Stürzlinger, W. (2018). Tactile Hand Motion and Pose Guidance for 3D Interaction. *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 1–10. DOI: 10.1145/3281505.3281526
Contribution: I was responsible for designing the theoretical framework with Kruijff and Maiero, assembled the hardware setup, implemented the software, conducted the user study, analyzed the data with Trepkowski and contributed to all parts of the manuscript.

III. Marquardt, A., Trepkowski, C., Eibich, T. D., Maiero, J., & Kruijff, E. (2019). Non-Visual Cues for View Management in Narrow Field of View Augmented Reality Displays. *2019 IEEE International Symposium on Mixed and Augmented Reality*, 190-201. DOI: 10.1109/ISMAR.2019.000-3
Contribution: I was responsible for designing the theoretical framework with Kruijff and Maiero, assembled the hardware setup, implemented the software with Eibich, conducted the user study, analyzed the data with Trepkowski and contributed to all parts of the manuscript.

IV. Marquardt, A.*, Trepkowski, C.*, Eibich, T. D., Maiero, J., Kruijff, E., & Schöning, J. (2020). Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics*, 26(12), 3389-3401. DOI: 10.1109/TVCG.2020.3023605. *Equal contribution.

Contribution: I was responsible for designing the theoretical framework, assembled the hardware setup, implemented the software with Eibich, conducted the user study and analyzed the data with Trepkowski. I contributed to all parts of the manuscript with Trepkowski in equal parts.

V. Trepkowski, C.*, Marquardt, A.*, Eibich, T. D., Shikanai, Y., Maiero, J., Kiyokawa, K., Kruijff, E., Schöning, J., & König, P. (2021). Multisensory Proximity and Transition Cues for Improving Target Awareness in Narrow Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics*, 28(2), 1342-1362. DOI: 10.1109/TVCG.2021.3116673. *Equal contribution.
Contribution: I was responsible for designing the theoretical framework with Trepkowski, assembled the hardware setup and implemented the software with Eibich. Together with Trepkowski, I conducted the user study, analyzed the data, and contributed to all parts of the manuscript in equal parts.

Not included in Thesis

VI. Kruijff, E., Marquardt, A., Trepkowski, C., Schild, J., & Hinkenjann, A. (2015). Enhancing User Engagement in Immersive Games through Multisensory Cues. *Proceedings of the 7th International Conference on Games and Virtual Worlds for Serious Applications*, pp. 1-8. DOI: 10.1109/VS-GAMES.2015.7295773

VII. Kruijff, E., Marquardt, A., Trepkowski, C., Lindeman, R. W., Hinkenjann, A., Maiero, J., & Riecke, B. E. (2016). On your feet! Enhancing Vection in Leaning-Based Interfaces through Multisensory Stimuli. *Proceedings of the 2016 Symposium on Spatial User Interaction*, pp. 149-158. DOI: 10.1145/2983310.2985759

VIII. Kruijff, E., Marquardt, A., Trepkowski, C., Schild, J., & Hinkenjann, A. (2017). Designed Emotions: Challenges and Potential Methodologies for Improving Multisensory Cues to Enhance User Engagement in Immersive Systems. *The Visual Computer*, 33(4), pp. 471-488. DOI: 10.1007/s00371-016-1294-0

IX. Marquardt, A., Trepkowski, C., Maiero, J., Kruijff, E., Hinkenjann, A. (2018). Multisensory Virtual Reality Exposure Therapy. *Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 769-770. DOI: 10.1109/VR.2018.8446553

X. Trepkowski, C., Eibich, D., Maiero, J., Marquardt, A., Kruijff, E., & Feiner, S. (2019). The Effect of Narrow Field of View and Information Density on Visual Search Performance in Augmented Reality. *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 575-584. DOI: 10.1109/VR.2019.8798312

Table of Contents

1 Introduction

The visual search for information is a vital task for humans and is used for many activities in daily life. Search requires attention by actively scanning the environment to find a particular object among other, distracting items [405, 449]. With recent advantages in Augmented Reality (AR) technologies, search can be facilitated through visual cueing methods [256]. This is achieved by projecting contextual visual information such as graphical overlays and text labels into the user's field of view (FOV) [17].

Visual cueing for guidance, hereafter referred to as visual guidance, is a well researched domain [449]. Guidance methods in AR visualize the location or direction of the target object in the environment of the user. Visual guidance typically involves overlaying abstract indicators such as arrows to provide information about the object to find [44]. Thus, visual guidance can direct the user's visual focus and facilitate rapid visual search [257]. Head-worn devices such as optical see-through (OST) head-mounted displays (HMD) are commonly used for experiencing AR [17]. However, using AR technology leads to certain issues related to perception [221] and cognition [25], that affect search in AR [44, 148]. These problems are referred to as sensory constraints and concern issues related to the user, the environment, and the device. For example, a narrow FOV is a sensory constraint that is typically associated with OST HMDs [221]. When using visual methods in a narrow FOV, guidance information can occupy a large portion of the available screen space. This results in occlusion issues and a cluttered view, which in turn affects performance in goal-oriented search tasks in AR [44, 148]. Sensory constraints are described in detail in Section 1.2 "Sensory Constraints".

In this thesis, we investigate the use of multisensory guidance to mitigate the effects of sensory constraints in AR. Guidance in AR is typically provided by visual cues [256]. However, human perception is highly multisensory, allowing the construction of a coherent picture of the external world from information provided by different sensory systems [391]. Therefore, we believe that multisensory cues that incorporate auditory and tactile stimuli are a beneficial approach to direct attention and support guidance in AR [90, 395]. Our strategy is based on a head-mounted AR device enhanced by a special multisensory setup. This configuration allows us to provide contextual auditory and tactile cues directed to the user's head in addition to the visual augmentations of the AR display. We use sensory substitution techniques to transpose spatial visual information into auditory and tactile sensations at the users head. For the transposition of sensory information, we

take advantage from the human perceptual sensitivities for the corresponding sensory modalities, namely the auditory frequency range and the tactile sensitivity of the skin [135]. Both sensory channels allow the use of a wide range of sensory stimuli in terms of perceptibility and discriminability [350, 408] which can be used to indicate the position of a target in the environment. In our approach, we exploit the potential of head-based auditory and tactile feedback to empower guided search. In this way, we directly address the issues that arise from sensory constraints in AR target search. In addition, we show how multisensory cues can be used to support attentional guidance to promote situation awareness (SA) when AR technologies are used.

1.1 Background and Motivation

This section presents the theoretical background that motivated the development of multisensory guidance under sensory constraints. First, we describe the technical foundations of AR and discuss typical devices associated with this technology. Second, the principles of real-world search, attention, and visual aids in AR are explained. Third, we explain the fundamentals of multisensory integration, including crossmodal interactions as well as approaches for sensory substitution. Finally, we explain the impact of SA on user performance in this context.

1.1.1 Technical Background

This subsection describes the technical background of the relevant technology used in this work, namely augmented reality using head-mounted display devices.

1.1.1.1 Augmented Reality. AR refers to the combination of real and digital information by superimposing virtual objects on the real world in a semantic context in real time. Virtual objects can be any computer-generated data, such as text, graphics, 3D, animation, audio, and video. Unlike virtual reality (VR) technologies, which completely immerse a user inside a synthetic environment created by computer graphics, AR supplements reality [17]. However, it is generally considered that the relationship between VR and AR is continuous, supplying purely virtual environments and purely real environments at opposite ends of a continuum (see Fig. 1.1). Environments in which real world and virtual world objects are presented together within a single display, that is, anywhere between extremes of this continuum is defined as Mixed Reality (MR) environment [287]. Thus,

MR covers most parts of the continuum except for the endpoints [401]. In the broader context, eXtended reality (XR) is a relatively new umbrella term that encompasses any sort of technology that alters reality by adding digital elements to the physical or real-world environment to any extent. Thus, XR includes AR, MR, VR, and any technology at any point along the virtuality continuum [400].

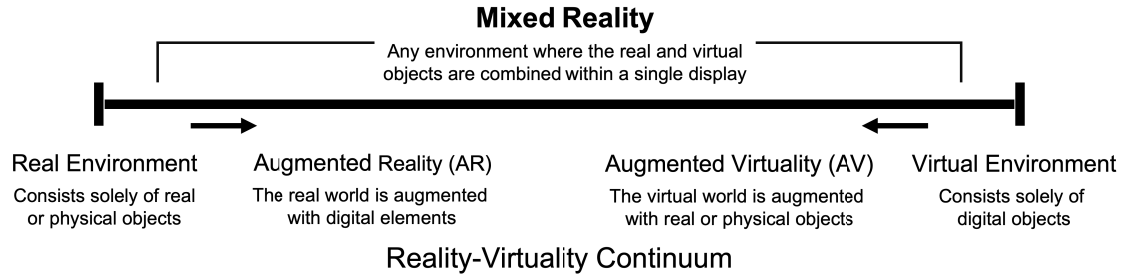


Fig. 1.1: Reality-Virtuality Continuum by Milgram and Kishino, adapted from [287, 401].

Mobile AR is one of the fastest growing research areas related to MR, partially due to the emergence of powerful and ubiquitous platforms for supporting mobile AR [15]. AR systems can be classified in three display categories based on the position between the viewer and the real environment: head-worn devices, hand-held and spatial. Head-worn devices can be subdivided into video and optical see-through head mounted displays, virtual retina displays and head-mounted projective displays. Hand-held AR displays include video or optical see-through displays and hand-held projectors. Spatial devices are placed statically within the environment and include screen-based video see-through displays, spatial optical see-through displays, and projective displays [420].

Although AR systems are not limited to sight and can apply to all senses, nearly all of the interest and development of AR to date has focused on the visual domain, such as virtual graphical objects and textual overlays [17].

1.1.1.2 Head-Mounted Displays. A head-mounted display is a display unit that is mounted on the user's head. An HMD consists of a helmet with small CRTs or liquid-crystal displays placed directly in front of the user's eyes [376]. HMDs can generally be divided into optical see-through (OST) and video see-through (VST) devices [465].

OST displays allow users to see the real world, overlaying graphics onto the user's view by using a holographic optical element. The main advantage of OST displays is that they offer a superior view of the real world by including a natural, instantaneous view of the real scene. VST displays, on the other hand, show a video view of the real world

containing overlaid graphics. The advantage of VST HMDs is the consistency between real and synthetic views. In addition, VST displays can accommodate occlusion issues better than OST displays due to the availability of image processing techniques [465].

Although AR can run on various platforms [17], head-worn devices offer several advantages over other form factors [210]. HMDs include sensors to detect head movement, which are able to adjust the view port of the moving head accordingly. Additionally, they can be equipped with eye-tracking technology to support gaze-based user interaction. Thus, HMDs enable a natural interaction with the environment. Finally, because the devices are head-worn and used near receptors for special senses, it becomes convenient to modulate sensations that a user perceives for augmented content while the hands can remain free for other tasks [210].

1.1.2 Guidance of Visual Search

This subsection examines methods for visually guiding search through attention. We first investigate which stimuli influence visual search in real-world search tasks. We then present selected methods for guided search in AR applications.

1.1.2.1 Visual Search and Attention. Visual search is a perceptual task that requires attention and typically involves actively scanning the visual environment for a specific object (target) among other objects (distractors) [455]. The ability to find one item in a visual world filled with other distracting items is an important routine of visual behavior [449]. It has become apparent that vision plays an important role for search, as the visual sense is considered to be the most dominant sense that humans use to perceive their environment [333]. Approximately 80% of the information extracted from the environment is perceived through the eyes [81]. Therefore, the processes of visual search and object recognition are closely related to human performance [439]. However, due to the limitations of human visual processing, it is impossible to recognize everything in the FOV at once. Therefore, the desired target must be searched for, even when it is located in the current FOV. This is because a large number of visual functions can only be performed in a limited part of the visual field at any given time, resulting in the need to direct attention to objects that could be the potential target [452].

The deployment of attention is guided to the most promising items and locations by one or more sources of information. Such sources are found in pre-attentive attributes such as color, motion, and size, which can be processed in parallel for the entire visual

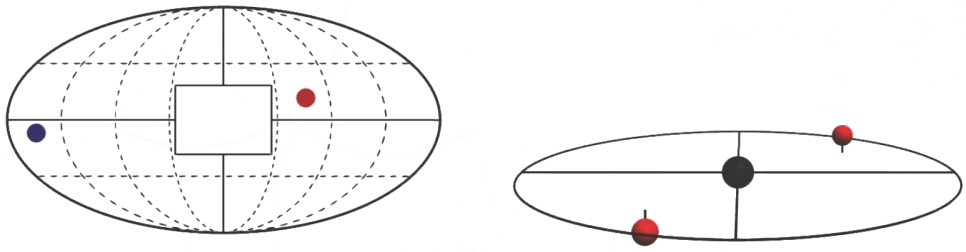
field in a single step [448]. This can be used to determine whether, for example, colors or movements are present at a particular location. Subsequently, attention can be guided in two ways: "bottom-up" stimulus driven and "top-down" user-driven. Top-down and bottom-up guidance can interfere with each other. First, attention is drawn in a bottom-up, stimulus-driven manner to the most salient items in view [450]. Here, attention is attracted to objects that differ from their surroundings when those differences are large enough and occur in a limited number of features that direct attention. This effect is called "pop-out" [405], which is known as an effective way of guidance. There are two underlying rules of bottom-up salience, namely, the salience of a target increases (1) with the difference from the distractors (target-distractor heterogeneity) and (2) with the homogeneity of the distractors along basic feature dimensions (distractor-distractor homogeneity). However, the bottom-up model does not perform well if the observer has a clear top-down goal, since it directs the attention first to the most salient areas in view (e.g., the most colorful, shiny objects) [452]. Second, attention is guided from the top-down in a user-driven manner. In top-down guidance, also known as feature guidance or feature-based attention, attention is directed toward objects with known features of targets. For example, the observer first searches for objects with the desired color and then determines the correct shape. Thus, top-down search benefits from knowledge of attributes such as color, size, and orientation. According to the target-distractor heterogeneity rule, search efficiency depends on the number of features that the target and distractors have in common. Moreover, observers seem to be able to focus their attention on multiple target features simultaneously; however, guidance toward multiple features does not appear to be an adequate account of how humans search for objects in the real world. The structure of the scene provides sources of guidance for real-world search, distinguishing between semantic and syntactic guidance. Syntactic guidance is based on physical constraints while semantic guidance refers to the meaning of the scene. For example, syntactic guidance implies that a person does not necessarily have to be searched for in the sky, because persons usually must be supported against gravity. In a search with semantic guidance, a person is less likely to be searched for on the roof of a building – not because the person could not be found there, but rather because understanding the scene makes it more likely that the person can be found on the ground [452].

In addition, visual search performance (search times) can strongly vary due to inhomogeneous processing across the visual field [103]. In particular, visual clutter caused by excess or disorganized display items (distractors) leads to the degradation of performance

in search tasks [354]. Furthermore, it is possible that the object to be found is currently not within the FOV of the observer [451]. This condition is due to the fact that human vision is limited by a binocular FOV of approximately 210° horizontally and 150° vertically [196], so that only parts of the environment can be perceived, while the rest remains partially unnoticed.

1.1.2.2 Visual Guidance in AR. Similar to guiding attention in real-world search, guidance can be assisted by overlaying digital information on top of the user's view through AR technologies. Visual guidance methods are used to draw user's attention towards augmentations to find real-world locations. By following visual cues for attention guidance, a user can find the shortest path between a starting orientation and a destination orientation in 3D space. However, due to the nature of OST devices, only a small portion of the environment is visible at any given time, resulting in augmentations that are frequently located outside the FOV (see Subsection 1.2.2.4 "Field of View") [44].

Existing visual guidance techniques for AR/VR environments can be classified into the following three categories [44]: Overview & Detail, Focus & Context, and Contextual Views. Overview & Detail describes approaches that provide two windows – one overview and one detail window. Both windows are typically shown on top of each other. The overview window conveys information about the user's environment, for example, in the form of a map. The detail window provides information regarding a local subspace, usually related to what the user is currently viewing. Focus & Context approaches describe methods that provide a distorted view of the user's environment, for example, using a fisheye projection [44]. Finally, contextual views overlay the view with abstract indicators, such as arrows, that provide information about the desired location. Furthermore, visualization techniques can be classified into 2D and 3D techniques based on the dimensionality of their information. However, 2D techniques have proven to be insufficient, especially for tasks such as navigation in an augmented environment. Prominent examples for 3D visualization techniques are 3D Halo projections [143], 3D arrows [366], attention funnels [37], EyeSee360 [144], and 3D Radar [44]. Fig. 1.2 shows the latter visualizations, which represent two of the current state-of-the-art in visual guidance [44]: The overview & detail method 3D radar (left) and the focus & context method EyeSee360 (right).



(a) EyeSee360. (b) 3D Radar.

Fig. 1.2: Two visualization methods used for guidance in AR. The visualizations are superimposed over the visible screen area (FOV), potentially occupying large amounts of the display real estate. *Note: Both methods in this example encode the 3D location of identical out-of-view objects.*

Conclusion for Multisensory Guidance in AR: The properties for visual search and attention are essential for multisensory guidance. However, search in AR always remains associated with visual properties, such as searching for a specific object or location in the environment and/or the associated overlaid augmentation. Search using multisensory guidance cues is usually accompanied by characteristics of visual search. Thus, models of visual search must also be considered in the design and implementation of multisensory guidance. These aspects include bottom-up and top-down guidance and, to some extent, guidance by environmental properties.

1.1.3 Multisensory Integration.

The processing of multisensory information is ubiquitous in daily life [383]. To perceive and understand the environment, humans rely on rich and complex sensory information. This multisensory experience is enabled by the brain's ability to combine signals from different sensory systems [457]. The incoming stream of – often ambiguous – sensory information is processed by the brain to reconstruct the environment into an unambiguous interpretation of the world [113]. This process is called multisensory integration, in which independent but timed signals originating from multiple sensory sources are combined into a coherent representation [270].

The interaction between multiple sensory signals can be described in two ways: (1) by redundant sensory signals and (2) by sensory combination with non-redundant cues. Redundant sensory signals arise within the same coordinate system and relate to the same environmental property. For example, both visual and auditory information can be transformed into craniotopic (head-based reference system) coordinates. Thus, vision

and audition can be used, for example, to receive redundant information about a person´s location [161]. Sensory combination refers to multisensory interactions for sensory signals that are not redundant and may be encoded in different coordinate systems [259].For example, vision and smell provide non-redundant information about a person's identity [161]. The perception of redundant sensory information is beneficial, as it is used to reduce the uncertainty in the estimation of the environmental property. In addition, the combination of complementary information can be beneficial in that it can expand the range and richness of the information available. However, it can also be superfluous and inadequate for the task. For example, olfactory cues may increase the richness of the representation but are not necessarily helpful in localization [161]. Because human information processing capacity is limited [247], interference or bottlenecks in attention could occur at a modality-specific or cross-modality level shared by different sensory modalities [171]. Finally, multisensory cues need to be spatially [389] and temporally [392] aligned. Otherwise, the information may be perceived to be derived from separate sensory sources, causing a depression in the multisensory response and instead leading to the perception of separate, unisensory responses [250].

In the following subsections, we discuss two aspects of multisensory integration that are important for this work, namely crossmodal links in spatial attention and sensory substitution.

1.1.3.1 Crossmodal Links in Spatial Attention.

When searching for something, multiple sensory impressions can be helpful. For example, a person is easier to find in a noisy crowd if that person is waving their arms (visual) and shouting loudly (auditory). In this way, sets of information from different sensory modalities interact with each other and help complete the search task more quickly [398]. Crossmodal links are situations in which the presentation of a stimulus in one sensory modality exerts an influence on the perception of, or ability to respond to, stimuli presented in another sensory modality. Prominent examples for crossmodal effects include the McGurk effect [278] and the ventriloquism effect [35]. The McGurk effect describes the influencing of the perception of an acoustic speech signal by the simultaneous observation of a lip movement or unconscious lip reading. The ventriloquism effect is a spatial interaction between auditory and visual inputs in which the perceived location of an auditory stimulus is attracted towards the location of a visual stimulus.

It has been demonstrated that there are robust crossmodal links between auditory, visual, and tactile sensations in spatial attention (see [96] for an overview). For example, an irrelevant but salient visual, auditory, or tactile event can attract covert spatial attention in the other modalities. When a person expects a target to appear in one modality (e.g., auditory) at a particular location, judgments at that location improve not only for the expected modality but also for other modalities (e.g., vision), even when events in the secondary modality may be more likely to occur elsewhere [94]. The efficiency of human multisensory information processing may be enhanced when relevant information is presented to different senses from approximately the same spatial location [169]. This effect suggests that it may be more difficult to selectively focus on a sensory signal when a concurrent signal from a different sensory modality is presented from approximately the same location. However, this also implies that it is more difficult to ignore a sensory signal when it is presented at the current focus of a person's spatial attention [96, 169]. These connections in spatial attention have been shown to influence both exogenous and endogenous attentional orienting [169]. Exogenous orientation refers to a stimulus-driven bottom-up shift in attention, with external stimulation resulting in a reflexive orientation. Endogenous orientation refers to a voluntary shift of attention that is internally controlled by top-down mechanisms [169, 214, 215, 332].

Research has shown that crossmodal cueing has the potential to facilitate a participant's visual search performance [301]. For example, presenting spatially informative non-visual cues – especially auditory cues colocalized with visual targets – has been shown to reduce visual search latency for peripherally located visual targets (e.g., [320, 321]). Furthermore, spatially uninformative auditory cues can reduce visual search latency for visual targets in the central field [101, 320]. Spatially informative auditory cueing of the target side can lead to improved discrimination of visual targets. This improvement occurs even when the cue side has not predicted the side on which the visual target is likely to occur [301]. Spatially uninformative auditory and vibrotactile cues have been shown to facilitate visual search performance when synchronized temporally with a change in the target stimulus [175, 301]. As bimodal extensions to the visual pop-out effect [405], auditory-visual pip-and-pop [414] and tactile-visual poke-and-pop [415] effects can modulate search by presenting temporally synchronous but also spatially informative cues regarding the likely location of the target [301]. The additional sensory signals boost the saliency of the concurrently presented visual event, resulting in a salient emergent feature that "pops out" from the cluttered visual environment [301]. When input and output are processed by

different sensory modalities, coordination or switching between the different modalities can become costly [171, 381]. This situation may be less efficient than the case in which a joint multisensory attentional resource account is assumed. However, auditory and visual spatial attention, for example, are not assumed to address separate resources but are instead interconnected [96, 169].

When considering methods for measuring attention, it is important to distinguish between overt and covert attention mechanisms. Overt attention involves physically directing the eyes, head, and hands toward a stimulus. Covert attention refers to a mental shift of attention without physical movement [264]. There are several ways in which attention can be measured; both qualitative methods, such as the use of questionnaires, and quantitative methods can be used to examine user feedback regarding a stimulus [264]. Attentive responses can be measured either directly in the brain or indirectly through user behavior. The work described here has focused on indirect, quantitative techniques and temporal information regarding attentive responses. A common indirect approach is to measure differences in task performance, such as reaction time [332] or accuracy [60]. Furthermore, eye-tracking is a widely used tool for measuring visual attention. Eye movement behavior exhibits a variety of features that indicate attention. Eye movements consist of saccades (rapid changes in position with a peak velocity greater than 100°/s), vergence (changes in the orientation of the two eyes), and smooth pursuit (slow movements, generally less than 100°/s, that track small, moving targets). In addition, eye fixations can be used to create scan paths or heat maps [264]. Because multiple brain areas have been shown to participate in the control of spatial attention in humans [164], attention-related activities can also be measured directly in the brain. Common brain imaging techniques such as electroencephalography (EEG), event-related potential (ERP), and functional magnetic resonance imaging (fMRI) are used for this purpose (see [264] for a review).

1.1.3.2 Sensory Substitution. Sensory substitution is a technique of transforming sensory stimuli for one sensory modality into stimuli for another sensory modality [233]. For example, a visual representation can be converted into a sound that can be heard or a tactile stimulus that can be felt [250]. Sensory substitution is considered a multisensory experience as it typically involves visual, auditory, or tactile processes [12].

Sensory substitution is generally used whenever the technology must present important sensory information that is not available in its native form [233]. For example, visual-to-tactile systems (e.g., [79, 438]) and visual-to-auditory conversion systems (e.g., [160, 283])

have been developed to assist people with visual impairments. In the field of perceptual augmentation, sensory substitution techniques can be used to access visual information either when there is an excess of visual information to process [170] or when perceptual conditions are degraded [63]. Tactile-visual sensory substitution is used, for example, to render haptic information to display textures in 3D [126], which is useful for supporting pattern-discrimination tasks [12, 157].

Sensory substitution requires user adaptation. The degree of user adaptation and training varies depending on the sensory mapping of modal inputs. It is recommended that mappings should use the strongest representation of the transposed channel to support easy user adaptation [157]. Spatial processing works best in the visual domain. Therefore, visual features are often of spatial nature, such as vernier acuity, orientation and texture, motion, and spatial frequency. Temporal processing occurs mainly in the auditory domain. The primary features of interest in the auditory domain are frequency (spectral) information and temporal information, such as the order, interval, or duration of stimuli [338]. The sense of touch is capable of processing both spatial and temporal information, although it is not as powerful as vision or hearing in either domain [422]. Typical applications for sensory substitution and examples for the corresponding mapping domain can be found in Table 1.1 [157].

Table 1.1: Examples for sensory substitution schemes (adapted from [157])

Initial Channel	**Input Domain**	**Transposed Channel**	**Mapping Domain**
Visual	Spatial	Tactile	Spatial
Auditory	Temporal (frequency)	Tactile	Sensorial intensity
Tactile	Sensorial intensity	Auditory	Temporal (frequency)

Substitution can also be used to produce sensory redundancy. For this purpose, the same sensory information is provided through different sensory channels in addition to the expected one. This process is performed to strengthen the original signal in order to increase the performance of users in complex tasks. Studies demonstrated the usefulness of redundant feedback, e.g., the provision of redundant visual, auditory, and haptic force feedback [114, 349]. However, this technique should be used carefully to avoid sensory contradictions or sensory overload. Rather than reinforcing the original signal, such methods can lead to confusion and cause reaction delays as the user copes with unexpected sources of information. The sensation of sensory overload caused by processing excessive sensory data can also impair human performance [157].

1.1.4 Situation Awareness

SA is defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [106]. SA originated in the field of aviation, but is also a basis for performance in many different areas, including air traffic control, education, military operations, and weather forecasting. SA is considered a foundation for effective decision-making and actions in complex systems [108] (Fig. 1.3). SA is composed of three different levels: (1) perception, (2) comprehension, and (3) projection. Level 1 of SA, perception, involves sensory detection of important environmental information. For example, operators must be able to see relevant displays or hear an alarm sound. In a broader context, other senses may also be relevant for information acquisition, such as smell. Level 2 of SA, comprehension, involves the understanding of the meaning or significance of this information in relation to one's own goals. For example, operators with strong Level 2 SA are able to see the immediate impact of an outage on other parts of the system. Level 3 of SA, projection, consists of extrapolating information into the future to determine how it will affect future states of the operating environment; an example would be the ability to predict the future impact on the system when an element is removed from service [108].

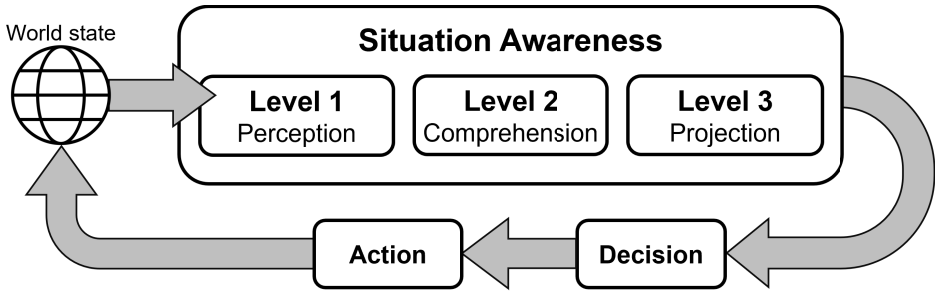


Fig. 1.3: Situation awareness in the decision-making process. Image adapted from [108].

There are several approaches to measure SA [106], including physiological measures, performance measures, subjective assessment techniques, questionnaires, and freeze-probe techniques. Physiological measures such as electroencephalographic measurements and eye-tracking methods have shown promise in determining whether information has been cognitively acquired. However, these methods cannot determine how much information has been retained in memory, whether the information has been registered correctly, or what understanding the subject has developed regarding these elements. Thus, although physiological measurements can provide useful data for other purposes, they do not seem very promising to measure SA [106]. Performance metrics have the advantage of

being objective and non-intrusive, as they are generated through the natural flow of the task. Specified performance data, such as speed and accuracy [107], can be recorded automatically in simulated systems, making it relatively easy to collect the necessary data [108]. However, there are some problems related to the relationship between SA and performance; for example, an expert participant may achieve acceptable performance even when their SA is inadequate [361]. Subjective techniques can be divided into self-rating and observer-rating. For self-rating, operators are asked to subjectively rate their own SA, usually on a 1–10 scale. Self-assessment methods are quick and easy to use and do not interfere with task performance because they are performed post-trial. However, this condition can lead to a rating that has been highly tainted by the outcome of the trial. In subjective observer-rating, independent, expert observers evaluate the quality of a subject's SA. The main advantages of using observer-rating scales to measure SA are that they are non-intrusive and can be applied "in the field". However, the extent to which observers can accurately rate participants' SA remains questionable [106, 361]. Questionnaires can be used to collect detailed information regarding SA that can be compared to reality, thus providing an objective assessment of operator SA. A detailed questionnaire can be completed at the end of each trial so that operators have ample time to answer a detailed list of questions about their SA during the trial. However, one disadvantage of post-test questionnaires is that they cannot reliably capture the subject's SA until the very end of the experiment [106]. Finally, freeze-probe techniques were introduced to overcome these limitations of SA reporting. In these techniques, the simulation is frozen at randomly selected points in time and operators are asked about their perception of the situation at that time. Operator perceptions are then compared to the real-world situation to provide an objective measure of SA. The Situation Awareness Global Assessment Technique (SAGAT) is one of the most commonly used SA measurement techniques, eliminating several problems associated with post-trial testing and subjective SA data. One drawback, however, is that freeze-probe techniques cannot be used "in the field" because it is not possible to freeze a real-world scenario [106].

Maintaining SA is also important for decision-making in many AR applications. It has shown that SA regarding the real-world environment decreases when visual tasks are performed in AR [194]. The shortage of SA is caused by increased distraction from the real world because AR requires a high level of concentration [13, 78] and human cognitive capacities are limited [194]. It is assumed that a higher cognitive load usually leads to decreased SA performance, which in turn increases the risk of an accident. Limited SA is

also increasingly becoming a social problem when AR applications are used in everyday life, as it can cause accidents [13, 78]. For example, collisions with incoming objects or traffic accidents can occur when an AR task requires movement or walking, demonstrating that SA is particularly important for performance and error prevention in safety-critical domains [453]. Therefore, a component is needed to notify the user about risks or further information regarding the environment so that the user can allocate cognitive resources accordingly [194].

1.2 Sensory Constraints

Sensory constraints are **intrinsic and extrinsic factors** that affect the perception of augmentations. Most research has addressed sensory constraints that affect perception in the visual processing and interpretation pipeline, also referred to as to perceptual pipeline (see [221] for an overview). Although AR is currently mainly focused on the visual sense [17], augmentations can also be applied to other sensory channels, including hearing and touch [14]. Because the present work has introduced the provision of augmented content through mainly audio-tactile methods, the auditory and tactile sensory channels may also be affected by perceptual impairments. Fig. 1.4 provides an overview of relevant sensory constraints that have been studied in this work and how they potentially affect the perception of augmentations. In addition, other constraints likely exist that affect perception, for example, those directly related to other technologies such as VST-HMD devices (see [221]). However, such constraints are not part of this work, as we are specifically investigating the factors that influence OST AR.

Intrinsic factors are user-related aspects. These factors include limitations in depth perception, disparity planes, and thresholds for sensory receptors of the auditory and tactile channels [176]. Extrinsic factors are divided into environment-related and device-related issues. Environment-related constraints consider the structure and layout of the scene, background features, and the influence of visual and auditory noise on augmentations. Device-related issues concern the FOV and display characteristics such as screen brightness and reflections.

We have noted that the problems arising from individual sensory constraints are often interrelated and thus can affect each other. This interplay can lead to a greater impairment in the perception and cognition of augmentations, which in turn can have a negative impact on performance in AR [221, 346]. Issues can affect perception across display technologies,

for example, head-worn displays, handheld mobile devices, and projector-based systems [221]. In the context of this work, sensory constraints have been specifically focused on OST HMDs. However, the results of this work can likely be applied to other display technologies as well. Sensory constraints and their effects are discussed below.

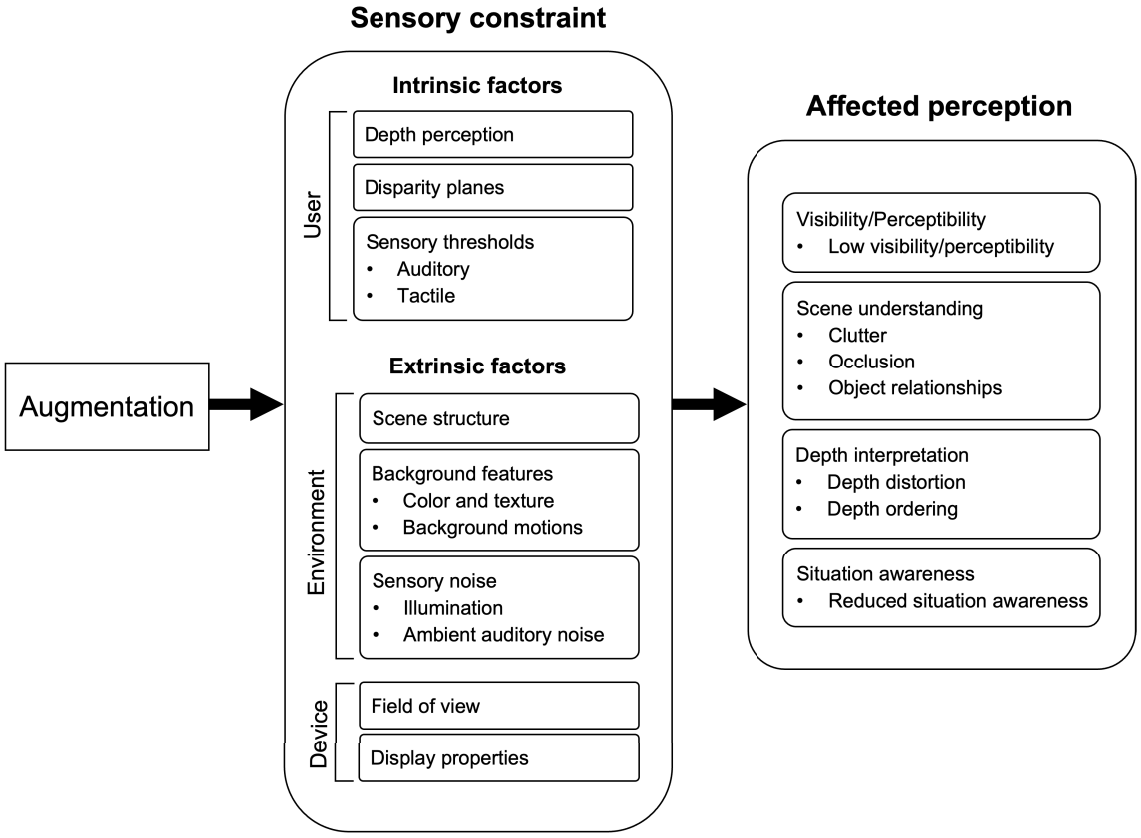


Fig. 1.4: Intrinsic and extrinsic factors of sensory constraints that can affect the perception of augmentations.

1.2.1 Intrinsic Factors

Intrinsic factors are issues associated with the user's perception of the augmented content. These issues occur at the final stage of the perceptual pipeline introduced in [221]. The augmented content presented through the display device is influenced by users themselves, resulting in highly individual differences among users. Factors include sensorial impairments in depth perception, binocular disparities, and limitations in auditory and tactile perception [221, 346].

1.2.1.1 Depth Perception. Incorrect depth interpretation is one of the most common problems related to perception in AR. This concept refers to the interpretation and

interaction of spatial relationships between the user's point of view, the objects in view, and the superimposed information. Problems with depth perception hinder users in correctly matching the overlaid information with the real world. Depth cues can be divided into pictorial depth cues, kinetic depth cues, physiological depth cues, and binocular depth cues. Pictorial depth cues include occlusion, height in the visual field, relative size, aerial perspective, relative density, relative brightness, and shadows. Kinetic cues provide depth information obtained by using relative motion parallax and motion perspective to change the viewpoint. Physiological depth information is provided by the muscular control systems of the eyes and includes vergence, accommodation, and pupil diameter. Binocular disparity provides depth cues by combining the two horizontally offset views of the scene provided by the eyes. Among all depth cues, occlusion is the most dominant. However, when only a limited number of depth cues are available, problems such as depth underspecification, inconsistencies, or contextual biasing may arise [221].

Impairment: Ambiguous depth cues lead to incorrect depth ordering of augmentations and depth distortions, making the overlaid information inconsistent with the real world.

Approach: Multisensory guidance can help in assessing the correct depth of target information by providing intuitive audio-tactile distance metaphors that convey depth information in a non-visual way. Furthermore, additional sensory cues in the longitudinal and latitudinal planes can assist in accurately matching the target in depth. This strategy can help restore object relationships in AR and mitigate typical problems associated with depth, such as underestimation.

1.2.1.2 Disparity Planes. Real and virtual objects can have different binocular disparities, which can lead to perceptual problems related to disparity levels and disparity areas. Human distance perception is based on the relative angle between the alignment of both eyes and their focal depth [93]. Common HMDs display virtual objects at different distances from the viewer while the focal plane remains constant (e.g., at two meters in the case of HoloLens) [309]. A disparity plane defines the depth disparity with which the content is viewed. Focal depth often refers to groups of objects that are in similar disparity planes, also called disparity areas. Depth disparities often occur in dual-view AR systems [221] in which augmentations exist in one disparity area and the real world exists in another. Since the areas are at different depths, users must switch their vergence between disparities to compare content. In head-worn systems such as OST devices,

different elements of augmented content may be placed at different depths, resulting in an offset between individual elements. This effect can require users to switch back and forth between disparities, which can lead to reduced awareness in currently unobserved disparities [268] and visual fatigue [221]. This switching between different depth planes also incurs access costs (e.g., the time to rotate eyes and adjust to different planes when switching between real and virtual content), which have a measurable impact on the usability of AR [36]. Methods have been introduced to resolve disparity-related issues to some extent (see [220] for an overview). Hardware-based approaches include, for example, multi-focal displays that use deformable membranes, tunable lenses, and parallax barriers. However, such systems simulate or support the perception of different focal distances only to a limited extent and are still not commercially available [36]. Furthermore, software-based methods such as gaze-contingent blurring [99] are not yet able to provide entirely correct focal cues [36, 220].

Impairment: Information located at different disparity planes leads to depth distortion problems. This can affect task performance and SA.

Approach: Multisensory guidance can potentially reduce depth distortion problems by substituting visual information across disparities with audio-tactile cues. This approach can potentially improve SA by reducing access costs and supporting the allocation of attention resources to relevant disparity areas.

1.2.1.3 Sensory Thresholds.

Sensory systems are used to receive stimuli regarding the environment and the internal state of the body [135, 445]. In terms of environmental stimuli, specialized sensory receptor cells convert external stimuli into neural impulses that the brain and nervous system can use; this process is called sensory transduction. Different kinds of stimuli require different kind of sensory receptors in order to be detected. Sensory thresholds can be divided into absolute thresholds and difference thresholds. An absolute threshold is the minimal stimulus necessary to be detected. The smallest difference that can be detected between two stimuli is called the difference threshold [135]. Below, we discuss the sensory thresholds of the sensory systems of hearing and touch, as the development of multisensory cues has been based mainly on these two modalities. We then discuss how the perception of each modality is affected.

Thresholds for Auditory Perception. Audio perception is based on pressure changes (sound waves) transmitted through the air or another medium. These pressure changes are converted into electrical activity in neurons in the auditory system and transmitted through the auditory nervous system [331]. Sound waves are characterized by their amplitude (size of the pressure change) and their frequency (number of times per second the pressure changes repeat). The amplitude is associated with the perception of the loudness of a sound and is reported on a logarithmic scale called decibels (dB). Decibel is an expression for the ratio between the amplitude of the primary sound and that of the background sound and provides a measurement of the ability to hear what is intended [135, 346]. Normal speech is perceived between 50 dB and 70 dB. The perceived loudness of a sound depends on both the sound pressure (dB) and the frequency (Hz), as shown in Fig. 1.5 [135]. The frequency is associated with the perception of pitch and is indicated in Hertz (Hz). Individuals with normal hearing can perceive frequencies ranging from approximately 20 Hz to 20,000 Hz [135]. The smallest frequency change that a normal hearing adult can perceive is approximately 0.2–0.3% at frequencies of 250–4000 Hz and increases rapidly with increasing frequency [254, 331]. The absolute intensity threshold of humans for the most sensitive mid-frequency range is up to 0 dB [331]. However, sounds below 20 dB are not distinguishable because the ear cannot detect frequency changes below this level [346]. In addition, tones below 10 dB at very high or very low frequencies cannot be heard [135].

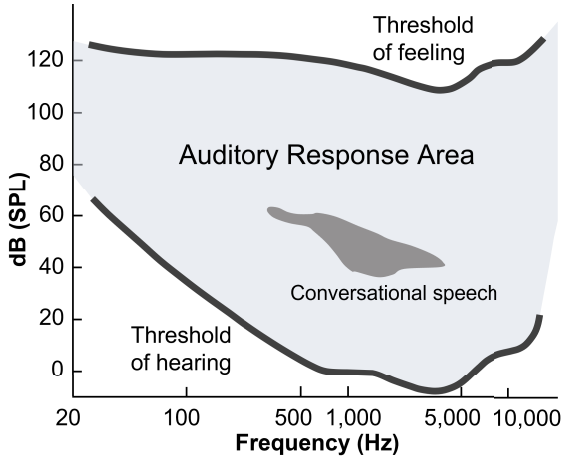


Fig. 1.5: Auditory response area: Hearing occurs in the highlighted area between the threshold of hearing and the threshold of feeling (adapted from [135]).

Binaural audio cues, namely the interaural time difference (ITD) and the interaural level difference (ILD), help to locate the position of a sound source in space by comparing the sound signals reaching the left and right ears [135]. The ITD describes the time difference

between when a sound reaches the left and right ears. When the sound is located in front of the listener, the sound reaches both ears simultaneously. When the sound is located laterally, the magnitude of the ITD increases toward that side. A sound located at 90° has an ITD of 0.6 ms. ITDs are dominant for frequencies below 1000 Hz [135, 251]. The ILD describes the difference in sound pressure level reaching the two ears. The difference between the two ears occurs because the head dampens the intensity of the sounds that reach the more distant ear compared to the ear that receives the sound unobstructed. ILDs are dominant for signals with frequencies above 1500 Hz and are ambiguous for angles larger than 60° [123, 135, 251].

In addition to the ITD and ILD, the head-related transfer function (HRTF) plays an important role in spatial hearing. Sound waves are affected by the head, pinnae, and torso before reaching the eardrum of the listener [179]. The HRTF describes the physical change in the sound wave in the frequency domain caused by this effect. Due to the asymmetrical shape of the head and pinnae, the HRTF varies with the direction of a sound source. The HRTF also varies per listener due to individual differences in shape of the head and the pinna. With the combination of ITD, IID, and the result of an (individual) HRTF, the brain deduces from which direction a noise has originated. In addition, acquired knowledge helps alongside the HRTF when ambiguities arise. However, for narrow-band signals, the sound is sometimes localized to the front although the sound source is actually at the back, and vice versa. This phenomenon is called front-back confusion [179].

Thresholds for (Vibro-)Tactile Perception. Tactile sensations are perceived via mechanoreceptive units in the outer layers of the skin, which transmit signals to the brain when activated. Therefore, the stimulation of mechanoreception allows tactile perceptions such as pressure and vibration. The distribution of receptors is not uniform across the skin and differs according to skin characteristics (e.g., hairy vs. hairless skin) [76].

Among the different classes of mechanoreceptors, Merkel cells are most commonly used for pressure sensation and respond to very low frequencies between 0.4 Hz and 100 Hz, working best at approximately 7 Hz. Meissner corpuscles respond to low-frequency vibrations between 10 Hz and 200 Hz, being most sensitive at 50 Hz [26, 76]. The Pacinian corpuscles are the most responsive receptors in the fast-acting receptor class because they respond quickly to changing stimuli. The perceived vibration intensity of the Pacinian corpuscles varies as a function of both frequency and amplitude. Their frequency range is from 40 Hz to 800 Hz, working best at about 250 Hz and rapidly decreasing at frequencies

below 50 Hz or above 600 Hz. Regarding the spatial resolution of the skin, studies have shown that the minimum distance between two vibration signals must be between 0.8 and 1.2 mm, depending on the signal frequency, in order to be correctly discriminated [76].

Impairment: Sensory thresholds affect the perception of the respective modality: the audibility of acoustic signals and the perception of vibrations.

Approach: We have addressed these sensory constraints by ensuring that sensory stimuli are always provided in a range that is easily perceivable by humans. Auditory signals consider sensitive ranges of human hearing in terms of frequency and amplitude and take advantage of localization capabilities such as the HRTF. Tactile methods consider the properties of the skin, such as the perceptible frequency range and spatial resolution, for vibrotactile stimulation.

1.2.2 Extrinsic Factors

Extrinsic factors are perceptual problems that originate in the environment and the display device. Relevant environmental issues include the structure and layout of the environment, background features, and sensory noise. Sensory constraints related to the display device refer to technical issues that are mainly associated with the screen. The resulting problems concern the available screen space, namely the FOV, and display properties, such as screen brightness and contrast [221].

1.2.2.1 Scene Structure. The richness of information of an environment, for example the arrangement of the contained objects, can lead to a cluttered view. A cluttered view contains too many salient features, affecting the general understanding of the scene. A large amount of information can also stretch or exceed the limits of short-term memory, making recall of the number of objects and their features difficult [354]. Clutter also causes occlusion problems, in which augmented information occludes real-world or other augmented information. Objects can be completely occluded or only partly visible, leading to incorrect depth ordering and reduced legibility of augmented content [221]. In-view labeling methods in AR can exacerbate the problem in dense environments, as they attempt to add additional labels within the FOV that refer to objects outside the FOV [224]. The problem of a dense scene structure in the context of AR devices is illustrated in Fig. 1.6. The excess of information leads to a degradation in task performance [354]. In this context, it has been shown that detection capacity degrades over time when there is an imperative

to search for infrequent targets that are embedded in more frequent, non-target distractors [159]. Related to search, clutter leads to decreased performance in the recognition of objects due to occlusions. Furthermore, clutter produces difficulties at both segmenting a scene and performing visual search, resulting in increased reaction times [28, 354].

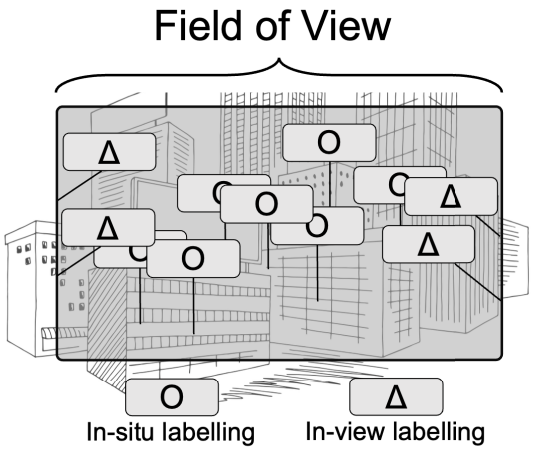


Fig. 1.6: Illustration of the problem of a dense information space. This problem is also affected by other constraints, such as FOV size. The visualization is compressed onto a (potentially small) displayable area, leading to problems such as overlapping and illegibility of the augmented information.

Impairment: A dense scene structure leads to reduced scene understanding, including a cluttered view and obscured information, that further affects depth ordering and visibility of augmentations.

Approach: Clutter and occlusion caused by dense scene structures can be reduced by substituting visual information into audio-tactile cues. This may reduce clutter and occlusion problems so that the relationships between augmentations and objects can again be represented in a more understandable way. In addition, multisensory guidance can be used to make target positions in dense information spaces more explicit by specifically guiding the user regarding longitude, latitude, and depth.

1.2.2.2 Background Features.

Background features describe issues related to color schemes and texture patterns as well as motions in the environment.

Color and Texture. Real-world color and background textures in the environment directly affect the visibility and legibility of augmentations. This is often the case with OST devices when the color or brightness of a real-world background conflicts visually or perceptually with the color or contrast of the augmented elements, resulting in poor

readability [129, 130]. This problem is especially critical problem in application domains in which color encoding is essential [128]. Surfaces with high variation in color and texture (patterns) can affect the visibility of projections. This is the case when the environment has a similar pattern to the augmentation pattern, leading to perceptual interferences [221]. Such problems become more significant under changing light conditions [129, 221]. Finally, studies have shown that user performance in a visual search task is significantly affected by the background color and texture as well as illumination at the background's position [130].

Background Motion. The visual feedback from the real world may include significant background motion and clutter (e.g., a busy city street), which can hinder user perception in AR [117]. In general, the search for a stationary target among moving distractors, such as a motion-rich background, is considered difficult. Reports have demonstrated that searches for a stationary target within a structured flow field are more efficient than searches for stationary targets among distractors moving in random directions [357]. Regarding AR, it has been shown that the perception of augmentations is partially affected by clutter and movement in the background [118]. Conflicting motion cues between background and augmentations can lead to more difficult judgements because there is no consistent point of reference. Background movement and clutter can also temporarily occlude the augmented content [268]. In addition, the presence of background motion can lead to distraction from the actual AR task [118]. Because real-world background distractions are expected to be on different disparities from virtual augmentations [221], users might unconsciously switch their vergence on this area. However, there is still a lack of fundamental understanding regarding how background motion affects perception in AR, requiring further research [118].

Impairment: Color and texture of the background, as well as motion, affect the visibility and legibility of augmentations. Background motion can lead to temporal occlusions of augmentations. Furthermore, motions in the background can lead to distractions, which in turn may affect search performance.

Approach: We have addressed background features essentially by recoding visual augmentations into audio-tactile stimuli. Thus, the search for information is not directly affected by the color, texture, and motions of the background. Focal switching between disparities caused by background distractions could be indirectly addressed by providing multisensory guidance cues. Salient audio-tactile stimuli could potentially support the allocation of attentional resources for target search in the presence of background distractions.

1.2.2.3 Sensory Noise.

In general, noise is defined as "random or irregular fluctuations or disturbances which are not part of a signal" or as "distortions or additions which interfere with the transfer of information" [312]. In the context of this work, sensory noise is associated with the exposure to environmental stimuli that may affect the perception of augmentations. This concerns exposure to ambient illumination as well as environmental auditory noise.

Illumination. The state of the environment can affect the perception of AR content drastically [221]. A major challenge in the presentation of augmented information involves uncontrolled environmental conditions, namely lighting and background conditions [129, 221]. Lighting conditions can be divided into indoor and outdoor or natural lighting. Indoor lighting is typically in a range of 100–1000 lux [110]. Although indoor lighting is of relatively low intensity, it can lead to various issues related to color and contrast representation as well as reflections on the screen [221]. On the other hand, outdoor lighting can cause large-scale fluctuations varying from 1–100,000 lux or more [129]. Highly varying light intensities affect the quality and correctness of imaging by underexposing or overexposing the augmented content. Very bright outdoor lighting can result in very low contrast ratios of 3% or less for current OST devices (tested on Microsoft HoloLens 2) [110]. Therefore, bright environments can limit projection, making it difficult to accurately distinguish virtual imagery presented on the HMD [110, 221]. For both indoor and outdoor lighting, strong ambient light can lead to reflections and lens flare [221].

Ambient Auditory Noise. AR technology can also be used to augment the physical world with auditory cues [14]. These cues can be generated, for example, by reading content aloud via text-to-speech or by playing alarm signals to attract the user's attention [346].

Ambient auditory noise is defined as noise emitted from all sources (with the exception of noise at an industrial workplace). The main sources of noise include road, rail, and air traffic, industry, construction and public works, and the neighborhood. The most common sources of indoor noise include ventilation systems, office machines, household appliances, and neighbors. According to the EU, high environmental (i.e., outdoor) noise is defined as a noise level above 55 dB during daytime. Noise pollution is especially severe in cities and is mainly caused by traffic along densely trafficked roads. Here, the equivalent sound pressure levels for 24 hours can reach 75–80 dB [33].

Noise is a complex pattern of sound waves that can originate from different sources and can be labeled, for example, as music or speech. However, noise is considered as unwanted sound. Most environmental sounds consist of a complex mixture of many different frequencies [33]. Typical sound levels include library (indoor, 40 dB), normal conversation (60 dB), and heavy road traffic (outdoor, 80 dB) [135]. Ambient noise can produce many difficulties in auditory perception, including impairing speech intelligibility, masking important acoustic signals such as alarms and warnings, causing annoyance, and acting as a distracting stimulus [33]. Thus, the perception of auditory feedback in AR may be impaired when exposed to ambient auditory noise.

Impairment: Sensory noise affects the visibility and perception of visual and audible augmentations.
Approach: The use of multisensory guidance through audio-tactile cues is largely unaffected by excessive lighting. Auditory and tactile cues are presented in an intensity and frequency range that is sensitive to human hearing, so the signals should remain perceptible even at higher noise levels.

1.2.2.4 Field of View.

The FOV refers to the extent of the observable world at any given moment [221]. The FOV related to AR describes the overlay FOV, in which computer-generated graphics are overlaid on the image of the real world. A wider FOV would result in the display of more information to the user in a single view [368]. The binocular FOV of human vision is approximately 210° horizontally and 150° vertically [196]. However,

OST HMDs usually provide only a relatively small FOV of approximately 60° [11, 210] to display the augmented content. For example, the widely used Microsoft HoloLens 2 has a 52° diagonal FOV [286]. In comparison, VR devices such as the HTC VIVE Pro 2 and the Oculus Quest 2 offer a wider FOV of approximately 110° diagonally [294]. Typical problems that occur in AR with small FOVs are occlusion and a cluttered view when too much data is presented at once, making comprehension of the data difficult [368]. Furthermore, augmented information is frequently located out-of-view in narrow FOV AR devices, leaving the user unaware of the presence of that information. This deficiency requires the use of additional methods, mostly in visual form, to locate objects that are outside the current FOV. However, such methods often occupy a large portion of the available screen space in narrow FOV systems, leading to further visual ambiguities and occlusion problems [44, 221, 224]. Although experimental OST HMDs with wider FOV ranges of approximately 100° diagonally theoretically exist, their development remains a technological challenge and they tend to be expensive and heavy [209]. Fig. 1.7 shows a comparison of the FOV of common OST AR display systems compared to human vision. The figure shows that even current head-mounted AR devices still cover only a fraction of human visual perception. Related to search, outcomes have indicated that the use of wider FOVs produces higher performance (shorter search times) compared to the use of smaller FOVs [11, 406]. However, negative effects induced by a narrow FOV can be alleviated by considering the use of appropriate view-management methods [208, 224].

Impairment: A typical small FOV can result in a cluttered view and object occlusions, especially in dense scene structures, impairing the understanding of the scene.

Approach: Multisensory guidance cues can be used to de-clutter the scene structure by substituting visual information with audio-tactile cues. This can help reduce occlusions and restore object relationships. In addition, multisensory guidance provides target localization information beyond the boundaries of the current FOV. This strategy can effectively guide users to locations outside the FOV.

1.2.2.5 Display Properties.

Relevant display properties for this work mainly include screen brightness, contrast, and resolution. Screen brightness refers to the luminance of a screen and varies between approximately 250 and 500 candela per square meter (cd/m^2). The brightness of the screen affects the visibility of the augmented content, especially when blended with ambient light. This effect results in poorly visible representations due to decreased contrast, which can be expressed as the ratio of the luminance of the

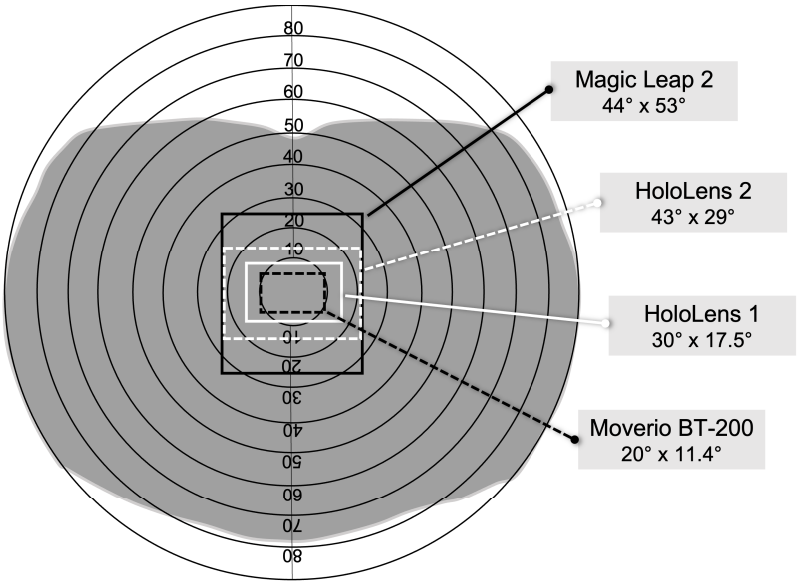


Fig. 1.7: FOV comparison of common head-worn AR devices. For the purposes of illustration and direct comparison, the FOVs are centered around the central vision. However, HMDs are not necessarily centered during real-life use. Image adapted from [406].

brightest color (white) to that of the darkest color (black) that the screen can produce. Under such conditions with increased light exposure, shiny objects in the environment can produce reflections on the screen such that the displayed content becomes nearly invisible. Furthermore, the resolution refers to the number of pixels a screen can display. High resolution results in the perception of a sharp image. However, sharp rendering can affect depth perception, as objects in focus are perceived to be closer than they actually are. Finally, displays with a high pixel density are able to render very small objects. This in turn can lead to problems regarding object detection and segmentation [221].

Impairment: The display properties mainly affect the visibility and legibility of augmentations, especially when exposed to higher levels of ambient light.

Approach: Multisensory guidance can be used independently of display properties by providing audio-tactile cues. As light exposure increases and display contrast and visibility decrease, audio-tactile cues can be used to supplement the visual representation to aid target guidance. If the lighting conditions exceed the display's capabilities, audio-tactile cues can perform the target guidance entirely until a stable visual presentation is again possible.

1.3 Research Questions and Contributions

The creation of a multisensory guidance system for use under sensory constraints in AR faces various problems that directly affect perception, cognition, and performance. For this purpose, we formulated RQ_{Main}, which is the overarching research question addressed in this dissertation:

RQ_{Main} :**What are the potentials and limitations of multisensory guidance to support search under sensory constraints in AR?**

To be able to answer RQ_{Main}, we defined subordinate research questions RQ_{1-3} to explore different foci that are related to leading research question. In the following paragraphs, we list the research questions and describe the approach used to address them. Fig. 1.8 illustrates the different stages of this thesis and how the individual chapters contribute to one another. In this way, we present how this dissertation contributes to answering each specific research question to address the overarching research question RQ_{Main}.

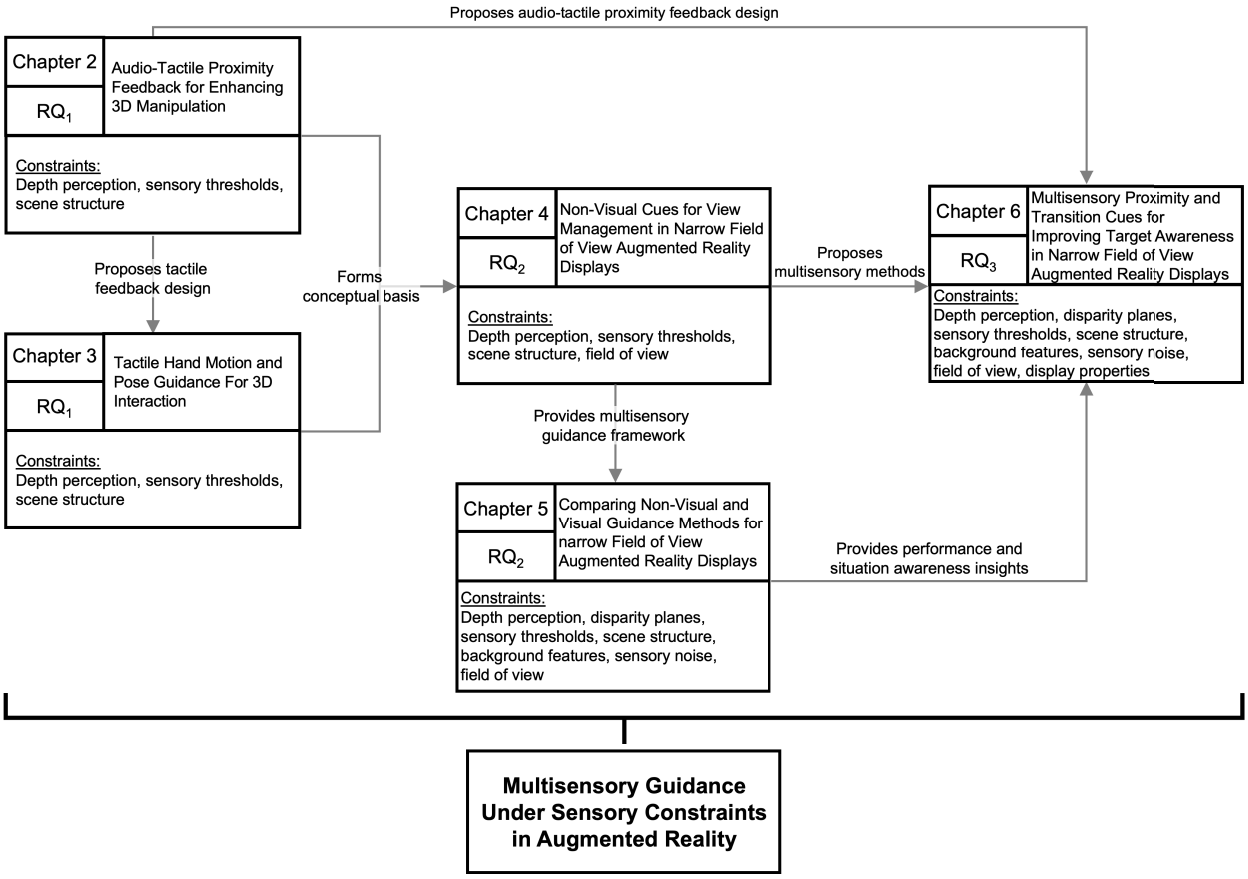


Fig. 1.8: Outline of the stages of this work. The chapters explore research questions under specific sensory constraints to collectively answer the overarching research question RQ_{Main}.

RQ_1 :What is the effect of hand-based non-visual guidance on task performance in visually complex environments?

Regarding RQ_1, we investigated possible multisensory guidance approaches that address primarily sensory constraints in the visual domain. Vision is often constrained in selection and manipulation tasks, for example, in assembly or maintenance scenarios that have a high visual complexity. Due to the increased occurrence of clutter or occlusion, the target position may not be visible at all times, resulting in unwanted object collisions or overshoot errors during interaction. To address these problems, we examined the limitations of existing non-visual approaches and explored how they could be extended in terms of directional guidance in dense scene structures. This process led to the development of two non-visual guidance approaches, audio-tactile proximity feedback and directional tactile cueing patterns, that were evaluated in 3D interaction and manipulation tasks. These approaches used a vibrotactile attachment for the hand that was enhanced by auditory cues to guide the user in visually complex scenes in the absence of visual cues.

Contribution 1

Contribution 1 is composed of the insights from Chapter 2 and Chapter 3 to answer RQ_1. The results of Contribution 1 can be summarized as follows:

- The provision of two non-visual guidance approaches: audio-tactile proximity feedback and motion guidance by tactile motion patterns.
- The finding that both guidance approaches contribute to improving task performance in 3D selection and manipulation tasks in potentially dense information spaces without visual aids:
  - Audio-tactile proximity feedback aids spatial awareness and helps to reduce errors (collisions and object pass-throughs) by providing audio-assisted, higher-resolution tactile feedback.
  - Tactile motion patterns help users perform finer 3D selection and manipulation tasks by triggering changes in hand posture and movement.
- The provision of a conceptual and technical basis for further research within the scope of this thesis.

It should be noted that the research presented in Chapter 2 was developed and evaluated largely as part of my master's thesis. This work explored the use of audio-tactile proximity cues in potentially dense information spaces. Within the scope of this work, the knowledge

gained has been further analyzed with respect to sensory constraints. In particular, this analysis concerns the application of the developed concepts to the sensory constraints of depth perception, sensory thresholds, and scene structure. In addition, technical limitations, perceptual and cognitive aspects of non-visual audio-tactile feedback mechanisms, and implications for task performance in AR guidance were further examined. The methods presented in this chapter establish important foundations for the research in Chapter 3. In Chapter 3, the use of directional tactile cueing patterns in environments with high information density is examined. The method developed in Chapter 2 informed the motion guidance feedback design described in Chapter 3. Although the methods in both chapters were evaluated under VR conditions, they also provided valuable insights into multisensory guidance methods under potential further sensory constraints. The results described in Chapters 2 and 3 established the theoretical basis for the multisensory guidance approaches that were refined for the methods presented in Chapter 4. Finally, the methods developed in Chapter 2 led to the initial audio-tactile proximity feedback design described in Chapter 6. These interrelationships are illustrated in Fig. 1.8.

Chapter 2: Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation. Chapter 2 considers the challenges of performing 3D selection and manipulation in the presence of conflicting or ambiguous visual cues in complex scenes, such as a virtual training assembly procedure. For this purpose, we created a novel, glove-based, tactile interface incorporating 18 vibration motors distributed over the hand. The tactile feedback was further enhanced by auditory cues to provide proximity guidance. Compared to other work in the field of audio-tactile, proximity-based, selection assistance [10], our approach contained a higher-density tactile grid to obtain directional information regarding multiple objects. As a result, we have developed two feedback models utilizing audio-tactile guidance: outside-in feedback for scene exploration and inside-out feedback for manipulation tasks. When outside-in feedback was used, each object in the scene emitted a signal. Thus, the feedback was spatially tied to the objects in the scene. In contrast, inside-out feedback provided cues relative to objects targeted for interaction. Objects radiated signals into the scene to provide spatial information to the user. Both models provided proximity information through directional and distance-based modulated vibrotactile signals and spatial audio cues. In two user studies (n = 12), we evaluated how such audio-tactile proximity cues could be used to inform the user about the presence of surrounding objects. The results showed that our approach with fully directional feedback

extended previous findings regarding single-point, non-directional proximity feedback [10], which had suffered from limited dimensionality. Both studies showed that the use of scene-driven outside-in and object-driven inside-out proximity cues could enhance spatial awareness significantly for exploration and manipulation tasks. In addition, performance improvements could be achieved by avoiding unwanted object collisions and reducing overshoot errors. These improvements became evident for our system, which provides higher-resolution feedback compared to previous work that used lower-resolution tactile grids [344]. Finally, we showed that audio-tactile guidance in the form of proximity cues could be highly useful for 3D interaction in applications that suffer from visual conflicts such as object occlusions.

Chapter 3: Tactile Hand Motion and Pose Guidance For 3D Interaction. In Chapter 3, we present how motor planning and coordination can be enhanced by using a forearm-and-glove tactile interface. The developed interface incorporated a high-resolution tactor grid of 21 vibration motors distributed over the forearm and hand of the user. By triggering different tactile patterns in specific areas, the user could be guided to take specific motion and pose actions related to selection and manipulation tasks. Such guidance can be particularly useful for 3D interaction, especially for applications that suffer from visual occlusions. We extended previous work on fine hand motion and pose guidance for manipulation actions, compared to vibrotactile cues for body and arm motions (e.g., [23]) or general body motions (e.g., [41]). Overall, 24 subjects participated in three user studies to validate the tactile guidance cues. In study 1 (n = 8), we showed that users could localize and differentiate tactile cues at different arm and hand locations reasonably well. In doing so, we also demonstrated that stimulation of rather unusual and infrequently used areas, such as the back of the hand, can also be useful locations for contact-driven feedback. In study 2 (n = 8), we explored tactile pattern interpretation and preferences. Although the recognition and interpretation of more complex tactile stimuli – tactile pattern as prompts to adopt a particular movement or pose – worked well, they were not without errors. This result was to be expected, as other work [380] has shown that users interpret patterns as either push or pull motions. However, most users were able to successfully match tactile patterns to be guided to the correct motion or bodily reconfiguration in study 3 (n = 8). To achieve a better interpretation of tactile patterns, we expected that lower-level abstraction design or personalized patterns would be beneficial. With respect to hand guidance, we showed through a Wizard-of-Oz experiment that our tactile pattern could

trigger fine-grained motions and poses that could support 3D selection and manipulation. Therefore, we achieved a similar granularity to EMS-based methods [386] while avoiding their disadvantages. With the results of the user studies, we obtained a robust basis for implementing further tailored guidance cues that could likely be coupled with visual cues, for example.

RQ_2 :How effective are head-based multisensory guidance cues to support search under sensory constraints?

RQ_2 specifically addresses the sensory constraints of depth perception, scene structure, and FOV for search tasks using head-mounted AR systems. Due to limited screen space, information of interest is frequently located outside the FOV or within a cluttered view in dense AR scenes. Furthermore, target localization in 3D space can be ambiguous, especially in dense scene structures that contain many distracting items. To approach this research question, we evaluated different approaches of non-visual sensory cue combinations to guide users in longitude, latitude, and depth within narrow FOV AR. To do so, we developed an audio-tactile guidance system to attach on OST AR devices, investigating the suitability of non-visual directional cues at the user's head. Thus, we aimed to select the best performing method to present multisensory directional cues that could direct attention to objects outside the FOV as well as specific locations in dense information spaces. We then compared this novel audio-tactile technique with a state-of-the-art visual guidance method called EyeSee360 to evaluate the effectiveness of non-visual guidance in search performance.

Contribution 2

Contribution 2 builds on the results from Chapter 4 and Chapter 5 to answer RQ_2. Chapter 4 explores the use of non-visual directional cues for object localization. Chapter 5 addresses the effectiveness of using multisensory guidance. The results of Contribution 2 can be summarized as follows:

- The provision of head-based audio-tactile cues for encoding longitudinal, latitudinal and distance information guidance in the absence of visual cues. The resulting system would form the methodological framework of this thesis.

- Effective multisensory guidance to support search when perception is impaired by sensory constraints, such as depth ambiguities, dense scene structures, and a narrow FOV.
- The comparison of audio-tactile guidance with a state-of-the-art visual guidance method, demonstrating that audio-tactile guidance, while generally slower, is comparatively reliable in terms of target finding.
- First findings that SA can be significantly higher when performing a guided search task with audio-tactile guidance compared to visual guidance. Results suggest that audio-tactile guidance is a viable alternative to traditional visual guidance methods for search tasks that require increased SA.

Chapter 4: Non-Visual Cues for View Management in Narrow Field of View Augmented Reality Displays. In Chapter 4, we directly address head-worn devices with a narrow FOV, which are a common commodity in AR technology. This technical characteristic can potentially lead to visual conflicts in view management, such as overlapping information and visual clutter. We considered the potential of using audio and vibrotactile feedback to guide searching and localization of directional information. For this purpose, we created a novel vibration feedback mechanism attached to the Microsoft HoloLens to provide vibrotactile feedback along the temples and forehead. Thus, non-visual cues regarding the location of augmented information could be provided by vibrotactile and spatial audio cues. This approach would be expected to be particularly useful in AR environments with high visual information density, as multisensory methods should potentially reduce visual ambiguities such as clutter and occlusion. To assess different aspects of non-visual guidance, we conducted three user studies ($n = 12$). The first study explored different cue combinations (referred to as modes) of audio and tactile cues for encoding longitudinal, latitudinal, and distance information guidance in the absence of visual cues. We found that the mode encoding latitude with audio cues and depth with vibrotactile pulse bursts exhibited the highest accuracy in latitude estimation as well as the highest subjective preference. In addition, users made accurate depth estimations of a target with only minor deviations using all modes tested. The second study examined the same modes in a guided search task in which numerous visual distractors were present to simulate scenes with a high information density. Results showed that latitudinal precision and performance time were significantly better when auditory cues were used. Of note, this insight contradicts previous conclusions [88] that unimodal vibrational feedback is superior. The third study

examined the usefulness of audio-tactile cues to determine absolute longitudinal position and distance (rather than the relative feedback used for guidance) for localizing information. Here, it was shown that target localization worked well using both auditory and vibrotactile pulse feedback. Finally, users were able to judge depth more precisely if the target was located nearby rather than far away.

Chapter 5: Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays. Chapter 5 considers the effectiveness of non-visual guidance compared to visual methods commonly used in AR. As problems with narrow FOVs still persist, visual guidance approaches tend to occupy a large part of the limited screen space. This condition can typically lead to search performance issues and decreased awareness of the physical environment. To evaluate the performance and SA capabilities of our non-visual guidance approach, we compared it with the state-of-the-art visual guidance technique EyeSee360. Three user studies (n = 16) were conducted to evaluate both approaches in solving a guided search task using a narrow FOV device. Targets and distractors were densely distributed in 3D space, including variations in distance, to further investigate the effect of depth perception between the two guidance approaches. In the first user study, audio-tactile guidance and EyeSee360 were used in solving a simple object-collection task to examine general task performance. It was shown that audio-tactile guidance could compete with EyeSee360's accuracy in terms of hit rate. However, search times were significantly slower for audio-tactile guidance than for visual guidance. In study 2, the difficulty of the task was increased by adding ambient auditory noise and background colors and motions to more closely approximate real-world conditions. Furthermore, a small noticeability test was implemented to indicate the influence on SA. Results showed that the increased difficulty due to ambient auditory noise and background features did not affect search performance for either guidance method. However, the noticeability test provided initial evidence of higher SA when the audio-tactile mode was used. In user study 3, the task difficulty was increased again by adding a secondary visual task while the main task remained unchanged. It was shown that audio-tactile guidance performed significantly better in measures of SA performance compared to the visual approach in the dual-task condition. This effect may be attributed to focal disparities, as users tend to focus on the AR plane to primarily follow visual guidance cues while blurring out the background [68]. Our results indicated that users were more aware of their environment when audio-tactile guidance was used as compared to well-

performing visual methods. Negative influences such as visual clutter and occlusions could be reduced and task performance was comparably reliable, albeit slower. These findings imply that contexts of use that require a higher level of safety combined with a de-cluttered visual FOV can benefit from audio-tactile guidance.

RQ_3 :What is the effect of multisensory guidance on situation awareness during search under sensory constraints?

The last research question, RQ_3, addresses the use of multisensory guidance cues to aid SA under sensory-constrained AR conditions. Especially in dense information spaces, it is difficult to become aware of new, potentially important information that appears inside or outside the FOV. Furthermore, sensory constrains such as sensory noise, background features, or limited FOV can affect perception, causing the user to miss newly emerging information. This research question asks how multisensory cue combinations can be used to inform users about moving out-of-view objects. For this purpose, we developed novel, multisensory proximity and transition techniques consisting of bimodal feedback in visual, auditory, or tactile form. We then evaluated their capability to enhance SA by guiding attention to out-of-view information using that feedback under the influence of further sensory constraints.

Contribution 3

Contribution 3 addresses the use of multisensory guidance to improve SA for moving out-of-view objects under sensory constraints. This topic is already partially addressed in Chapter 5, which presents our initial findings that multisensory guidance can help to increase SA alongside a guided search task. Chapter 6 expands on this issue to answer RQ_3 in detail. The key points of Contribution 3 can be summarized as follows:

- The provision of multisensory proximity and transition cues to inform users of emerging information in the scene in visual, auditory, and tactile manners.
- Improved SA in AR search when perception is affected by sensory constraints. Constraining factors include limited depth perception, disparity planes, dense scene textures, background features, sensory noise, a narrow FOV, and display properties.
- Findings of a high usefulness and user acceptance of bimodal proximity and transition combinations compared to unimodal modes.
- Modes with tactile transition cues were found to be particularly helpful under conditions of increased sensory noise. In particular, the audio-tactile mode was

found to be the most effective under conditions with high influence of sensory constraints.

Chapter 6: Multisensory Proximity and Transition Cues for Improving Target Awareness in Narrow Field of View Augmented Reality Displays. In Chapter 6, we address the problem that it can be difficult to detect newly emerging information that appears inside or outside a potentially narrow FOV. Especially in dense information spaces, the problem is further aggravated by typical visual conflicts in the visible screen space. For this purpose, we developed and evaluated multisensory cue combinations for narrow FOV devices to inform the user regarding moving out-of-view objects. To do so, we distinguished between proximity and transition cues in a visual, auditory, or tactile manner. Proximity cues were used to enhance spatial awareness of approaching out-of-view objects, while transition cues informed the user that the information had just entered the FOV. These cues were finally combined into a seamless feedback stream called a "mode", starting with a proximity cue when the augmented information was approaching out-of-view and triggering a short transition cue as soon the augmentation had passed the border of the FOV. Two user studies were conducted to examine the multisensory cue combination in terms of preferences and effectiveness in task performance and SA under different conditions of sensory noise and background features. In study 1, users (n = 10) were asked to state their personal preferences for a variety of six different modes via forced-choice decisions. Users were also asked to evaluate the usefulness of the two approaches: proximity cues to draw attention to an augmentation outside the FOV and transition cues to inform when the information is transitioning into the FOV. It was shown that, in general, bimodal cues that combined cues of different modalities received higher preference scores than unimodal cues that combined cues of the same modality. Moreover, when one perceptual channel has been blocked by noise, the capacities of another sensory channel remain potentially free. In addition, modes that incorporated tactile transition cues were preferred under conditions of higher noise. In study 2 (n = 14), the three modes with the highest preference scores from study 1 – namely Audio-Tactile, Visual-Tactile, and Visual-Audio – were evaluated in a divided attention task. In this task, users were asked to react to out-of-view objects that entered the FOV while performing a concurrent visual task in the central visual field. Results showed faster reactions under low and high noise conditions with the Visual-Tactile and Audio-Tactile modes compared to the Visual-Audio mode. Furthermore, we found an increase in reaction times when the noise level was

increased for the Visual-Audio and Visual-Tactile modes but not for the Audio-Tactile mode. The high performance in the secondary task for all tested modes demonstrates that the use of proximity and transition cues left sufficient cognitive capacities free to perform concurrent tasks. Furthermore, the stronger impairment of visual compared to auditory cues during increased noise conditions suggests that audio noise was manageable with a good design of auditory cues. We also showed a high usefulness of tactile transition cues in environments with increased noise levels, which emphasizes the noticeable yet unintrusive character of tactile feedback. Overall, preference and performance results showed that users could effectively use proximity and transition cues to raise their awareness off incoming out-of-view targets.

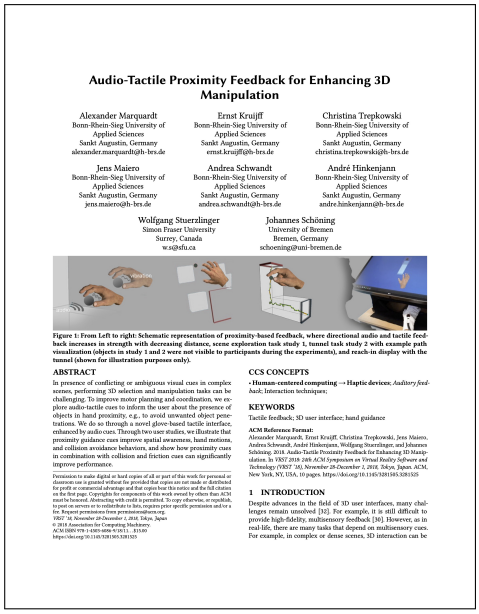
1.4 Structure of the Thesis

This work comprises five publications, whose contributions are described in Chapters 2–6. The general structure of the dissertation and the relationships of the individual chapters to each other are shown in Fig. 1.8.

Chapter 2 provides insights into the use of novel audio-tactile proximity feedback to enhance spatial awareness for 3D interactions in visually complex scenes. Chapter 3 presents tactile pose and motion guidance cues, which are beneficial for motor planning and coordination in potentially dense 3D interaction scenarios. The findings and methods from Chapters 2 and 3 informed the work in the following chapter. Chapter 4 provides insights into the provision of head-based, non-visual guidance cues under the influence of sensory constraints in AR, such as a narrow FOV and dense scene structures. Thus, Chapter 4 represents the methodological framework of this work, which is further developed in Chapter 5. Here, the search performance of the final multisensory guidance approach from Chapter 4 is examined by comparing it to a modern visual guidance method under further sensory constraints. In addition, initial studies of SA performance using multisensory guidance were conducted. Insights from the results described in Chapters 4 and 5 informed the research presented in Chapter 6, which addresses how multisensory guidance can improve SA of out-of-view objects under sensory constraints. In Chapter 7, the findings from the previous chapters are discussed in relation to the proposed research questions. In this way, we highlight the potentials and limitations of multisensory guidance under sensory constraints to support search guidance in AR. Finally, we summarize our results and provide a brief outlook for future work.

A note on the writing style. Although this thesis has been written by a single author, the first-person plural *we* is consistently used when referring to research activities. This has been done to avoid passive phrases, which are more difficult to read. Furthermore, the work described in the upcoming chapters was conducted in collaboration with others. Finally, *we* is used to engage the author and the reader in the transfer of knowledge induced by this scientific work. Nevertheless, this thesis contains only original research that was planned and conducted by the author.

2 Audio-Tactile Proximity Feedback



Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation

Alexander Marquardt
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
alexander.marquardt@h-brs.de

Ernst Kruijff
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
ernst.kruijff@h-brs.de

Christina Trepkowski
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
christina.trepkowski@h-brs.de

Jens Maiero
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
jens.maiero@h-brs.de

Andrea Schwandt
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
andrea.schwandt@h-brs.de

André Hinkenjann
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
andre.hinkenjann@h-brs.de

Wolfgang Stuerzlinger
Simon Fraser University
Surrey, Canada
w.s@sfu.ca

Johannes Schöning
University of Bremen
Bremen, Germany
schoening@uni-bremen.de

Figure 1: From Left to right: Schematic representation of proximity-based feedback, where directional audio and tactile feedback increases in strength with decreasing distance, scene exploration task study 1, tunnel task study 2 with example path visualization (objects in study 1 and 2 were not visible to participants during the experiments), and reach-in display with the tunnel (shown for illustration purposes only).

ABSTRACT

In presence of conflicting or ambiguous visual cues in complex scenes, performing 3D selection and manipulation tasks can be challenging. To improve motor planning and coordination, we explore audio-tactile cues to inform the user about the presence of objects in hand proximity, e.g., to avoid unwanted object penetrations. We do so through a novel glove-based tactile interface, enhanced by audio cues. Through two user studies, we illustrate that proximity guidance cues improve spatial awareness, hand motions, and collision avoidance behaviors, and show how proximity cues in combination with collision and friction cues can significantly improve performance.

VRST '18, November 28-December 1, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6086-9/18/11…$15.00
https://doi.org/10.1145/3281505.3281525

CCS CONCEPTS

• **Human-centered computing** → **Haptic devices**; *Auditory feedback*; Interaction techniques;

KEYWORDS

Tactile feedback, 3D user interface, hand guidance

ACM Reference Format:

Alexander Marquardt, Ernst Kruijff, Christina Trepkowski, Jens Maiero, Andrea Schwandt, André Hinkenjann, Wolfgang Stuerzlinger, and Johannes Schöning. 2018. Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation. In VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), November 28-December 1, 2018, Tokyo, Japan. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3281505.3281525

1 INTRODUCTION

Despite advances in the field of 3D user interfaces, many challenges remain unsolved [31]. For example, it is still difficult to provide high-fidelity, multisensory feedback [30]. However, as in real-life, there are many tasks that depend on multisensory cues. For example, in complex or dense scenes, 3D interaction can be

This chapter represents the first stage of designing and evaluating multisensory cues to be used for attention and guidance cueing. This work was mainly developed and evaluated as part of my master's thesis entitled "An Audio-Tactile Glove for Improved Object Interaction in 3D Scenes". The initial methods and results of this work were further explored and analyzed in terms of sensory constraints for the purposes of this thesis. Specifically, we closely examined task performance under the sensory constraints. Influencing factors explored in this work concerned depth perception, sensory thresholds, and scene structure.

In this work, we present a new audio-tactile approach to provide proximity cues to inform about objects in the close vicinity of the hand. This approach is particularly useful in visually complex environments, such as assembly or maintenance scenarios, in which components may occlude each other and thus remain hidden from the user. We have demonstrated the usefulness of audio-tactile proximity feedback for spatial exploration and manipulation through a novel glove-based tactile interface that is enhanced by audio cues. By using this method, we can improve hand motor planning and action coordination during 3D interaction. Through two user studies (n = 12), we found that proximity guidance cues could improve spatial awareness, hand motions, and collision avoidance behaviors. Finally, proximity cues in combination with collision and friction cues could significantly improve task performance.

The material in this chapter originally appeared in: Marquardt, A., Kruijff, E., Trepkowski, C., Maiero, J., Schwandt, A., Hinkenjann, A., Stürzlinger, W., & Schöning, J. (2018). Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation. *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 1–10. DOI: 10.1145/3281505.3281525

2.1 Introduction

Despite advances in the field of 3D user interfaces, many challenges remain unsolved [233]. For example, it is still difficult to provide high-fidelity, multisensory feedback [223]. However, as in real-life, there are many tasks that depend on multisensory cues. For example, in complex or dense scenes, 3D interaction can be difficult: hand motions are hard to plan and control in the presence of ambiguous or conflicting visual cues, which can lead to depth interpretation issues in current unimodal 3D user interfaces. This, in turn, can limit task performance [233]. Here, we focus on 3D manipulation tasks in complex scenes. Consider a virtual reality training assembly procedure [42], in which a tool is selected and moved through a confined space by hand, and then using the tool to turn a screw. Here, multiple visual and somatosensory (haptic) cues need to be integrated to perform the task. A typical problem during manipulation in unimodal interfaces in such scenarios is hand-object penetration, where the hand passes unintendedly through an object. Such object penetrations can occur frequently, especially when users cannot accurately judge the spatial configuration of the scene around the hand, making movement planning and correction difficult. However, similar to real-world scenarios, multisensory cues can disambiguate conflicting visual cues, optimizing 3D interaction performance [412]. Cues can be used proactively and adaptively, affording flexible behavior during task performance [412].

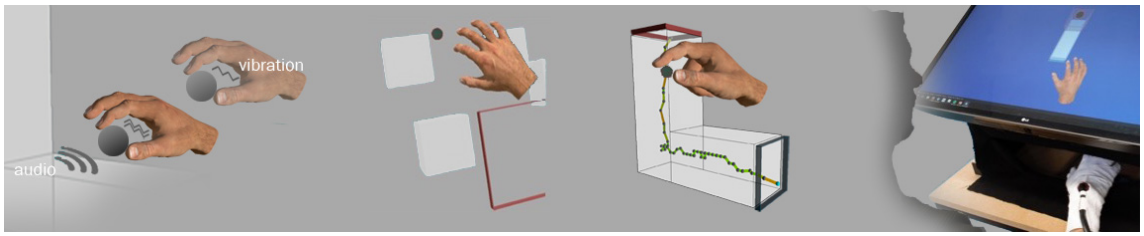


Fig. 2.1: From Left to right: Schematic representation of proximity-based feedback, where directional audio and tactile feedback increases in strength with decreasing distance, scene exploration task Study 1, tunnel task Study 2 with example path visualization (objects in Study 1 and 2 were not visible to participants during the experiments), and reach-in display with the tunnel (shown for illustration purposes only).

2.1.1 Motor Planning and Coordination

Planning and coordination of selection and manipulation tasks is generally performed along a task chain with key control points. These control points typically relate to contact-driven

biomechanical actions [190]. As such, they contain touch cues that relate to events about touching objects to select them (selection) or move along a trajectory (manipulation). This may contain various hand motion and pose actions that are performed within the scene context, e.g., for steering the hand during manipulation tasks. There should be sufficient indication as to where the hands touches objects upon impact (collision contact points) or slides along them (friction contact points), while other indications, such as object shape or texture, can also be beneficial [193].

Multisensory stimuli enable learning of sensorimotor correlations that guide future actions, e.g., via corrective action patterns to avoid touching (or penetrating) an object [190]. In real-life, to steer hand motions and poses, we depend typically on visual and physical constraints. E.g., lightly touching a surrounding object might trigger a corrective motion. However, manipulation tasks are also performed independent of touch cues, namely through self-generated proprioceptive cues [288]. Such cues may have been acquired through motor learning [369]. Although not the main focus of this work, motor learning can be an important aspect for skill transfer between a 3D training application and the real-world [71, 219], thereby potentially also "internalizing" proprioception-driven actions for later recall.

2.1.2 Research Questions

Our novel guidance approach, which is described in more detail in Section 2.3 "Approach", is based on audio-tactile proximity feedback to communicate the direction and distance of objects surrounding the user's hand. Feedback is used to plan and coordinate hand motion in 3D scenes. Our research is driven by the following research questions (RQs) that assess how we can guide the hand motion before and during 3D manipulation tasks using such feedback.

RQ1: *Do scene-driven proximity cues improve spatial awareness while exploring the scene?*

RQ2: *Can hand-driven proximity cues avoid unwanted object penetration or even touching proximate objects during manipulation tasks?*

In this paper, we measure the effect of proximity cues in combination with other haptic cue types (in particular collision and friction). Towards this goal, Study 1 (scene exploration) explores the general usefulness of proximity cues for spatial awareness and briefly looks at selection, while Study 2 looks specifically at the effect of proximity on 3D manipulation

tasks. In our studies, we specifically look at touch and motion aspects, while leaving support for pose optimization as future work. As a first step, we focus on feedback independently of visual cues, to avoid confounds or constraints imposed by such cues.

2.1.3 Contributions

Our research extends previous work by Ariza et al. [10] that looked into low resolution and non-directional proximity feedback for 3D selection purposes. We provide new insights into this area of research by looking at higher-resolution and directional cues for manipulation (instead of selection) tasks. Our studies illustrate the following benefits of our introduced system:

- In the scene exploration task, we show that providing proximity feedback aids spatial awareness through a higher number of tactors (18 vs. 6), which improves both proximity feedback (20.6%) and contact point perception (30.6%). While the latter is not unexpected, the results indicate the usefulness of a higher-resolution tactile feedback device.
- We illustrate how the addition of either audio or tactile proximity cues can reduce the number of object collisions up to 30.3% and errors (object pass-throughs) up to 56.4%.
- Finally, while friction cues do not show a significant effect on measured performance, subjective performance ratings increase substantially, as users thought that with friction (touch) they could perform faster (18.8%), more precisely (21.4%), and react quicker to adjust hand motion (20.7%).

2.2 Related work

In this section, we outline the main areas of related work. Haptic feedback has been explored for long, though is still limited by the need for good cue integration and control [223, 384], cross-modal effects [313], and limitations in actuation range [163]. The majority of force feedback devices are grounded (tethered). Such devices are often placed on a table and generally make use of an actuated pen that is grasped by the fingertips, e.g., [416]. Only few glove or exoskeleton interfaces exist that enable natural movement, while still providing haptic feedback, such as grasping forces, e.g., [47]. In contrast, tactile methods remove the physical restrictions of the aforementioned actuation mechanisms, and thus afford more flexibility, by substituting force-information in tactile cues, not

only for 3D selection and manipulation tasks [195, 225], but also for other tasks like navigation [222]. In 3D applications, recent research looked at smaller, handheld (e.g., [32]) or glove-based (e.g., [131, 372]) tactile actuators [9, 69]. Instead of stimulating only the fingertips and inner palm using a limited number of tactors, researchers have also looked into higher-density grids of vibrotactors to stimulate different regions of the hand [136, 269, 344], but these approaches are currently limited to localized areas.

Some researchers have explored proximity feedback with a haptic mouse [173], using vests for directional cues [242], to trigger actions [29], and for collision avoidance using audio feedback [2]. Most relevant to our tactile proximity feedback is a system called SpiderSense [272], which uses tactors distributed over the body to support navigation for the visually impaired. This kind of feedback is similar to a distance-to-obstacle feedback approach [162] and a glove-based approach for wheelchair operation [410]. Furthermore, tactile guidance towards a specific target [310] or motion and pose [266] has shown promise. Yet, both the usage context and approaches differ fundamentally from our tactile guidance approach, which aims to increase spatial awareness to better support manipulation of objects in 3D interaction scenarios. Finally, Ariza et al. studied non-directional feedback for selection tasks, showing that different types of feedback affect the ballistic and correction phases of selection movements, and significantly influence user performance [10].

Challenges

Providing multisensory cues – in particular haptics – to complement visual-only feedback has benefits for 3D manipulation tasks. However, while haptic cues aid in guiding hand motion and poses, their inclusion in 3D user interfaces is challenging. Traditional grounded haptic interfaces (force feedback devices) provide cues that support the user in performing fine manipulation tasks, for example by guiding the hand by constraining its' motion. As such, haptics potentially ameliorate any negative effect of visual ambiguities [53] and has been shown to improve selection tasks [70]. However, haptic devices often have limitations, such as operation range, the kind of contact information being provided, and issues related to the type of the used feedback metaphor. For example, popular actuated pen devices, such as the (Geomagic) Phantom, do not necessarily comply to how users perform actions in the real world, as they support only a pen grip instead of full-hand interaction. Such interfaces do not provide contact point feedback across the full hand, which limits the feedback that users can use to plan and coordinate selection and manipulation tasks: users

will be unaware where the hand touches another object, even though this information may be required to steer hand motion and poses. While full-hand interfaces exist, they are often expensive, have mostly a limited operation range, and can be cumbersome to use.

Tactile interfaces are an interesting alternative to traditional grounded haptic (force feedback) devices, as they provide portable solutions with good resolution and operation range [424]. However, designing effective tactile cues is challenging, as haptic (force) stimuli cannot be fully replaced by tactile cues without loss of sensory information [195]. Furthermore, simulating contact has its limitations, as untethered systems cannot restrict physical motion. As a result – similar to visual-only feedback conditions – users may still pass through virtual objects unintentionally, as users often cannot react quickly (enough) to tactile collision cues [77]. During selection tasks, and before colliding (selection) with an object, the hand typically goes through a fast (ballistic) motion phase, followed by fine, corrective motor actions [248]. Similarly, once the hand touches an object in the scene during a manipulation task, a corrective movement may be performed, e.g., to steer away from this object. However, as movement is typically not perfect, the users' hand will often move into or through the object even though a tactile collision cue is perceived, especially when a corrective movement is initiated too late. The presence of any depth perception issues or other visual ambiguities typically make this situation only worse. During selection tasks, this may for example lead to overshooting [10]. Furthermore, especially for thin objects, users may move (repeatedly) through the object during manipulation, as such objects trigger only short bursts of collision feedback.

2.3 Approach

We aim to overcome limitations associated with the untethered nature of many tactile devices – in particular the inability to constrain human motion – by guiding the hand through proximity feedback. This kind of feedback can improve spatial awareness about objects surrounding the hand to guide the motion, which helps to avoid contact before it happens. While proximity cues have been introduced to optimize pointing behavior during 3D selection tasks [10], we expect such cues are also beneficial for manipulation tasks that are driven by steering behaviors. Yet, we are unaware of work that has explored proximity cues for manipulation tasks. Our proximity feedback provides continuous, spatio-temporal audio-tactile feedback about objects surrounding the hand, independent from contact events. This feedback is coupled to object collision and friction cues that

relate to the biomechanical control (contact) points, to enrich task chain-driven feedback. In our approach tactile feedback only provides indications about distance to other objects, while directional information is provided through audio. We made this choice based on the results of pilot studies, described in Subsection 2.4.1 “Pilot Studies”. Audio extends the tactile feedback by providing sound upon impact (collision), directional and distance cues to objects around the hand (proximity), and texture cues during friction. Coupling audio to tactile cues can be beneficial as there is evidence for good multisensory integration of both, especially with regards to temporal aspects [306]. However, while audio and vibration have been shown to improve performance in 2D interfaces [77], there is surprisingly little evidence for performance improvements for 3D selection and manipulation tasks.

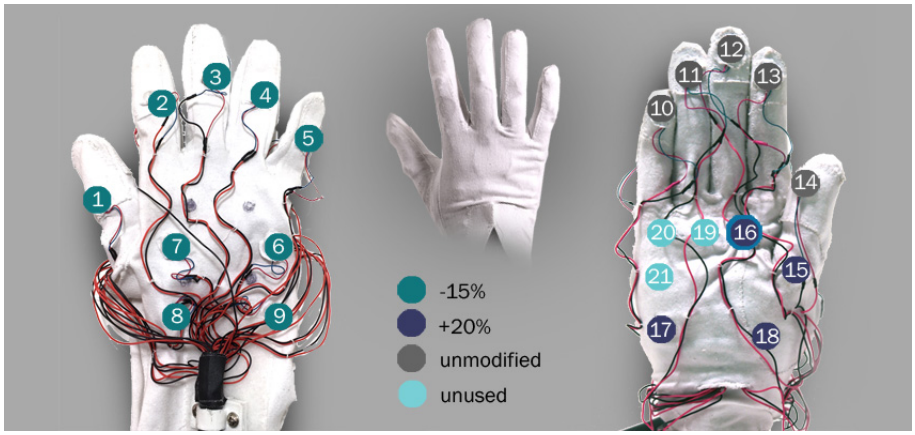


Fig. 2.2: Tactor IDs and balancing of our tactile glove (inner glove only), glove with protective cover.

Our feedback mechanism differs from previous work on audio-tactile proximity-based selection assistance [10] in multiple ways. There the authors used only non-directional cues and focused on selection, not manipulation. Also, non-directional cues can only encode distance to a single object, which is insufficient in scenes where users can collide with multiple surfaces/objects around the hand. In contrast, our approach uses a glove-based interface developed in-house that contains a higher-density grid of vibrotactors across both sides of the hand and as such provides contact information across the full hand. Moreover, we use directional cues to elicit directional information about objects in hand proximity.

Tactile Glove

We developed a vibrotactile glove (see Fig. 2.2) whose operation range supports full arm motions. Hand pose and motion is tracked through optical methods, in our case a Leap

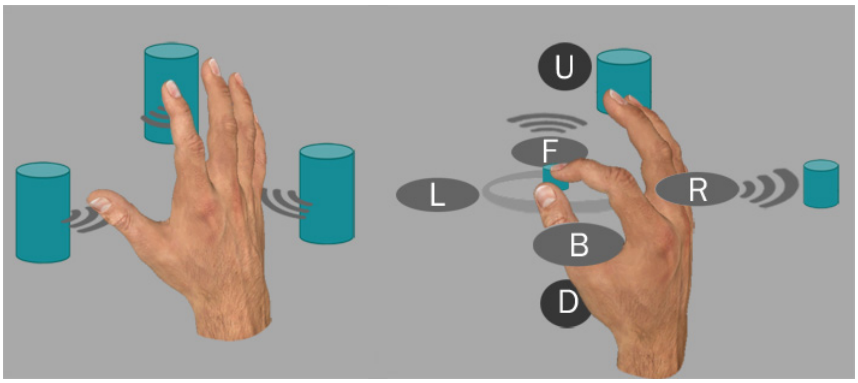


Fig. 2.3: Outside-in proximity cues, where audio-feedback is spatialized in the scene (Left). Inside-out proximity cues, where sound localization is tied to the hand (Right).

Motion. The glove has also been used for other purposes, namely hand motion and pose guidance. In [266] we illustrated how tactile patterns can guide the user, by triggering hand pose and motion changes, for example to grasp (select) and manipulate (move) an object.

The glove is made of stretchable, comfortable-to-wear cotton. In the glove, tactors are placed at the fingertips (5 tactors), inner hand palm (7), middle phalanges (5), and the back of the hand (4), for a total of 21 tactors (Fig. 2.2). An outer cotton cover fits exactly over the inner glove to protect the cables and lightly press the tactors against the skin. We use 8-mm Precision Microdrive encapsulated coin vibration motors (model 308-100). In our pilot studies, we identified that tactors #19-21 lie too close to the tactor used for proximity feedback, #16. Especially during grasping, this leads to misinterpretation of cues, as tactors move closely together. Thus, we used only tactors #1-18 in our studies, to avoid confusion between collision and proximity feedback. With 18 tactors, we simulate many contact points that are associated with grasping objects (palm, fingertips) while also supporting collision feedback at the back of the hand. This is a novel feature, as back-of-the-hand feedback is generally not supported in tactile interfaces. Even though we do not cover the full hand surface with tactors, we still cover most areas and can benefit from phantom effects by interpolating between tactors, similar to [183]. The cable ends and master cable are attached at the back of the wrist through a 3D printed plate embedded in the fabric. All tactors are driven by Arduino boards. To overcome limitations in motor response caused by inertia (up to ~75 ms), we use pulse overdrive [269], which reduces latency by about 25 ms.

2.3.1 System and Implementation

The system was implemented in Unity3D V5.6, using NVidia PhysX 3.3 for collision detection. Hand tracking was performed with a Leap Motion, through the Orion SDK. We used the Uniduino plugin to control four Arduino Boards to trigger the tactors. The system ran on a graphics workstation (Core i7, 16GB RAM, NVidia 1080GTX, Windows 10) to guarantee fluid performance. During the first study, interaction was performed below a reach-in display, a 20-degree angled 32" display (Fig. 2.1, right). Replicating standard virtual hand metaphors [233], we only showed the hand, not the wrist or arm, using a realistic 20,000 polygon hand model from the Leap Motion SDK. The index finger and thumb were used to grab (pinch) an object. Once an object is pinched, the user receives a short tactile burst at the thumb and index fingertip. While the user holds the object, no further tactile cues are provided at these locations, to avoid habituation as well as confusion between pinch and scene collision cues.

2.3.1.1 Proximity Feedback Modes. We explored two modes that combine tactile and audio feedback for proximity feedback, see Fig. 2.3. With outside-in feedback, each object in the scene emits signals, i.e., feedback is spatially tied to the objects in the scene. In contrast, with inside-out feedback, feedback is provided relative to the grasped object in the hand – directions are divided into zones. The hand “sends” signals out into the scene, and “receives” spatial feedback about which zones around the hand contain objects, similar to radar signals. Both modes are implemented analogous to car park assistant technologies to indicate where (direction) and how close (distance) surrounding objects are. Tactile cues are represented by vibration patterns, starting with slow and light vibrations and, as the distance to neighboring objects shortens, ending with stronger and shorter-cycle vibrations.

Vibrotactile proximity cues are provided for the closest available object collider as soon the users hand is close enough. As discussed above, we use the pulse overdrive method to quickly activate the corresponding tactor. To stably drive the motor, we then reduce the voltage via pulse width modulation (PWM) to the lowest possible amount, about 1.4V (a duty cycle of 28%). As the user is getting closer to the collider, the duty cycle is adjusted inversely proportional to the collider distance, creating the maximum vibration intensity with a duty cycle of 100% right at the object. We use a single tactor in the palm (tactor #16 in Fig. 2.2) to provide vibrotactile proximity cues and use audio to communicate the direction and distance to surrounding objects. This design decision was based on pilot

studies that showed that full-hand proximity cues are difficult to separate from collision cues. Furthermore, we introduce a deliberate redundancy between tactile and auditory distance cues, as we aim to strengthen the amount of "warning" before potential object penetrations. To provide audio cues, we used the Audio Spatializer SDK of the Unity game engine. This allows to regulate the gains of the left and right ear contributions based on the distance and angle between the AudioListener and the AudioSource, to give simple directional cues.

For *outside-in* proximity feedback, each object contains a spatially localized audio source: hence, users can hear the location of the objects over the used headphones. The audio "objects" are characterized not only by their location relative to the hand, but also by volume and pitch to provide 3D spatial cues. The adjustment of volume depends on the relative distance to the hand with a linear roll-off within a specified radius. As long the hand is within the roll-off threshold, the sound starts at neutral pitch level and gets higher the closer the hand gets to an object. As it is scene-driven, we assumed this model would be beneficial for general spatial exploration tasks: the feedback provides a general indication about objects in vicinity of the hand, instead of targeting more precise cues related to a single (grasped) object.

To support *inside-out* proximity feedback, we located six audio sources around the hand that define unique directions along the coordinate system axes. If an obstacle is detected at a certain direction, the corresponding proximity sound is played with the same volume and pitch characteristics as in the selection phase. Different abstract ("humming") sounds are used for up/down proximity compared to forward/backward/left/right proximity, in order to make the cues more distinguishable. This method is similar to parking aids in cars. Motivated by previous work [351], the pitch of a sound indicates the approximate position in the vertical direction: higher pitched sounds are perceived as originating above lower pitched sounds. As this model provides highly granular proximity cues in relation to the hand (and grasped object), we assumed that it can be beneficial for manipulation tasks in which an object is moved through a scene.

2.3.1.2 Collision and Friction Feedback.

Once the user actually touches an object, we provide location-specific collision cues, based on a mapping between contact point and an adjacency list of the tactors. All motors are given an individual weighting factor (see Fig. 2.2) which were fine-tuned through a pilot study reflecting on the local mechanoreception properties of the skin [191]. We calculate the distance of the collision point in relation to

the closest tactor on the glove. If a collision point is in between two tactors, this results in interpolation of vibration strength, similar to a mechanism described previous work [183]. Beyond the mechanoreception weighting factor, modulation of the tactor is then also affected by the distance to objects and hand velocity, resulting in a higher intensity when the collision occurs at a higher speed.

We use the Karnopp model, a stick-slip model that describes friction forces as the exceedance of the minimal tangential velocity relative to the object surface to provide friction cues [197]. Friction cues are triggered by the combination of object penetration and velocity and are represented through both vibration and audio feedback [225]. We use the PhysX API to determine penetration and its depth. Similar to proximity, friction cues consist of localized auditory and vibrotactile feedback, while tactile cues are directly dependent on the sound waveform that represents the material properties, similar to the method presented in [225]. For auditory friction feedback, we take the penetration depth and the velocity of the penetrating object into account. A material-conform friction sound is assigned to each object in the scene and is faded in or out depending on penetration depth. The intensity and pattern of the vibration feedback is based on the spectrum of the played friction sound, similar to [225].

2.4 User Studies

In our user studies we explored how different audio and tactile cues affect *touch* and *motion* by looking how proximity cues influence spatial awareness in a scene exploration task (RQ1, using the outside-in model) and precise object manipulation performance in a fine motor task (RQ2, with the inside-out model). All studies employed the setup described above. With consent of the users, demographic data was recorded at the start. For Study 1, we only analyzed subjective feedback, while for Study 2 we logged task time, object collisions, penetration depth and the number of tunnel exits in between start and end position (errors). After the study, participants rated their level of agreement with several statements related to concentration, cue usefulness, perceptual intensity, and spatial awareness on a 7-point Likert scale (7 = “fully agree”). It took between 45 and 75 minutes to complete the whole study.

2.4.1 Pilot Studies

We performed several pilot studies during the design and implementation process of our glove interface prior to the main ones. The first pilot aimed to verify our feedback approach, coupling proximity, collision and friction cues. Nine users (1 female, aged between 25 and 30 years) interacted with an early design of the glove. Users performed a key-lock object manipulation task, selecting a target object and moving it into another object. The objects were small and partly visually occluded. The pilot confirmed the utility of the proximity-driven approach, but identified limitations in tactile resolution and audio feedback. This informed the design of a higher-resolution glove. Based on an near-complete version of the glove, the second pilot fine-tuned feedback cues and probed study parameters for the main studies. Through multiple tests performed with four people we tuned the weighting factors of the tactors, with the results shown in Fig. 2.2. A third pilot with six users (one female, aged between 26 and 39) explored various design parameters of our main studies. This pilot included a tunnel task and a search task to find an opening and was used to make final adjustments to the glove feedback mechanisms, in particular the proximity based feedback approach in the reach-in display system (Fig. 2.1).

2.4.2 Study 1 - Scene Exploration

In this study, we explored how the number of contact points afforded by the glove and the enabling or disabling of proximity cues affects spatial awareness in relation to hand motion constraints, i.e. hand-scene constraints, during scene exploration.

For the task, we showed a start position and the position of an object to select, which defined the end position. We located several invisible objects (cubes) between the start (front) and end position (back), creating an environment through which the hand had to be maneuvered without touching or passing through obstacles (see Fig. 2.1, second image from Left). Before selecting the object, users had to explore the scene while receiving collision, proximity, and friction cues, which enabled them to understand the scene structure. As the Cybertouch is currently a quasi-standard in vibrotactile gloves, the glove was either used with full resolution for collision (18 tactors) or simulating the Cybertouch II (6 tactors, one at each finger tip, one in the palm, ID 16, Fig. 2.2, Right). In both conditions proximity cues were only felt at the tactor at the palm of the hand. In our simulated low-resolution Cybertouch condition, collision cues were remapped to match the limited number of tactors. We compared this condition with our high-resolution tactor

configuration to assess if increasing the number of tactors enables better performance. In other words, we investigated if quasi full-hand feedback instead of mainly finger-tip and palm feedback provides more benefits compared to somewhat higher technical complexity of additional tactors.

The study was performed within-subjects and employed a 2 (low or high resolution feedback) x 2 (proximity feedback on / off) x 2 (different scenes) design, totaling 8 trials. All scenes had to be explored for about one minute each and feedback was based on the scene-driven outside-in proximity model. Participants were asked to evaluate if they could more easily judge where their hand would fit between objects depending on proximity cues (off vs. on) and the resolution of the feedback (high vs. low).

2.4.3 Study 2 - Object Manipulation

In this study, we looked into the effect of proximity cues on user performance during a manipulation task that involved steering the hand (with a grasped object) through a scene. We used a tunnel scene analogy as it is quite common to assess steering tasks using paths with corners [456], while it also shows resemblance to assembly tasks where a grasped object needs to be moved through space. Users were asked to move a small object (2 cm size) through an *invisible* tunnel (from top front to lower back). Participants were instructed to move as fast as possible, while reducing collisions and penetrations with or pass-throughs of tunnel walls. In this study we always used all 18 tactors - 17 for contact information and one for proximity. The focus of our research was on the usefulness and performance of the different feedback conditions, i.e., collision, proximity, and feedback cues, during fine object manipulation. We aimed to isolate the effect of each feedback method through three blocks and also looked into potential learning effects. The tunnel contained two straight segments connected by a 90 degree corner (main axis). The "bend" was varied by changing the angle of the two connected tunnel segments (10 degrees variations from the main axis - tunnels with more angled segments were expected to be more difficult). Tunnels had a wall thickness of 1.5 cm, which was used to calculate penetration depth and pass-throughs. We only showed the start and end positions of the tunnel and the object to be selected, while the rest of the tunnel remained invisible. This forces users to focus on the tactile cues in isolation and has the additional benefit that it avoids any potential disadvantages of any given visualization method (such as depth ambiguities associated with transparency). When users exited the tunnel by more than 1 cm between start and end, users had to restart the trial. Users wore the glove (Fig. 2.2),

while interacting underneath the reach-in display. To avoid the potential confound of external auditory cues during the user studies and to remove the effect of potential audio disturbances, we used Bose 25 headphones with active noise cancellation.

This study used the object-driven inside-out proximity model. It deployed a within-subject design and consisted of three blocks. Block 1 (*collision only*) included 9 trials, defined by the nine tunnel variants (3 variants of segment one x 3 variants of segment two). Subjects performed the task solely with collision feedback. This block implicitly also familiarized participants with the procedure. Block 2 (*collision and proximity*) employed a 9 (tunnel variants) x 2 (with and without audio proximity cues) x 2 (with and without vibration proximity cues) factorial design, totaling 36 trials. Collision feedback was always enabled. Block 3 (*collision, proximity, and friction*) employed a 9 (tunnel variants) x 2 (with or without friction) factorial design, totaling 18 trials, where collision and audio-tactile proximity cues were always enabled. We split the experiment into blocks, as a straight four-factor design is statistically inadvisable. Instead, our blocks build on each other, which enables the comparison of trials with and without each cue. Between blocks participants were introduced to the next feedback condition in a training scene. As friction cues alone do not help to avoid collisions, they were only presented in combination with proximity cues in the third block. It took around 35 minutes to finish this study.

Table 2.1: Mean ratings (standard deviations in brackets) during scene exploration, for hand-scene constraints with proximity cues ("does the hand fit through") and contact points.

	Feedback Resolution		
Perceived constraints	low	high	Improvement
– off	4.08 (0.90)	4.92 (0.90)	+20.6% **
– on	5.33 (0.88)	6.25 (0.62)	+17.3% **
Improvement	+30.6% ***	+27.0% ***	
Perceived contact point			
– overall hand	4.08 (0.90)	5.33 (1.37)	+30.6% *
– fingers	4.50 (1.24)	5.67 (1.37)	+26.0% **
– back of hand	3.42 (1.08)	5.0 (1.04)	+46.2% **
– palm	4.33 (1.37)	4.92 (1.24)	+13.6%, n.s.

* $p < .05$, ** $p < .01$, *** $p < .001$

2.5 Results

The sample for Study 1 and 2 was composed of 12 right-handed persons (2 females, mean age 31.7, SD 11.11, with a range of 23–58 years). Five wore glasses or contact lenses and 7 had normal vision. The majority played video games regularly, 6 persons daily (50%), 5 weekly (41.7%) and one only monthly (8.3%). All participants volunteered and entered into a drawing (with a shopping voucher).

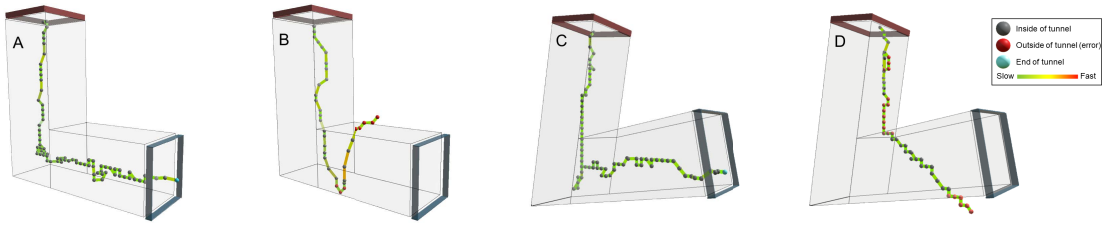


Fig. 2.4: Example paths from Study 2. The first tunnel is simple, with a 90° bend (A & B). The second variant is moderately difficult, with a 70° turn (C & D). Tunnel walls were not visible to participants in the studies.

2.5.1 Study 1

In this part of the study, participants explored a scene to gain spatial awareness of the scene structure. As this task was not performance driven, we only report on subjective ratings from the questionnaire, analyzed using paired t-tests.

Table 2.1 shows mean ratings and standard deviations as well as statistically significant differences. The mean level of agreement was significantly higher for high resolution than for low resolution feedback, both with proximity cues and without. Comparing the ratings for the same statement between proximity cues (off vs. on), the level of agreement was higher with proximity cues than without in both the high and low resolution feedback conditions. The point of collision could be better understood with high than with low resolution feedback on the overall hand, fingers, and the back of the hand, but not in the palm.

2.5.2 Study 2

For the analysis of blocks 1 to 3, we used in each case a repeated-measures ANOVA with the Greenhouse-Geisser method for correcting violations of sphericity, if necessary. Dependent variables were time to finish a trial successfully, collisions, penetration depth and errors in each block. Independent variables differed between blocks, in the first block

we examined the effect of tunnel variants, in the second block the effect of tunnel variants, proximity audio and vibration cues and in the third block the additional use of friction. The effect of the factor cue on different questionnaire ratings for block 2 was examined using a one-way repeated measures ANOVA. Post-hoc comparisons were SIDAK corrected. For block 3, paired t-tests were used to compare questionnaire ratings for trials with and without friction cues. All tests used an alpha of .05. Below, we only report on the main results.

Time to finish a trial increased between blocks (31.06 s, SD=23.4 for block 1, 36.65 s, SD=27.52 for block 2, 40.15 s, SD=28.64 for block 3). In block 1 (collision cues only) there was no effect of the tunnel variants in terms of collisions, penetration depth or errors, except that there was an effect on time, F(8,88) = 2.16, p = .038, η^2 = .16. As expected, tunnels with angled segments took longer.

In block 2 we analyzed collision and proximity cues. The time required to pass tunnels was not affected by the tunnel variant and was also not influenced by cues. Yet, the tunnel variants significantly influenced the number of collisions, F(8,88) = 2.64, p = .012, η^2 = .19. Most tunnels produced a limited range of collisions, 3.51 (SD = 3.25) to 5.71 (SD = 3.57), except for the most complex one that produced 7.50, (SD = 6.57). For proximity cues we observed that most collisions occurred when both cues were off and fewest when only audio cues were on (Table 2.2 shows mean values and significances).

Table 2.2: Study 2, block 2: Mean performance values depending on proximity cues and % change against baseline. *Prox* stands for proximity, *A* for audio, *V* for vibration

	Collisions	Penetration depth	Errors
Collision (baseline)	6.17	0.145	1.56
$Prox-A$ only	4.3 * (-30.3%)	0.125 ** (-13.8%)	0.68 ** (-56.4%)
$Prox-V$ only	4.94 * (-19.9%)	0.142 n.s. (-2.1%)	1.13 n.s. (-27.6%)
$Prox-A+V$	4.56 n.s. (-24.6%)	0.113 ** (-22.1%)	0.9 n.s. (-42.3%)

n.s. not significant, * $p < .05$, ** $p < .01$

Audio and vibration proximity cues showed no main effect on the number of collisions, but there was a tendency to an interaction effect of proximity cues, F(1,11) = 4.76, p = .052, η^2 = .30 (see Fig. 2.5). Post-hoc comparisons revealed that audio or vibration proximity cues alone significantly affected the number of collisions when the other proximity cue was

turned off ($p < .05$). Furthermore, mean penetration depth was significantly smaller with audio cues (Table 2.2, $F(1,11) = 14.57$, $p = .003$, $\eta^2 = .57$). Penetration depth was also influenced by the tunnel variant ($F(4.50,49.48) = 4.34$, $p = .003$, $\eta^2 = .28$) – again, the most complex one lead to the largest penetration depth ($M = 0.155$, $SD = 0.036$). Regarding errors there was a tendency to an interaction effect of audio and vibration proximity cues, $F(1,11) = 4.55$, $p = .056$, $\eta^2 = .29$ see Fig. 2.5. When vibration proximity cues were turned off, audio proximity cues significantly influenced the number of errors as less errors occurred with audio proximity cues than without (Table 2.1, $p = .035$). The presence of vibration cues did not significantly reduce the number of errors when audio was turned off ($p = .093$).

Block 3 focused on collision, proximity, and friction cues. There was a significant effect of tunnel variant on the number of collisions ($F(8,88) = 4.38$, $p < .001$, $\eta^2 = .29$), but no effect on time, mean penetration depth and errors. Again, the most complex tunnel stood out, with the most collisions ($M = 8.17$, $SD = 6.24$). Friction cues did not affect any of the dependent variables and there was also no interaction effect of tunnel variant and friction.

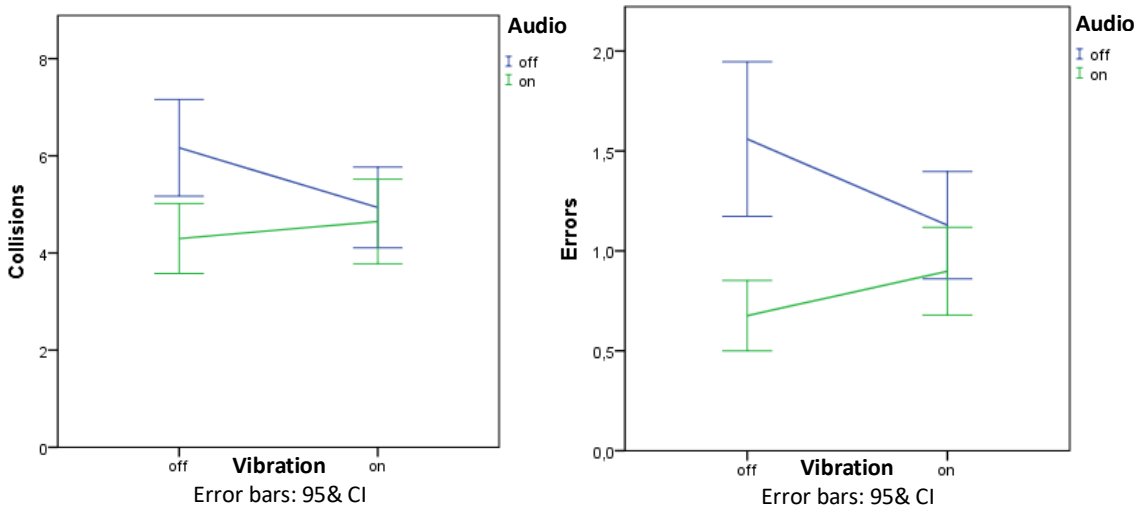


Fig. 2.5: The effect of audio and vibration proximity cues on collisions and errors.

2.5.3 Path Analysis

To better understand participant performance during the trials, we sampled the dataset by selecting best and worse trials from different tunnel conditions (easy and more difficult ones, as defined by the variable angle between both tunnel segments). Here, we present the most relevant examples of this process to exemplify path behavior. Fig. 2.4A & 2.4B show examples of an easy task (90° bend) in visual comparison to a more challenging

one (70° turn, Fig. 2.4C & 2.4D). With all activated proximity cues (collision, proximity, and friction cues, Fig. 2.4A & 2.4C) participants found it easier to stay within the tunnel, while this was harder when only collision cues were present (Fig. 2.4B & 2.4D). In the latter cases the path shows only a partial run until the first error occurred (which required a restart of the trial). Samples and measurements taken at various points along the path of the examples paths show that proximity cues can help the user to move the object closer along the ideal path for both easy task (M = 0.69, Fig. 2.4A) and difficult task (M = 0.71, Fig. 2.4C). In contrast, however, without proximity feedback, the distance to the ideal path increased drastically for the simple task (M = 0.86, Fig. 2.4B), as well as for the difficult task (M = 1.22, Fig. 2.4D). This resulted in a higher error rate, through participants (unintentionally) leaving the tunnel.

Manipulation behavior is different from selection. Selection is a pointing task that exhibits a ballistic, fast phase before a corrective, slower motion phase. In contrast, a manipulation is a steering task in which motion velocity is far more equalized [316, 456]. As such, manipulation performance – and difficulty – is affected by the steering law, instead of Fitts's law [316]. Like Fitts's law, steering difficulty is defined by path width and curvature, yet is linear instead of logarithmic. The absence of velocity difference due to ballistic and corrective motions hand motions can be clearly seen in our examples. While velocity varies from about 14.14 mm/s to 67.12 mm/s in the shown samples, fast movements are only performed rarely and not in patterns that conform to rapid aimed pointing movements. Of course, steering still exhibits corrective motions, as can be seen for example in Fig. 2.4B at the lower end of the path. What is also striking is the behavior of steering through corners: the path does not necessarily adhere to the shortest path (hence, cutting the corner), rather the ideal path is defined by staying clear of the corners [316], even though Fig. 2.4D shows this is not always successful. This is somewhat in contrast to behavior in 2D interfaces, as noted in [316], where corners tended to get cut. We assume that in our case, cutting was avoided as proximity cues encourages the user to stay away from surrounding objects and thus also corners.

2.5.4 Subjective Feedback

Questionnaire ratings indicated that all cues facilitated to perform the task faster and more precisely, aided understanding of the tunnel shape, and made movement adjustments easier (Table 2.3). However, there was no significant difference between cue ratings. Interestingly, participants thought they performed the task faster ($t(11) = -2.59$, $p = .025$), more precisely

Table 2.3: Mean level of agreement on 7-point Likert items and standard deviations for cue usefulness in Study 2, block 2 & 3. *Prox* stands for proximity, *A* for audio, *V* for vibration, *Fric* for friction.

	Performed faster	Performed more precisely	Understood the tunnel shape better	Reacted more quickly
Collision	4.92 (1.78)	5.17 (1.27)	4.83 (1.53)	5.58 (1.73)
$Prox-A$	5.58 (1.38)	5.58 (1.83)	5.75 (1.49)	5.92 (1.08)
$Prox-V$	5.33 (1.16)	5.33 (1.44)	5.08 (1.44)	5.17 (1.12)
$Prox-A+V$	5.42 (1.62)	5.67 (1.78)	5.75 (1.49)	5.75 (1.55)
$Prox-A+V$	4.42 (1.08)	4.67 (1.23)	4.92 (1.31)	4.83 (1.40)
$\ldots+Fric$	5.25 (1.22)	5.67 (1.23)	6.0 (1.35)	5.83 (1.27)
Improvement	+18.8% *	+21.4% *	+22% *	+20.7% *

* $p < .05$

(t(11) = -2.71, p = .02), understood the shape of the tunnel better (t(11) = -2.86, p = .015), and reacted more quickly to adjust the object movement in the scene (t(11) = -2.45, p = .032) while using friction. In the open comments it was also striking that half of the participants reported that it was easier to focus on a single proximity cue at any given time. Some users stated they experienced a limited form of information overload when both proximity cues were activated simultaneously, which distracted them. Finally, we also evaluated the overall usability, comfort, and fatigue in the questionnaire (see Table 2.4). Most ratings were positive to very positive, though tracking errors and cabling issues were noted. As the experiment took some time, we were particularly interested in user fatigue. Fortunately, participants rather disagreed that they got tired while wearing the glove interface.

Table 2.4: Mean level of agreement with comfort and usability statements on 7-point Likert items and standard deviations.

Statement	Mean Rating (SD)
Sitting comfort	5.33 (1.14)
Glove wearing comfort	6.42 (0.67)
No disruption through the cable	3.25 (1.71)
Match of virtual to real hand	5.25 (1.14)
Hand tracking problems	4.41 (1.78)
Ease of learning the system	5.5 (1.24)
Ease of using the system	5.58 (1.17)
Expected improvement through exercise	6 (0.74)
Getting tired wearing the glove interface	3.25 (1.49)

2.6 Discussion

In our studies, we investigated the effect of proximity cues for hand touch and motion associated with scene exploration and manipulation actions. Here, we discuss our main findings.

RQ1: *Do scene-driven proximity cues improve spatial awareness while exploring the scene?*

Overall, our scene exploration study provides positive indications about the usage of scene-driven outside-in proximity cues to enhance spatial awareness. It also indicates a positive effect of increasing the number of tactors, as both the awareness of hand-scene constraints and contact (touch) points across the hand improved. The performance improvements provide a positive indication for higher numbers of tactors in novel glove-based or other types of full-hand interfaces. With our high-density tactor design, the localization of contact points across the hand improved about 30% in comparison to a Cybertouch-like configuration. It is also interesting to contrast our results to the hand-palm system TacTool that uses six vibration motors [344]. There, directionality (mainly of collision cues) was not always easily identified, whereas in our system, the simulated contact point was always well differentiated. While a contact point alone does not indicate an exact impact vector, it enables at least an identification of the general impact direction. Potential explanations for our different finding include the different locations and numbers of tactors, as well as a different hand posture. Finally, as the inside-out model partitions surroundings into zones irrespective of the amount of objects, we assume that our approach is resilient towards increasing object density in a scene, but have not yet verified this.

RQ2: *Can hand-driven proximity cues avoid unwanted object penetration or even touching proximate objects during manipulation tasks?*

In our manipulation task, we showed that audio-tactile proximity cues provided by the object-driven inside-out model significantly reduced the number of object collisions up to 30.3% and errors (object pass-throughs) up to 56.4%. With touch cues users thought they could perform faster (18.8%), more precise (21.4%), and adjust hand motion quicker (20.7%). Interestingly, audio cues alone also produced surprisingly good results, which is a useful finding as it potentially frees up vibrotaction for purposes other than proximity feedback. As fewer errors were made, we assume that proximity cues can enhance motor learning. Also, as haptic feedback plays a key role in assembly procedures [326], additional cues may not only optimize motion, but also hand poses. While we only indirectly

steer hand poses in this work, explicit pose guidance might a worthwhile extension [42]. Interestingly, our results somewhat contradict previous findings that identified bimodal feedback to be less beneficial in terms of throughput [10]. While we cannot calculate throughput for the steering task users performed, it would be interesting to investigate the measure on simpler tasks with our proximity models. Also, while we currently have a uniform tunnel width, it will be interesting to contrast our results to other tunnel widths in future work. Furthermore, users noted in their subjective feedback that single cues were, not entirely unexpected, easier to focus on than coupled cues. However, while cognitive load may pose an issue, it is not uncommon for multimodal interfaces to increase load [425]. In this respect, it is worth to mention related work [373] that has looked into bimodal (audio-tactile) and unimodal (tactile) feedback in touch related tasks. Results revealed a significant performance increase only *after* a switch from bimodal to unimodal feedback. The authors concluded that the release of bimodal identification (from audio-tactile to tactile-only) was beneficial. However, this benefit was not achieved in the reverse order. The interplay between modalities also gives rise to potential cross-modal effects. Previous work in the field of object processing using neuroimaging methods [199] has shown multisensory interactions at different hierarchical stages of auditory and haptic object processing. However, it remains to be seen how audio and tactile cues are merged for other tasks in the brain and how this may affect performance.

Overall, through our fully directional feedback, we extend previous findings on single-point, non-directional proximity feedback [10] that elicit constraints on dimensionality. We confirm that directional feedback can improve performance, in particular through a reduction of errors. We also improve on previous work by investigating fully three-dimensional environments. In this context, it would be interesting to assess performance differences between non-directional and directional feedback in the future, also for selection tasks, while also looking more closely at potential learning effects. While we focused on the usefulness of proximity feedback in manipulation tasks, we expect our inside-out feedback to also have a positive effect on selection tasks. Another open area is the trade-off and switching between outside-in and inside-out proximity feedback models based on the usage mode (selection versus manipulation versus exploration). Such switching has the potential to confuse users and thus necessitates further study.

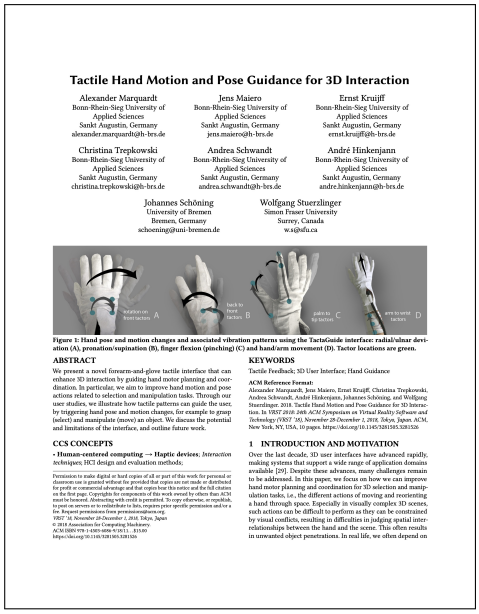
Similar to Ariza et al. [10], we studied the feedback methods in the absence of additional visual feedback in this work. This poses the question how our methods can be used in combination with visual feedback and what dependencies any given visualization

technique introduces in a real usage scenario. Naturally, information about objects around the hand is usually communicated over the general visual representation of the rendered objects, as will be the case during, e.g., learning assembly procedures. Yet performance may be affected by visual ambiguities. While visual and haptic stimuli integration theories [112] underline the potential of a close coupling of visual and non-visual proximity cues, ambiguities may still affect performance. Researchers have looked into reducing such ambiguities, for example through transparency or cut-away visualizations, where spatial understanding may vary [8]. Another approach to address ambiguities might be to provide hand co-located feedback, where first attempts have been presented previously, e.g., [336]. For example, portions of the hand could be color coded based on their level of penetration into surrounding objects. Hence, we are considering to verify performance of our methods in combination with visual feedback in the future, using both standard or optimized visualization methods.

2.7 Conclusion

In this work, we explored new approaches to provide proximity cues about objects around the hand to improve hand motor planning and action coordination during 3D interaction. We investigated the usefulness of two feedback models, outside-in and inside-out, for spatial exploration and manipulation. Such guidance can be highly useful for 3D interaction in applications that suffer from, e.g., visual occlusions. We showed that proximity cues can significantly improve spatial awareness and performance by reducing the number of object collisions and errors, addressing some of the main problems associated with motor planning and action coordination in scenes with visual constraints, which also reduced inadvertent pass-through behaviors. As such, our results can inform the development of novel 3D manipulation techniques that use tactile feedback to improve interaction performance. A logical next step requires integrating our new methods into actual 3D selection and manipulation techniques, while also studying the interplay with different forms of visualization (e.g., [385]) in application scenarios. In due course, the usage and usefulness of two gloves with audio-tactile cues is an interesting venue of future work, e.g, to see if audio cues can be mapped to a certain hand. Furthermore, we currently focused only on haptic feedback to eliminate potential effects of any given visualization method, such as depth perception issues caused by transparency. Finally, we are looking at creating a wireless version of the glove and to improve tracking further, e.g., by using multiple Leap Motion cameras [186].

3 Tactile Patterns for Motion Guidance



Tactile Hand Motion and Pose Guidance for 3D Interaction

Alexander Marquardt
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
alexander.marquardt@h-brs.de

Jens Maiero
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
jens.maiero@h-brs.de

Ernst Kruijff
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
ernst.kruijff@h-brs.de

Christina Trepkowski
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
christina.trepkowski@h-brs.de

Andrea Schwandt
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
andrea.schwandt@h-brs.de

André Hinkenjann
Bonn-Rhein-Sieg University of Applied Sciences
Sankt Augustin, Germany
andre.hinkenjann@h-brs.de

Johannes Schöning
University of Bremen
Bremen, Germany
schoening@uni-bremen.de

Wolfgang Stuerzlinger
Simon Fraser University
Surrey, Canada
w.s@sfu.ca

Figure 1: Hand pose and motion changes and associated vibration patterns using the TactaGuide interface: radial/ulnar deviation (A), pronation/supination (B), finger flexion (pinching) (C) and hand/arm movement (D). Tactor locations are green.

ABSTRACT

We present a novel forearm-and-glove tactile interface that can enhance 3D interaction by guiding hand motor planning and coordination. In particular, we aim to improve hand motion and pose actions related to selection and manipulation tasks. Through our user studies, we illustrate how tactile patterns can guide the user, by triggering hand pose and motion changes, for example to grasp (select) and manipulate (move) an object. We discuss the potential and limitations of the interface, and outline future work.

CCS CONCEPTS

• **Human-centered computing** → **Haptic devices**; *Interaction techniques*; HCI design and evaluation methods;

KEYWORDS

Tactile Feedback; 3D User Interface; Hand Guidance

ACM Reference Format:

Alexander Marquardt, Jens Maiero, Ernst Kruijff, Christina Trepkowski, Andrea Schwandt, André Hinkenjann, Johannes Schöning, and Wolfgang Stuerzlinger. 2018. Tactile Hand Motion and Pose Guidance for 3D Interaction. In *VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), November 28-December 1, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3281505.3281526

VRST '18, November 28-December 1, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6086-9/18/11…$15.00
https://doi.org/10.1145/3281505.3281526

1 INTRODUCTION AND MOTIVATION

Over the last decade, 3D user interfaces have advanced rapidly, making systems that support a wide range of application domains available [29]. Despite these advances, many challenges remain to be addressed. In this paper, we focus on how we can improve hand motor planning and coordination for 3D selection and manipulation tasks, i.e., the different actions of moving and reorienting a hand through space. Especially in visually complex 3D scenes, such actions can be difficult to perform as they can be constrained by visual conflicts, resulting in difficulties in judging spatial interrelationships between the hand and the scene. This often results in unwanted object penetrations. In real life, we often depend on

In this chapter, we present a novel forearm-and-glove tactile interface that can enhance 3D interaction by guiding hand motor planning and coordination. This interface extends the technical principles and results that were developed as part of my master's thesis and are described in Chapter 2. Influencing sensory constraints examined in this work included depth perception, sensory thresholds, and scene structure.

This work was designed to improve hand motions and postures in selection and manipulation tasks by providing tactile cue patterns. This approach is considered particularly useful in visually complex scenes in which target positions may not be visible to the user at all times. By providing guidance based on vibrotactile cue patterns, task performance can be supported, for example, in assembly tasks that require fine-grained movements and postures. In Study 1 (n = 8), we showed that users were able to localize and differentiate tactile cues reasonably well on different arm and hand locations. Study 2 (n = 8) examined the interpretation and preference of tactile cues and patterns, showing that the recognition and interpretation worked well but was not completely error-free. Study 3 (n = 8) used a Wizard-of-Oz system to assess the cues in a simulated selection and manipulation task. Here, we found that tactile patterns could be used to trigger fine-grained motions and poses that could support 3D selection and manipulation.

The material in this chapter originally appeared in: Marquardt, A., Maiero, J., Kruijff, E., Trepkowski, C., Schwandt, A., Hinkenjann, A., Schöning, J., & Stürzlinger, W. (2018). Tactile Hand Motion and Pose Guidance for 3D Interaction. *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 1–10. DOI: 10.1145/3281505.3281526

3.1 Introduction and Motivation

Over the last decade, 3D user interfaces have advanced rapidly, making systems that support a wide range of application domains available [233]. Despite these advances, many challenges remain to be addressed. In this paper, we focus on how we can improve hand motor planning and coordination for 3D selection and manipulation tasks, i.e., the different actions of moving and reorienting a hand through space. Especially in visually complex 3D scenes, such actions can be difficult to perform as they can be constrained by visual conflicts, resulting in difficulties in judging spatial interrelationships between the hand and the scene. This often results in unwanted object penetrations. In real life, we often depend on complementary haptic cues to perform tasks in visually-complex situations. However, including haptic cues is not always straightforward in 3D applications, as it often depends on complex mechanics, such as exoskeletons or tactor grids.

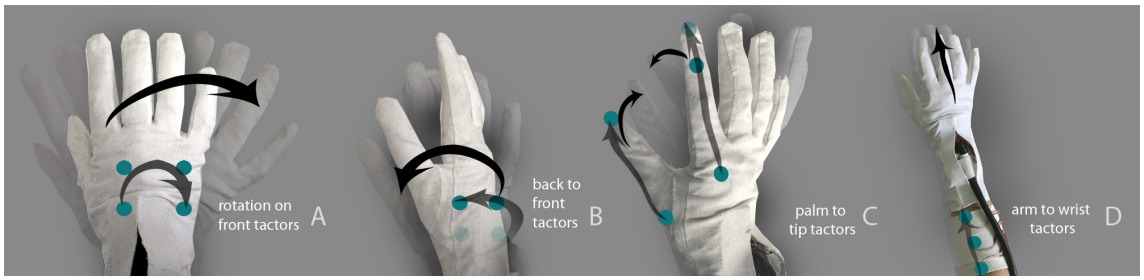


Fig. 3.1: Hand pose and motion changes and associated vibration patterns using the TactaGuide interface: radial/ulnar deviation (A), pronation/supination (B), finger flexion (pinching), (C) and hand/arm movement (D). Tactor locations are green.

3.1.1 Cues for Motor Planning and Coordination

Motor planning and coordination of selection and manipulation tasks is generally performed in a task chain with key control points that relate to biomechanical actions [190]. These actions contain contact-driven touch events that can inform the planning and coordination of hand motion and pose actions. For example, a user may grasp an object (touch informs hand pose to grasp) and change its rotation and translation in space by moving and reorienting the hand (motion, pose) while avoiding touching other objects (touch) [233]. As the hand-arm is a biomechanical lever system, hand motion can be accomplished by arm motion, but also by wrist rotation. Within this article, we specifically focus on motion and pose guidance, and reflect on interrelationships with touch in our discussion. Pose not only relates to the orientation of the hand itself but also to its specific postures needed to select and manipulate an object, e.g., to grasp or move an object through a

tunnel. While contact-point feedback on a user's hand may provide useful feedback to avoid touching other objects during pose and motion changes, such actions can also be performed independent of (or even avoid) touch contact. To do so, both in real life and in 3D applications we may rely on proprioceptive cues, which are typically acquired through motor learning [369]. However, cues beyond proprioception and visual feedback about the scene may be required to perform (or learn) a task correctly. So-called augmented feedback – information provided about an action that is supplemental to the inherent feedback typically received from the sensory system – is an important factor supporting motor learning [238]. While learning how to optimally perform a task – regardless of whether it is in a purely virtual environment or a simulated real-world task – most interfaces unfortunately do not provide feedback to encourage correct hand motions and poses, i.e., no form of *guidance*. However, selection and manipulation tasks, and potentially subsequent motor learning, likely will benefit from such guidance. For example, consider training users for assembly tasks where knowledge acquired in a virtual environment needs to be transferred to the real world [42].

3.1.2 Limitations of Haptic Devices for Pose and Motion Guidance

Traditional haptic interfaces, such as the (Geomagic) Phantom, can guide hand motion to a certain extent to improve selection and manipulation task performance, often in a contact-driven manner. As such, haptics can potentially overcome limitations caused by visual ambiguities that, for example, make it difficult to judge when the hand collides with an object [53]. However, there are certain limitations that directly affect motion and pose guidance. Most common haptic devices depend on a pen-based actuation metaphor instead of full-hand feedback. How we hold an actuated pen does not necessarily match how we interact with many objects in real life. Furthermore, while typical contact-driven haptic feedback models support overall motion guidance, they do not aid users in achieving a specific pose, unless a full-hand interface like an exoskeleton is used. Finally, most haptic devices are limited in operation range, imposing constraints on the size of training environments.

3.2 Related Work

Haptic feedback for 3D interaction has been explored for many years, though is still limited by the need for good cue integration and control [223, 384], cross-modal effects

[313], limitations in actuation range [163], and fidelity issues [281]. The majority of force feedback devices provide feedback through a grounded (tethered) device. These devices are often placed on a table and generally make use of an actuated pen that is grasped by the fingertips, instead of full hand operation, e.g., [271]. In contrast, glove or exoskeleton interfaces can provide feedback such as grasping forces and enable natural movement during haptic interactions [39, 364]. Few haptic devices provide feedback for the full hand. An example is the CyberGrasp (CyberGlove systems), a robot-arm actuated glove system that can provide haptic feedback to individual fingers. Tactile methods afford more flexibility by removing the physical restrictions imposed by the actuated (pen-)arm or exoskeleton construction. However, they can be limited as haptic cues have to be "translated" within the somatosensory system [195]. While substituted cues have been found to be a powerful alternative [222, 225], they can never communicate all sensory aspects. In 3D applications, research has mostly revolved around smaller tactile actuators that are hand-held, e.g., [32], or glove-based, e.g., [131]. Some work has explored the usage of a dense vibrotactors grid at or in the hand, e.g., [136, 269, 344], which is related to our glove design.

Some systems provide guidance cues to trigger body motions and rotations. Most approaches focus on corrective feedback with varying degrees of freedom. The majority of systems focuses on some form of motor learning, which may be coupled with visual instructions of the motion pattern [241]. Effective motion patterns have yet to be found, as illustrated by the variety of patterns in the different studies [23]. However, one common insight is that the spatial location of vibrations naturally conveys the body part the user should move and that saltation patterns are naturally interpreted as directional information [380]. Such saltation patterns are a sequence of properly spaced and timed tactile pulses from the region of the first contactor to that of the last, allowing for good directionality perception [7]. Yet, there is no conclusive answer for rotation patterns. Researchers have provided cues at arms, legs and the torso [330] to train full-body poses that, for example, help with specific sports like snowboarding [380]. Research has also focused specifically on guiding arm motions [370, 411] in 3D environments. Further variants of this work look at arm [72] or wrist rotation [386] for more general applications. All these methods target only general motions and are not particularly useful for hand pose and motion guidance for 3D selection and manipulation. In contrast, other systems use electromuscular stimulation (EMS) to control hand and arm motions to produce finer motions and poses [396]. The most closely related work looked at triggering muscular actions at the hand and arm via

EMS [253]. Yet, EMS systems are awkward to use, and often have limited usage duration or user acceptance. Also, receptors or muscles may get damaged through use of EMS [303]. For hand guidance, the usage of proximity models to improve spatial awareness around the body to indirectly trigger hand motion and pose adaptations is another related area. Some researchers have explored proximity cues with a haptic mouse [173], the usage of proximity to trigger actions [29], and auditory feedback for collision avoidance [2].

Extending the state of the art, we introduce a novel set of vibrotactile cues that can guide hand motion and pose configurations that have high relevance for 3D selection and manipulation.

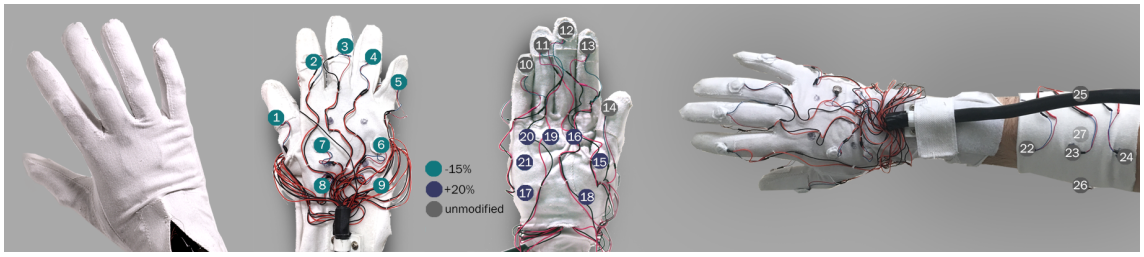


Fig. 3.2: Tactor IDs and balancing of TactaGuide glove, based on pilot study results. The tactors at the arm sleeve were unmodified.

3.2.1 Research Questions

To design an effective tactile interface for motion and pose guidance, we need to address several challenges. In this paper, we examine how we can guide the user to perform specific motion and pose actions along key control points in the task chain, ideally independent of contact events. Doing so, we can identify the following three research questions (RQ).

RQ1: *How well can tactors be localized and differentiated across the hand and lower arm?*

RQ2: *How do users interpret tactile pose and motion patterns and what are their preferences?*

RQ3: *How does tactile pose and motion guidance perform in a guided selection and manipulation task?*

In this paper, we assess each RQ through a respective user study. In Study 1, we measure the effects of vibration on localization/differentiation, which informed Study 2, which looks into the interpretation of tactile cues on pose and motion changes, while analyzing user preference for patterns. Study 3 takes the main user preferences and uses a Wizard-of-Oz methodology to assess the cues in a simulated selection and manipulation task, where we

measured the effectiveness of operator-controlled cues. This study is designed to illustrate cue potential in real application scenarios.

3.2.2 Contributions

In this paper, we present the design, implementation and validation of a tactile pose and motion guidance system, TactaGuide, which is a vibrotactile glove and arm sleeve interface. We show that our new guidance methods afford fine hand motion and pose guidance, which supports selection and manipulation actions in 3D user interfaces. We go beyond the state of the art that mainly focused on vibrotactile cues for body and arm motions [23, 211, 275, 370, 380], or general poses [41, 463]. In that, we extend previous work to fine hand manipulation actions through a set of vibrotactile cues provided via TactaGuide, through the following findings:

- Localization and differentiation: we show that tactors can be well localized at different hand and arm locations and illustrate that simultaneous vibration works best. We also show that the back of the hand (normally used infrequently) scored as good as the index finger and is a useful location for contact-driven feedback.
- Pattern interpretation: Based on the biomechanical constraints of various hand/arm parts, we illustrate that most users successfully match patterns to the right motion or bodily reconfiguration.
- Selection and manipulation guidance: through a Wizard-of-Oz experiment we show that vibration patterns support finer-grained 3D selection and manipulation tasks, confirming the validity of our approach.

We deliberately performed all studies in the absence of visual cues to reliably identify the effect of tactile guidance in isolation, with an eye towards eye-free interaction scenarios. We reflect on the potential for combinations of visual and tactile patterns for guidance in the Section 3.5 “Discussion”.

3.3 Approach

To overcome these limitations, we investigate the use of tactile feedback, even in non-contact situations. Tactile feedback is unique in that it directly engages our motor learning systems [241], and performance is improved by both the specificity of feedback and its immediacy [7]. Deliberately, we give tactile feedback independent of visual cues, to avoid confounds or constraints imposed by such visual cues. Normally, designing tactile cues is

challenging, as haptic (force) stimuli cannot be fully replaced by tactile ones without loss of sensory information [195]. To avoid this issue, we provide instructional tactile cue patterns, instead of simulating contact events. Also, tactile devices can provide light-weight solutions with good resolution and operation range [261, 424]. Current touch-based vibrotactile approaches typically do not provide pose and motion requirement indications. In our study, we look specifically at feedback that addresses these issues, by providing feedback to guide the user to move in a particular way or assume a specific hand pose. Our methods use localized vibration patterns that trigger specific bodily reconfigurations or motions. Previous work, e.g., [358, 370, 380], indicates that vibration patterns – independent of touch actions – can aid in changing general body pose and motion, which we extend in this work to support more fine-grained selection and manipulation actions.

Pose and Motion Guidance Feedback

We provide tactile feedback through our new TactaGuide system, a vibrotactile glove and arm sleeve (Fig. 3.2). The device affords a full arm motion operation range, tracked by a Leap Motion. Both glove and sleeve are made of stretchable eco-cotton that is comfortable to wear. In the glove, tactors are placed at the fingertips (5 tactors), inner hand palm (7), middle phalanges (5), and the back of the hand (4), for a total of 21 tactors (Fig. 3.2). Cables are held in place through a 3D printed plate embedded in the fabric on top of the wrist. The arm sleeve consists of 6 tactors, positioned to form a 3D coordinate system "through" the arm. We use 8-mm Precision Microdrive coin vibration motors (model 308-100). All tactors are driven by Arduino boards. To overcome limitations in motor response caused by inertia (tactors can take up to ~75 ms to start), we use pulse overdrive [269] to reduce the latency by about 25 ms. After that, pulse width modulation (PWM) is used to reduce the duty cycle to the desired ratio under consideration of the corresponding tactor balancing (Fig. 3.2) to generate different tactile patterns. The system was previously used for another purpose, namely proximity feedback [265], where we showed that proximity cues in combination with collision and friction cues can significantly improve performance.

Many selection and manipulation tasks depend on fine control over hand motion and poses. However, in complex 3D scenes, such motor actions maybe be difficult to plan and coordinate. For example, consider training the hand to move behind an object, to grasp a small and occluded object (or part). While adjusting the visualization may solve some issues – x-ray visualization has been used to look "through" an occluding object [21] – the

associated visual ambiguities can make performing the task challenging. To overcome such visual limitations, we assume that tactile cues are valuable to guide hand motion and poses. Inspired by related work, e.g., [72, 386], the basic premise of our hand motion and pose guidance system is centered around providing various pattern stimuli – activating tactors in a specific region in a specific sequence – using a specific vibration mode (Fig. 3.1 and 3.4). Previous work indicates that such patterns are well interpretable by the user, while cue location and directionality inform the user about the specific body part or joint that should be actuated [241]. These cues can be triggered independent of contact events, i.e., events that relate to touching an object. For example, stimulating three tactors in a serial manner from hand palm to fingertip may indicate to the user that they should stretch that finger (Fig. 3.1C). Similarly, a forward pattern over the arm may indicate the arm needs to be moved forward (Fig. 3.1D). Further details on the patterns are discussed in Section 3.4.3. By focusing on motion and pose adjustment for selection and manipulation, which requires finer control over hand and fingers, we extend previous work [72, 386], that focused only on arm or wrist rotation. Our target actions are closer to EMS-based work [396], though without their aforementioned limitations.

We looked closely at the different actions undertaken by the hand during 3D selection and manipulation. Each of these actions is generally associated with a specific hand or arm region. The different posture/motion actions refer to fundamental hand movements (Fig. 3.1) and thus to biomechanical actions that involve various joint/muscle activations:

- Radial/ulnar deviation: turning of the hand (yaw).
- Pronation/supination: rotation of the hand (roll).
- Move: arm movement to move the hand in the scene, including abduction and adduction (moving arm up and down), forward/backward and left/right motion afforded by the arm lever system.
- Finger flexion/extension: straightening of fingers to pinch or grasp an object.

While flexion and extension can also refer to orienting the hand around the wrist (pitch), we did not support this motion in our work, as it is used infrequently in the frame of selection and manipulation tasks. For fingers, we use different patterns for closing (palm to fingertip vibration) and opening gestures (fingertip to palm vibration), while hand rotations simply involve directional patterns. With respect to arm movement, the arm is a biomechanical lever system as bones and muscles form levers in the body to create human movement – joints form the axes, and the muscles crossing the joints apply the force to move the arm.

Based on ease of detection of location, direction, and guidance interpretation (which hand motion or pose change does the pattern depict?), we implemented three different vibration modes, which we then assessed in our user studies. The location of a stimulus guides the biomechanical action. E.g., when a finger needs to be bent, the vibration pattern is provided at the finger [380]. The three modes were continuous (a continuous vibration stimuli), stutter (a pulsed vibration stimuli), and mixed (a mixture of both). We assumed that the stutter at the end of the mixed mode pattern could indicate direction. Prior to the studies, we performed a pilot study, where we verified stimuli with 5 users and fine-tuned the system.

3.4 Experiment

Pose and motion guidance was examined in three studies, 1, 2 and 3, which investigated how well different vibration patterns and modes trigger hand pose and motion changes, to potentially guide the design of haptic selection and manipulation techniques. These studies were designed to show if hand pose and motion guidance is principally possible, and to investigate its potential and limitations. As noted before, we deliberately did so independently of visual cues, to avoid confounds or constraints imposed by such cues.

Different user samples were recruited for each study. In each study users wore the complete TactaGuide glove and arm sleeve setup. Post-hoc questionnaires for each study were composed of 7-point Likert items (0 = “fully disagree” to 6 = “fully agree”), related to mental demand, comfort, usability, and also task-specific perceptual issues. Users were seated at a desk and could rest their elbow on the armrest of a chair in Study 1 and 2, while vibrotactor locations (IDs) were shown on a 27" desktop screen. In Study 1 we examined if and to what extent our glove enables users to accurately localize tactile feedback and their ability to discriminate between different tactors. Study 2 focused on the user’s interpretation of vibration patterns into assuming hand poses and performing motions. In Study 3, the user’s hand pose and motion were guided through vibration patterns that were chosen on the basis of the previous studies. Study 3 deployed a Wizard-of-Oz methodology to overcome finger tracking limitations associated with the LeapMotion, which cannot reliably detect the hand once it is rotated vertically. Yet, this pose is required for many grasping actions.

3.4.1 Study 1 - Tactor Localization and Differentiation

This study focused on the ability of users to locate and differentiate between tactors to ensure that users can detect the actual region that receives biomechanical actuation. As higher-resolution tactile gloves are scarce, there is no information in the literature about the detectability of individual tactor locations (stimuli), especially with respect to our particular locations at the TactaGuide glove. Also, while sensitivity is well studied for the inside of the hand, sensitivity at the back of the hand has hardly been studied [191].

In task 1, participants were asked to locate a single actuated tactor. A within subjects 2 x 2 factorial design was employed to study the effect of factor feedback mode (stutter, continuous) and hand pose (straight, fist) on feedback localization performance (mean hits per trial). Vibration feedback was provided at all 21 different hand locations of the TactaGuide glove, resulting in 84 trials. Two feedback modes were also compared at 6 locations on the wrist, resulting in 12 additional trials. The total of 96 trials were randomly presented. Participants were informed that only a single tactor provided feedback at any given time. In each trial feedback was provided for two seconds, after which the participant selected a tactor (ID) from the overview shown on a desktop monitor showing the hand with tactor locations.

In task 2, combinations of two or three actuated tactors had to be located and differentiated. A 2 x 4 x 7 factorial design was used to study the localization of tactors depending on their number (two or three tactors), feedback mode (simultaneous, continuous; simultaneous, stutter; serial, continuous; serial, stutter) and zone (thumb, index, pinkie, palm, back of the hand, from the back to the inner hand, wrist). Each factor combination was repeated, resulting in 112 trials, presented in randomized order. Before starting the task, participants were informed that either two or three tactors would be actuated. Feedback was always provided for two seconds. As in the first task, participants responded with tactor ID displayed at the screen. Together, both tasks took around 45 minutes to complete.

3.4.2 Results of Study 1

Eight right-handed persons (2 females, mean age 39 (SD 15.7), with a range of 25–65 years) volunteered. Six wore glasses or contact lenses and two had normal vision. Within subjects repeated-measures analysis was used to study task specific main and interaction effects of factors on dependent measures.

In task 1, a total of 768 trials were analyzed. For each trial the actually activated tactor and the participant's choice were compared, to record a hit as the correct tactor was chosen (1) or a miss if not (0). As expected, the hand pose but not the mode affected hit rate (hits/trials), which was significantly higher with a straight hand pose (M = 0.82, SE = 0.02) than with a fist (M = 0.69, SE = 0.4), F(1 ,7) = 13.44, p = .008, η^2 = .66. With a fist, tactors are closer together, making it more difficult to localize a stimulus. In a secondary analysis, tactors were grouped into six zones across which we compared hit rates (thumb; middle fingers: [index,middle,ring]; pinkie; back of the hand; palm; wrist). The zone affected the hit rate, F(5,35) = 6.48, p < .001, η^2 = .48. Post-hoc comparisons showed that only the pinkie with the lowest hit rate (M = 0.61, SE = 0.05) differed significantly from the back of the hand, which had a high hit rate (M = 0.85, SE = 0.03), p = .015.

In task 2, a total of 896 trials were analyzed. In this task activated tactors were compared to participants' responses. Depending on their perception, participants could either name three tactor IDs or they could name less than three and state there were no more activated tactors. We scored a hit for each correctly named tactor and also for correctly stating that no more tactor was activated. That is, the maximum number of hits per trial was always three. Mean hits depended significantly on the stimulated zone F(6,42) = 2.62, p = .03, η^2 = .27 (see Fig. 3.3 for mean values and standard errors), the feedback mode F(3,21) = 10.81, p < .001, η^2 = .61 and its interaction with the number of activated tactors F(3,21) = 22.98, p < .001, η^2 = .77 (see Table 3.1 for mean values and standard errors). A post-hoc test showed that the mean number of hits was higher when feedback was provided at the back of the hand compared to the thumb, the pinkie, and the palm. Performance on the back of the hand was also marginally better than feedback transitioning from the back to the inner hand (p = .058). There were also more hits when feedback was provided at the index finger than at the palm (p = .048). In trials with two activated tactors and for both simultaneous feedback modes, participants got more hits compared to both serial activations (p < .01). When three tactors were activated, differences became non-significant.

Performance was best at the index finger and the back of the hand. While the mean differences between zones were statistically significant, they were relatively small (up to 0.23 = 8% of the maximum score). This outcome might be related to the distribution and sensitivity of mechanoreceptors of glabrous skin [191], where the density of low threshold mechanoreceptive units at the fingers is principally higher than in the palm. Therefore, vibrations are in general harder to differentiate inside the palm, especially in

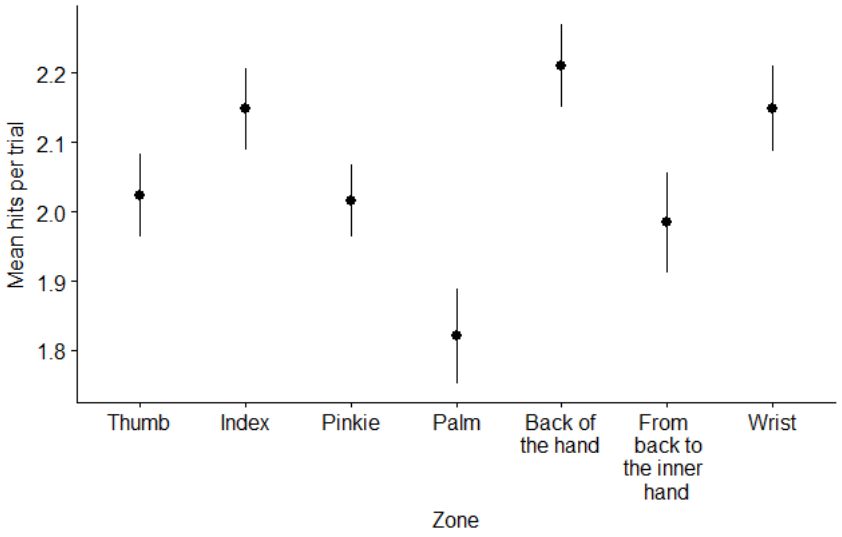


Fig. 3.3: Study 1, task 2 (tactor localization and differentiation): Mean number of hits per trial by stimulated zone with standard errors (SE) hits per trial, hit range = [0;3].

Table 3.1: Study 1, task 2 (tactor localization and differentiation): Mean hits per trial by number of activated tactors and feedback mode with standard errors (SE), hit range = [0;3].

Number of tactors	Feedback mode	Mean (SE)
Two	Si-C	2.33 (0.09)*
	Si-S	2.45 (0.09)*
	Se-C	1.72 (0.08)
	Se-S	1.87 (0.09)
Three	Si-C	1.89 (0.08)
	Si-S	1.94 (0.09)
	Se-C	2.15 (0.16)
	Se-S	2.05 (0.14)

Si = Simultaneous, Se = Serial, C = Continuous, S = Stutter

case of adjacent, nearby located tactors. Simultaneous activations led to better performance compared to serial continuous activation when two tactors vibrated. Mean differences ranged from 0.46 to 0.72 (=15% to 24% of the maximum score). However, when three tactors were activated, participants generally achieved a good hit rate for serial feedback, as they correctly identified two out of three tactors on average. There was no interaction effect between feedback mode and stimulated region, that is, the optimum feedback mode was not region specific.

3.4.3 Study 2 - Pattern Interpretation and Preference

We explored motion interpretation and preferences in this observational study in two different tasks. In task 1, we focused at how users would interpret a certain trigger (pattern + mode) by adjusting their hand pose or motion, while task 2 investigated which vibration mode was preferred for a stated hand pose or motion change.

For task 1 of Study 2, feedback was provided at the same six hand zones as in the second task of Study 1 (localization and differentiation), as well as at the wrist and at an additional hand zone that includes the thumb and index. A specific feedback pattern with varying numbers of involved tactors depending on the zone, see Table 3.2. We actuated the tactor-vibrations serially in three modes: Stutter, continuous and a mixed mode (see Fig. 3.4).

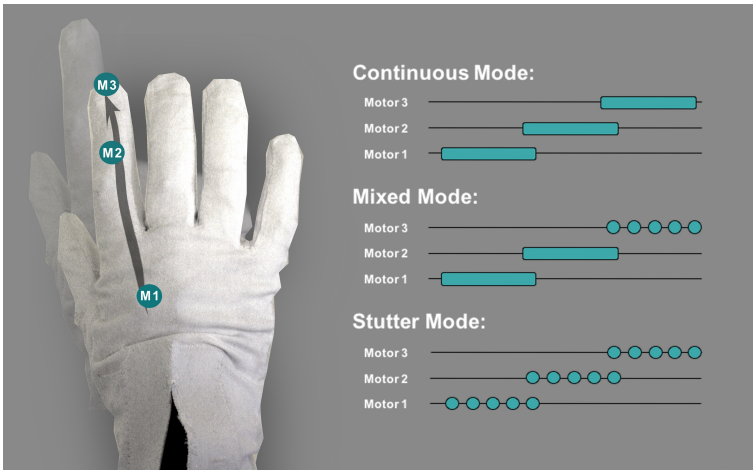


Fig. 3.4: Activation sequence of different feedback modes using the example of finger pointing motion (index finger) with three involved tactors.

In mixed mode, the first tactor(s) was in continuous mode, while the last one was stuttering. Unlike Study 1, simultaneous feedback modes were not used in Study 2, as we provided directional feedback cues through serial activation. Feedback patterns at each zone were provided using zone-specific vectors in two opposite directions (forward/clockwise and backwards/counterclockwise), except for the wrist at which three vectors with opposite directions were provided (forward/backwards; up/down; left/right). Feedback was provided and randomized blockwise. Participants completed one block of 36 trials with feedback at six hand zones first (6 regions x 3 modes x 2 directions), followed by 18 trials for the wrist (3 modes x 3 vectors x 2 directions) and finally 6 trials involving the thumb and index at the same time (3 modes x 2 directions), for a total of 60 trials per participant. Participants were told to change their hand pose in a way that

they felt matched the provided pattern best. The starting pose for each trial was resting the elbow on the armrest of a chair while the hand was hanging down in a relaxed manner (i.e., a pose between a fist and fully stretched hand gesture). No further instructions were given and users could choose their movements and gestures freely. The experimenter recorded the resulting motions.

For task 2, the zone-specific feedback patterns and directions were the same as in task 1. We pre-defined specific hand poses for each zone-specific feedback pattern and direction, see Table 3.2. In each trial, the experimenter first demonstrated which movement or hand pose should be initiated by the feedback that followed. Then the corresponding feedback was provided in three different modes (continuous, stutter, mixed mode), presented in randomized order. The modes were not examined as factor but functioned as response options: that is, the user had to choose which cue was most suitable for initiating the previously shown movement or pose. The suitability of the feedback for the respective movement/pose was also rated on a 7-point Likert scale (6 being "totally suitable"). As in task 1, six hand zones, the wrist, and the zone including thumb and index were tested and randomized blockwise. With one repetition 24 trials were presented for the six hand zones (6 zones x 2 directions x 2 repetitions), 18 trials for the wrist (2 repetitions x 3 vectors x 3 directions) and 4 trials for thumb and index zone (2 repetitions x 2 directions), resulting in 46 trials. The experimenter recorded the choice of mode and suitability rating for each trial. Eight participants (7 right-handed, 2 females, mean age 29.6, SD 5.3, with a range of 23–40 years) volunteered. Three wore glasses or contact lenses and five had normal vision.

3.4.4 Results of Study 2

For task 1, 480 trials were analyzed. All feedback-dependent interpretations were listed and counted if they occurred sufficiently often, that is, were used by at least three of the participants. When feedback was provided at the thumb, the back of the hand, palm, or wrist, resulting movements were diverse for each feedback direction and mode and no coherent movement/gesture could be observed. Feedback provided once at the index, pinkie and at thumb and index, or repeatedly from the back to the inner hand resulted in "successful" movements/gestures, which correspond to our interpretation of the respective feedback. That is, forward/backward feedback at the index and pinkie resulted in stretching/bending respective fingers, simultaneous feedback at the thumb and index was

Table 3.2: Study 2, task 1 (pattern interpretation) and 2 (preference): Pre-defined hand movements depending on zone, activated tactors and feedback direction. The + symbol indicates simultaneous activation of concatenated numbers.

Zone	IDs of activated tactors (see Fig. 3.2) and order of activation	Movement for tactor activation → from left to right, ← from right to left
Thumb	7, 1, 14	→ stretch ← bend
Pinkie	6, 5, 10	
Index	7, 2, 13	
Thumb and Index	7, 1 + 2, 13 + 14	→ pinch ← release
Hand inner	18, 16, 20, 17	→ ulnar deviation ← radial deviation
Back of the hand	8, 7, 6, 9	
From back to inner hand	7, 6, 20, 16	→ supination ← pronation
Wrist	24, 23, 22	→ forward ← backward
	26, 23, 25	→ right ← left
	27, 23	→ up ← down

interpreted as pinch movement and feedback provided from the back to the inner hand resulted in supinations.

For task 2, 384 trials were analyzed. Mode preferences for hand and wrist were analyzed separately, as three instead of two directional vectors were used for the wrist. For each participant and factor combination we calculated how many times each mode was preferred. With one repetition each mode could maximally be preferred two times for a given combination. Generally, the continuous mode was preferred at the hand, M = 1.21, SE = 0.1, over the stutter, M = 0.2, SE = 0.06, p = .001, and mixed mode, M = 0.6, SE = 0.06, p = 0.18, F(1.15,8.03) = 30.09, p < .001, η^2 = .81. Nevertheless, this preference was not consistent across zones as, especially at the back of the hand and the palm, the mixed mode was chosen more often than the continuous mode, but not significantly so. At the wrist the continuous mode was also preferred, F(2,14) = 8.71, p = .003, η^2 = .56. Post-hoc comparisons showed that the continuous mode, M = 1.27, SE = 0.19, was significantly superior to stutter vibration, M = 0.25, SE = 0.1, p = .02. Mode preferences in percent by zone are listed in Fig. 3.5.

The direction (at hand zones and wrist) and the vector (at the wrist) did not affect mode preference. Suitability ratings were generally slightly positive, while feedback patterns

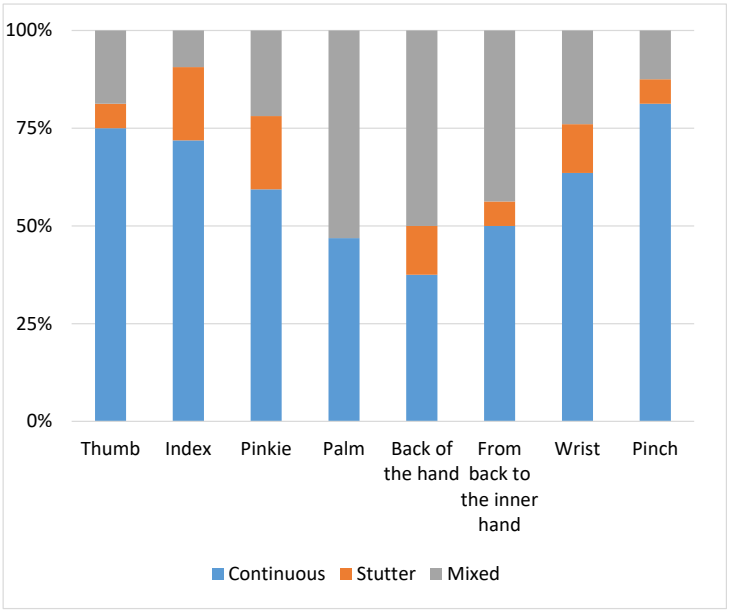


Fig. 3.5: Task 2: Vibration feedback mode preferences by zone in percent.

that were provided on the wrist to trigger up/down and left/right movements got more neutral ratings.

Results from task 1 indicate, that in principle patterns can be reasonably well interpreted, i.e., users did perform the intended main action. However, the interpretation of direction was often an issue. Most likely, the generally good detection of the main action can be associated to the biomechanical limitations and prime actions of hand and fingers, e.g., fingers are mainly bent, not rotated. Still, as we did not inform users what kind of action a pattern could potentially trigger, they had little possibility of learning a pattern. For task 2, it is not clear why the mixed mode was preferred for some areas. One possible explanation is that both areas (inner, back of hand) are quite flat, and exhibit different mechanical properties compared to, for example, the fingers. Suitability ratings indicated that feedback patterns used at the hand zones and wrist are generally appropriate for guidance.

3.4.5 Study 3 - Hand Pose and Motion Guidance

Based on the outcomes of the first two studies (1 and 2), we performed a Wizard-of-Oz [139] study to assess the cues for controlling finer-grained hand selection and manipulation actions. We deliberately chose a Wizard-of-Oz methodology to overcome some of the evident limitations of the hand tracking system we used (Leap Motion), which cannot track fingers precisely when the hand is held vertically, due to the occlusion of the fingers in the camera image. This study investigated user performance in six selection and manipulation tasks that cover hand pose changes and hand motions. Grids were used to control and

measure performance on the horizontal and vertical plane with 25 x 16 grid fields on each plane and a grid field size of 2 x 2 cm, see Fig. 3.6.

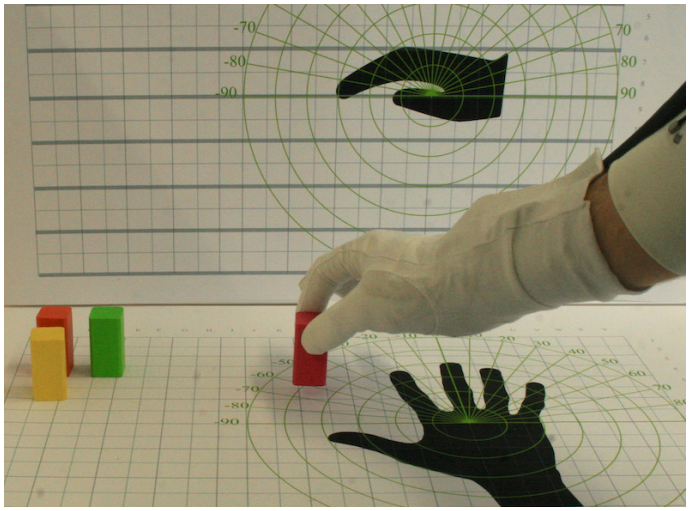


Fig. 3.6: Apparatus for Study 3, showing the measurement grids used for observing performance in the tasks.

The six tasks involved 1) moving the hand to a specific field in straight horizontal directions on the grid and 2) on the vertical plane using the shortest path, 3) performing supination/pronation, 4) radial/ulnar deviation, 5) pointing and 6) grasping one of four wooden blocks that were arranged on the horizontal plane in a 2 x 2 matrix. We included pointing in addition to selection and manipulation, as it is often used for cohesion in training tasks. To trigger actions, we applied a pattern that we also used in Study 2 and that corresponds to a pre-defined motion, see Table 3.2: pinching was used to grasp blocks. We decided to use the continuous vibration mode as it was preferred overall in Study 2. Before starting the actual experiment, participants received a 5-minute training session to learn the association between vibration feedback patterns and corresponding actions.

Each participant performed the six tasks in random order. The experimenter acted as operator who had an overview about the tasks and the order and "controlled" each action of the participant step by step, using a visual interface to trigger the predefined patterns. The operator started and stopped the specific feedback that was required for the respective task. False movements were not corrected, that is, if the user's hand moved too far, the operator provided feedback, as if the hand was at the correct position. After a task was finished, an observer (assistant of the experimenter) who was not aware of the targeted position and who could only see the participant, recorded the final position of the hand, noted any further observations and took pictures. After having finished all six tasks, the participant started a new trial that required him/her to do the six tasks again in a random order. All

tasks were the same for the second time, except for grasping the block. When participants encountered a task for the first time, blocks had a distance of two fields between each other. The second time around the difficulty was raised by reducing the distance to one field. Study 3 was video-recorded with permission of the users. After having completed the study, participants rated feedback perception, task easiness, needed concentration, ease of remembering movements/gestures that correspond to a respective feedback, suitability of feedback and their performance. Eight right-handed participants (2 females, mean age 35.8 (SD 16.4), with a range of 23–65 years volunteered.

3.4.6 Results of Study 3

For Study 3, we analyzed 48 trials. The comparison of the targeted and the actually reached grid field showed that participants could be guided quite precisely to a specific grid field on the horizontal plane. In the first trial, the reached field had only an average deviation of M = 1.88, SD = 1.36 fields from the targeted one, and M = 2.25, SD = 2.05 in the second trial. Deviations on the vertical plane were even smaller: M = 0.88, SD = 0.35 in the first and M = 0.63, SD = 1.06 in the second trial. Pointing and grasping the bricks at the two difficulty levels was always successful. Nevertheless, sometimes participants confused radial/ulnar deviation with supination/pronation, radial with ulnar deviation and up/down with left/right feedback. Participants' ratings were compared between different tasks. Generally, all ratings were positive, especially concerning pointing and grasping. While ratings for the tasks that targeted supination/pronation, radial/ulnar deviation and moving the arm around received slightly positive feedback, ratings for pointing and grasping were strongly positive. Suitability ratings for moving the arm up/down and left/right were also slightly positive and higher than in Study 2, task 2. Grasping and pointing required even less concentration than the other tasks and the assignment of the vibration feedback to the movement/gesture seems to have been easier to remember. Participants thought they performed better in pointing and grasping than in the other tasks and that the pattern initiating pointing and grasping fitted "better" compared to other patterns. Overall comfort and usability ratings are listed in Table 3.3. In general ratings are rather positive, only the cable seemed to have disrupted users slightly, which could be due to the weight of the cable as users also felt somewhat exhausted after wearing the glove for some time.

While results are generally encouraging, hand rotation guidance was not followed reliably. As noted below in the discussion section, based on previous work [24], we can

Table 3.3: Overall comfort and usability ratings for Study 3.

Statement	Mean (SD)
Glove wearing comfort	5 (1.2)
Sitting comfort	5.13 (2.03)
No disruption through the cable	3.88 (2.17)
Noticeability of vibrations	4.5 (0.93)
Not exhausted	3.75 (2.38)
Ease of learning the system	5.5 (0.03)
Ease of using the system	4.88 (1.13)
Expected improvement through exercise	6.5 (0.54)

assume that the combination of tactile and (non-ambiguous) visual cues could address this and further improve performance.

3.5 Discussion

Here, we will discuss findings with regards to the research questions.

RQ1: *How well can tactors be localized and differentiated across the hand and lower arm?*

We showed that users can reasonably well localize and differentiate cues. Especially interesting is the good performance of cues at the back of the hand, which performed about as well as the index finger (which is highly sensitive, in contrast to the back of the hand). This result is useful as the back of the hand can also be used for other purposes, like the provision of touch-driven events that can be coupled to guidance, e.g., touching a wall with the back of the hand while moving a grasped object.

RQ2: *How do users interpret tactile pose and motion patterns and what are their preferences?*

While tactors could be localized well, the interpretation of more complex stimuli – in particular direction – was not without errors. For several reasons, this is not surprising. First, a previous study also found that users interpret some patterns as either push or pull motions [380]. That is, the direction a pattern refers to may be interpreted differently by different users. While recognition of the dominant biomechanical action (e.g., flexion of the finger, or rotation of the hand) was reasonably high, we assume that personalizing patterns will result in a higher percentage of correct motions. In our study we observed that the efficiency of our system likely improves with learning. This means that over time, users will likely be able to interpret the patterns more easily and reliably. Previous work already noted that the level of abstraction likely influences learning rates [275], with lower abstraction resulting in quicker learning. Here, we assume that our guidance patterns are

at a medium to low abstraction level, as patterns are (a) easily localizable and (b) have good directional information that can be associated with a dominant biomechanical action. Learning will likely also be required to separate different types of feedback. Currently, we did not focus on touch cues, which would involve vibration at specific contact points. Depending on the context of operation, such vibration could be misunderstood, especially in cases where the user touches an object while receiving a pose change guidance pattern. While we could use different vibration modes to encode different events, the ability of users to actually differentiate among them requires further study, especially if we want to integrate pose and motion guidance methods with haptically supported selection and manipulation techniques.

RQ3: *How does tactile pose and motion guidance perform in a guided selection and manipulation task?*

With respect to hand guidance, we showed that our guidance methods can trigger motions and poses that can support finer-grained 3D selection and manipulation, independent of touch cues that may normally drive hand guidance. Our results extend previous tactile methods that only support general motion and pose guidance [72, 386], while our granularity is similar to EMS-based methods [253, 396], but without their disadvantages. Furthermore, while our current patterns only triggered start-to-end motions, e.g., to move the finger from a stretched to a bent configuration, guidance to intermediate stages is possible by running the pattern as long as needed. We might also use the strength of the feedback to provide a further indication about when to stop a motion, e.g., by making the feedback proportional to the error being made [97].

In all studies, we decoupled tactile cues from visual feedback. We deliberately did not use visualization aids to isolate the performance of our tactile guidance method, without interference from any given visualization method. However, related work, e.g., [24], has established that cross-modal feedback, such as the combination of visual and haptic cues, may also reduce error rates. We assume that tactile guidance methods can be visually enhanced to reduce ambiguities, based on visual and haptic stimuli integration theories [112]. The challenge is to do this in an unambiguous manner and to avoid visual conflicts. An example of visualization techniques that can aid to this respect are see-through visualization techniques [8], like transparency or cut-away. While cut-away techniques may limit spatial understanding as inter-object spatial relationships may be more difficult to understand (as objects are not rendered), transparency has been shown to maintain a

reasonable level of spatial understanding [8]. Such visualization could be combined with feedback co-located with the hand (instead of embedded in the scene) that provides motion and pose guidance. We assume co-located feedback – for example by overlaying a second hand / finger animation over the virtual hand to provide guidance – will likely have a higher success rate, to avoid ambiguity issues. However, this requires further study. Furthermore, coupling of feedback in multiple modalities may increase cognitive load [425]. In our case, the somatosensory system performs complex processes involving multiple brain areas to interpret the haptic cues, while cognitive load can vary based on different haptic properties [417]. Still, cognitive load likely decreases through learning, [466], an issue we plan to follow up in future work.

3.6 Conclusion

We presented a novel tactile approach to improve hand motor planning and action coordination in complex spatial 3D applications, by guiding hand motion and poses. Such guidance can be highly useful for 3D interaction, especially for applications that suffer from visual occlusions. Extending previous work on tactile cues that only worked on more general body motions, we showed that finer-grained pose and motion adjustments can be triggered. While learning and visual cues are expected to further improve performance, e.g., by reducing some interpretation errors, the results of our user studies already provide a solid basis for implementing tailored 3D selection and manipulation techniques that can be used in the frame of applications that require fine motor control, such as assembly training.

Future work includes full integration of guidance methods into 3D applications and study thereof. An important next step is the coupling of tactile guidance with hand co-located visual cues, which will likely lead to further improvements and a better understanding of the full potential of guidance support methods. We also want to investigate task chain variations to see when and how guidance feedback affects performance in complex situations, including tactile cues for guidance and touch-related events (collision, friction) in combination with visual feedback. For real training applications, guidance methods need to be coupled to behavior and ideal path analysis to dynamically guide users through, for example, training scenarios. To address the finger detection problems with a single sensor, hand tracking must be improved, e.g., via a multi-sensor setup [186]. Finally, we like to point out that due to the independence from visual cues, our system can be used in other domains, such as guiding visually-disabled people [241].

4 Non-Visual Guidance for Narrow Field of View Augmented Reality

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 [52]. 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 [44]. 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 [52]. 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 [52].

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 [45] 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 [44]. 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.xxxxxx

This chapter explores the potential of head-based audio and vibrotactile feedback to guide search and information localization in AR. For this purpose, we investigated the effects of audio-tactile guidance on search performance under sensory constraints. The sensory constraints investigated in this work involved depth perception, sensory thresholds, scene structure, and the FOV.

The results of this chapter provide the foundational framework for this thesis in both theoretical and practical terms and were informed by the outcomes presented in Chapters 2 and 3. In three user studies (n = 12), we showed that users could be guided with high accuracy using our novel audio-tactile feedback, even when perception was affected by a limited FOV. In study 1, we explored different auditory and tactile cues for encoding longitudinal, latitudinal, and distance information guidance without visual cues to determine which mode exhibited the highest accuracy. Study 2 examined audio-tactile cues for a visual search task. The results showed that for both studies, the mode that encoded latitude with audio and depth with vibrotactile pulse exhibited the highest accuracy and was preferred most by the users. Finally, study 3 revealed how audio-tactile cues could be used to determine the absolute longitudinal position and depth for localizing information. The findings of the three user studies revealed that non-visual, audio-tactile guidance was effectively supported search when perception was impaired by sensory constraints.

The material in this chapter originally appeared in: Marquardt, A., Trepkowski, C., Eibich, T. D., Maiero, J., & Kruijff, E. (2019). Non-Visual Cues for View Management in Narrow Field of View Augmented Reality Displays. *2019 IEEE International Symposium on Mixed and Augmented Reality*, 190-201. DOI: 10.1109/ISMAR.2019.000-3

4.1 Introduction

When increasing the density of information displayed in Augmented Reality (AR) applications, view management – the presentation and layout of augmentations – becomes challenging [406]. Conflicting visual cues and high density information can eventually lead to various degrees of sensory overload [247]. While visual attention is not necessarily affected by the number of distractors in visual search [449], the abundance of labels with potential visual conflicting cues can be difficult to process [221]. 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 [247]. 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 [57], 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 [447]. Generally, human reactions such as performance fluctuations or frustration are preceded by increasing cognitive load [247].

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 [116], 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 [224] 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 [104]. This may cause visual conflicts such as those related to visibility, legibility, depth ordering, scene distortion and object relationship issues [221]. 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 transferring some information towards another sensory channel [213]. 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 [260]. 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 another 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 [244], they are used infrequently and with few exceptions (e.g., the audio notes presented in [231]) 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 [368]. 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.

4.2 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

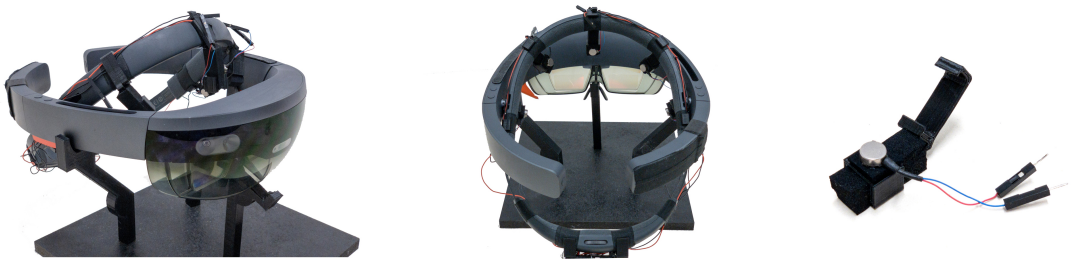


Fig. 4.1: The novel head-worn vibration feedback mechanism attached to the Microsoft HoloLens. In our studies, we used five vibration motors touching the forehead and temples. Three further vibration elements can be attached to the back of the head. The left and middle image show the arrangement of the 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.

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 [88] 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 [202]. 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.

4.3 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 [30], optimizing the layout and appearance of information. Among others, researchers have looked into label placement for size and position [16, 30] and depth-placed ordering [324, 325]. The appearance of labels has also been focused upon, for example in relation to foreground-background issues [138], or the legibility of text [129, 239]. While view management for wide FOV displays has found some interest [208, 224], with few exceptions (e.g., [345] and [406]) 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 [449], also in relation to more complex search tasks [307]. Visual search is affected by the types of features the search target and distractors elicit, which have been widely discussed in various theories [341], while specific aspects relevant for AR such as target eccentricity, orientation [62] and depth [282] have also been focused at. In general, search behavior has been studied widely, also specifically in AR by using eye tracking [127]. While visual cues such as the pop-out effects [154] have found reasonable wide application to draw the user's direction towards an item [347], also less obtrusive methods have been studied. Examples include subliminal cueing [327] and saliency modulation, also with specific application in AR [421]. Furthermore, the usage of specific pointers to targets, like arrows or attention tunnels have been studied [371]. 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, 17, 298], pathway guidance [125] and to increase situation awareness of pilots [121, 122]. Similar to that, windshield HUDs are becoming more common in cars, where navigation [206, 297] and attention factors [359, 402] have been studied. Furthermore, HUDs can be used to guide through assembly tasks and manufacturing,

[67, 397] and maintenance processes [165]. Finally, traditional visual overview methods like 3D Arrows and modern approaches such as EyeSee360 and 3D Radar [44, 144] 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 [242, 410], 3D selection [10, 265], target finding on mobile AR devices [3] and visual search tasks [235]. Of direct influence to our physical setup are the ring-based tactile guidance systems around the user's head [34, 88], and the top head/forehead system with a higher resolution tactor grid resembling an EEG setup [202]. Audio has also been used to guide visual search [292] and navigation [200]. 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% [279]. Audio cues have also been adapted in visual search tasks based on gaze direction [255]. Finally, within the frame of visual search tasks, cross-modal effects have been studied, including audio-tactile effects [175, 301] and conflicts between audio and visual cues [217]. 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 [98]. 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 [315]. Furthermore, this method can be applied for spatial data exploration and guidance [201, 374] and to support navigation tasks for visually impaired people in AR [200]. Another application area of sonification is the improvement of accuracy during the performance in high precision tasks [38, 352], e.g., in medical AR without obstructing the visual field with additional information.

With respect to our vibration methods, insights of [88] 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 user´s 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 [88] indicated that the overall mislocalization of a target was about 7% on the horizontal position and 4.5% on the vertical position.

4.4 System Approach and Implementation

Within this paper, we compare 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 Fig. 4.1. The extension consists of a row of 5 vibrotactors along the temples and the forehead in 45° intervals, schematically depicted in Fig. 4.2.

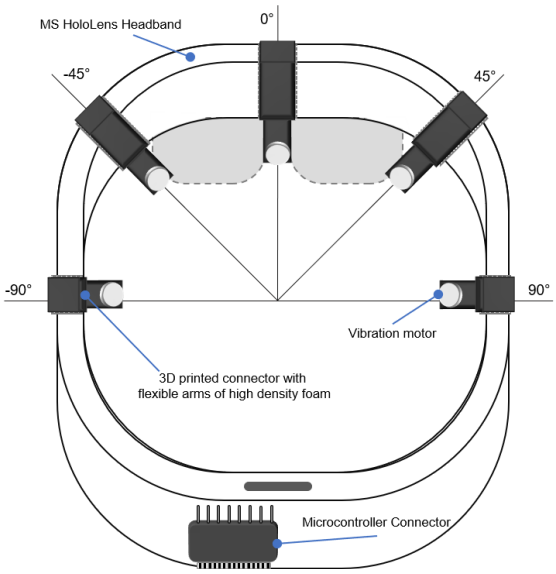


Fig. 4.2: Custom made tactor attachment on the Microsoft HoloLens headband with 5 vibrotactors placed in 45° intervals. A connector attached at the rear of the headband ensures a flexible an easy connection to the microcontroller.

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 Fig. 4.1, 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 audio-tactile feedback for multisensory view management. The methods reported in the next subsections are the result of pilot testing (see Subsection 4.4.4 “Pilot Study”).

4.4.1 Longitudinal Feedback

For Study 1 and 2 we reimplemented the metaphor for longitudinal feedback from [88], see Fig. 4.3, 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 vibrate. 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 [183].

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 [45, 207] and specific details on front-to-back confusion in [187]).

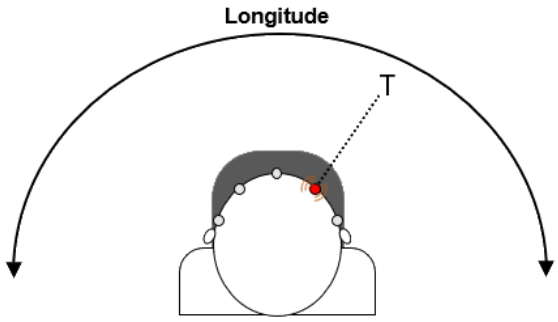


Fig. 4.3: Longitudinal feedback method by motor position adapted from [88].

4.4.2 Latitudinal Feedback

For latitudinal feedback, we created two different modes, namely vibrotactile (Fig. 4.4a) and auditory (Fig. 4.4b). Both methods use the adapted modulating function with a quadratic growth of [88]:

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

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

In case of vibrotactile feedback by intensity modulation (Fig. 4.4a), 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 [265].

In case of latitudinal audio feedback, 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 (see Fig. 4.4b). Unlike [88], 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) [158]. In contrast it has been shown that users are able to discriminate a maximum of only 9 levels of frequency on the skin [50]. Therefore, we expected it to perform better than the frequency adjustment of the vibrotactile cues on the users forehead.

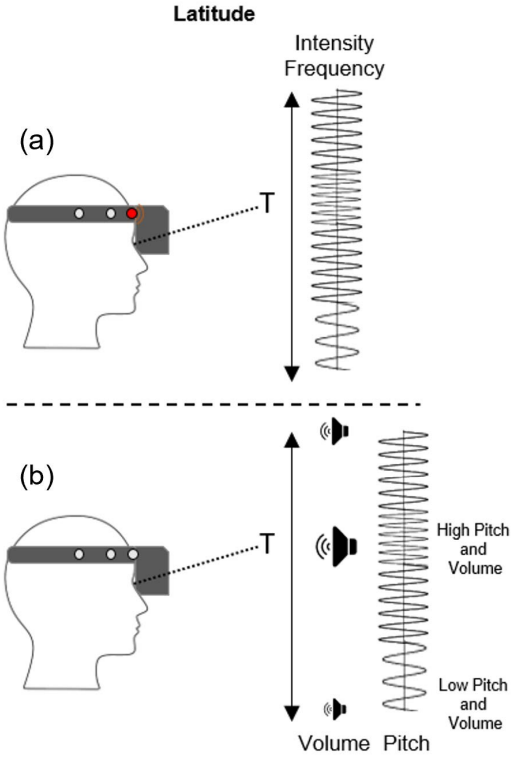


Fig. 4.4: Two variants of latitudinal feedback: (a) analogous to [88]; (b) shows the same feedback but the frequency modulation of the vibration motor is replaced with a sound. Volume and pitch adjusted by the target elevation.

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 [64]. Additionally, the volume level of the sound increases in a similar manner, with closer objects sounding stronger.

4.4.3 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 (Fig. 4.5a), and using a variable on/off pattern (Fig. 4.5b) 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 one and three 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 Fig. 4.5a. 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 Fig. 4.5b. 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 [265].

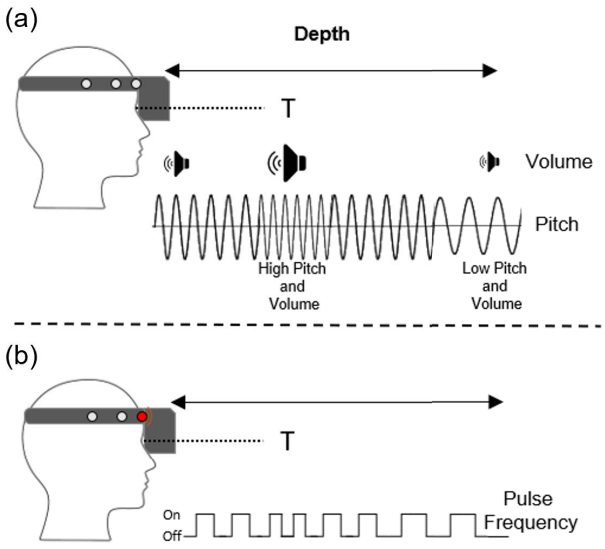
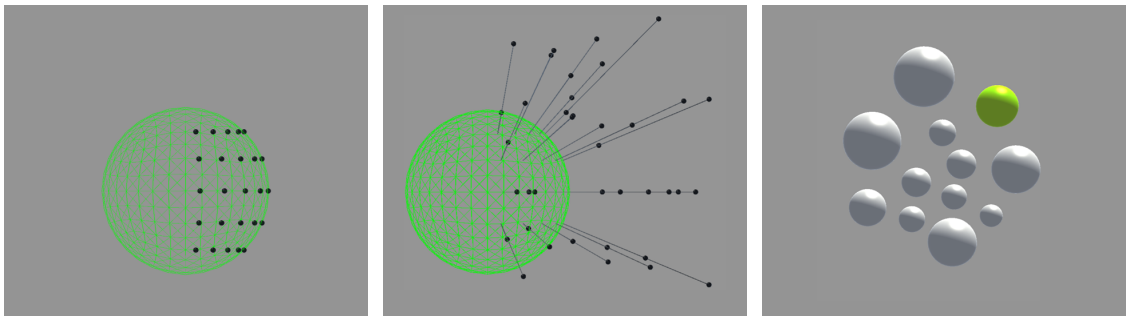


Fig. 4.5: Two variants of depth feedback: (a) analogous feedback like in Fig. 4.4a, adapted to depth. (b) pulse frequency adjustment depending on the target depth.



(a) Item population like in [88] on a hemisphere around the user (side view).

(b) Adding factor depth by assigning the items a random depth position between one and three 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.

Fig. 4.6: Item population, depth distribution, and item clustering for target items in the user studies.

4.4.4 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 Subsection 4.4.1 "Longitudinal Feedback" (direction indication by tactor position) since it already delivered good results in [88], 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 [237], especially using non-individualized HRTFs [436, 440]. 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 4.1.

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 [88] 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

Table 4.1: 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)

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 4.2. This also provided the advantage that users would not get strained or confused by the need to learn too many different feedback modes.

Table 4.2: Three isolated cue combination modes with longitudinal, latitudinal and depth feedback for Study 1 and 2. Each mode uses the longitudinal metaphor presented in Fig. 4.3.

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 [88], 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 Fig. 4.6c). 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 [88] 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.

4.5 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.

4.5.1 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 4.2 on accuracy performance). All possible target items were placed analogous to [88] around the user, in our case within the grid cells of a unit sphere´s surface with a radius of one meter on five elevation angles (45°, 22.5°, 0°, -22.5°, -45°), see Fig. 4.6a. 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 between one and three meters (Fig. 4.6b). As in the outcome of [88], 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.

4.5.2 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 4.2) on search time performance. Users were guided towards a visible cluster of spheres (see Fig. 4.6c) 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 [88], 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 4.2). 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.

4.5.3 Study 3 - Information Localization

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

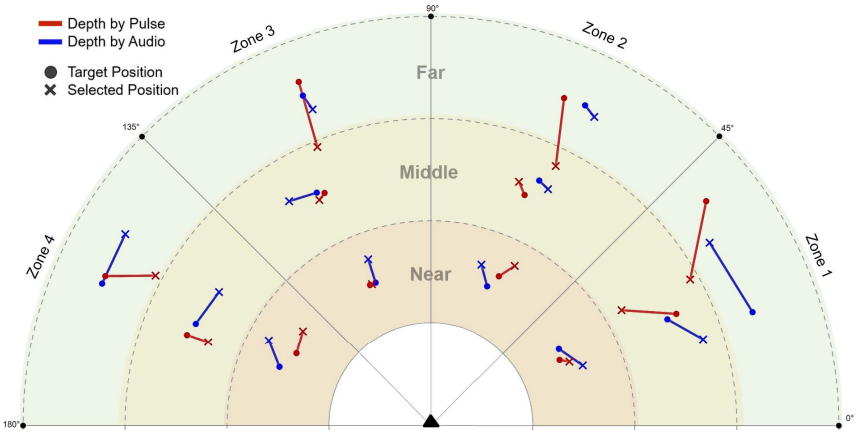


Fig. 4.7: Zones and depth areas in Study 3. The lines depict the average offset of all ratings per zone which each method. For illustration purposes we used the average target points per each zone as the start point of the lines. Interestingly, the ratings are not completely symmetrical. The near depth 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.

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 Fig. 4.7 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. 4.7), 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 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.

4.6 Results

Friedman test was used 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.

4.6.1 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 4.3 and Fig. 4.8.

Table 4.3: 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

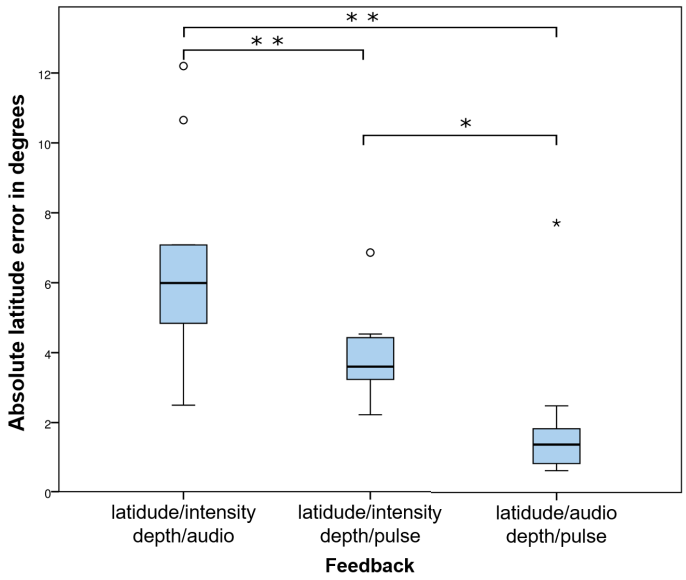


Fig. 4.8: Absolute latitude error in degrees by mode in Study 1.

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. 4.9). The latter indicates that showing the correct position of the target after each trial positively affected learning.

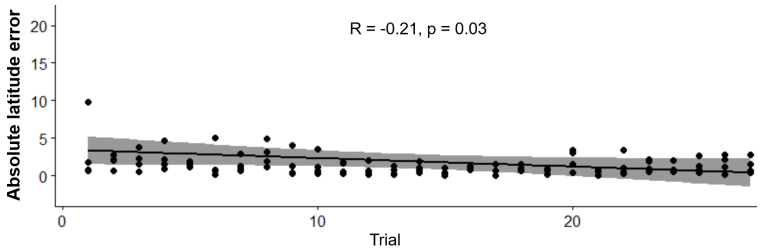


Fig. 4.9: Absolute latitude error in degrees by trial number in Study 1.

4.6.2 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. 4.10). 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 ($M = 21.2, IQR = 16.9 - 24.9, Z = 3.06, p = .002, r = 0.62$).

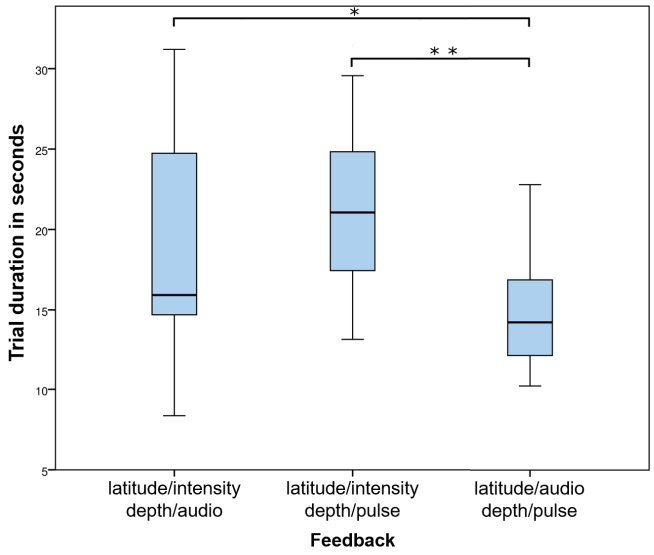
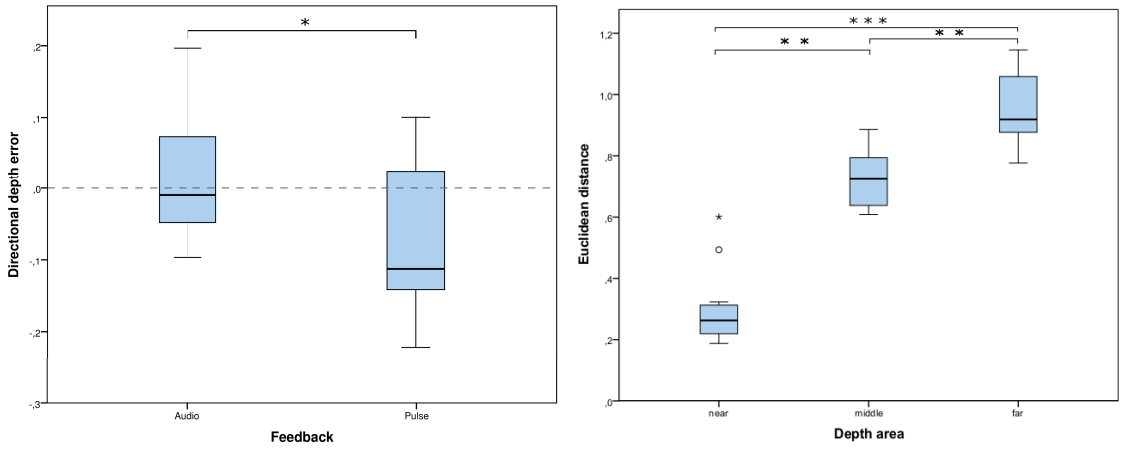


Fig. 4.10: Time to complete a trial in seconds by mode in Study 2.

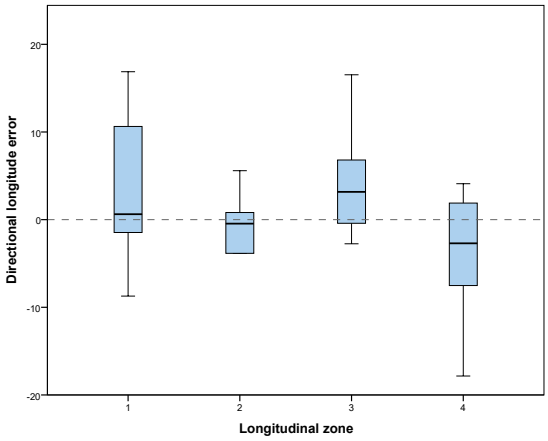
There was a marginal effect of the feedback mode on the rate of the correctly chosen cluster ($X^2(2) = 5.82, p = .055$ and absolute latitude error ($X^2(2) = 5.91, p = .052$). Post-hoc pairwise comparisons showed no significant differences regarding absolute latitude error. Descriptive values indicate 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. 4.12.



(a) Directional depth error by feedback mode.

(b) Euclidean distance by depth area.



(c) Directional longitude error by longitudinal zone.

Fig. 4.11: Directional errors and euclidean distances for Study 3. See Fig. 4.7 for depth areas and zones.

4.6.3 Information Localization

Seite 1

There was no difference between depth/audio and depth/pulse coding regarding the absolute longitude error, the euclidean distance, trial duration (see Table 4.4) 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 4.4) and the directional depth error ($Z = 2.59, p = .01, r = 0.75$, see Fig.4.11a). 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. 4.7) 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. 4.11b and 4.7). 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 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. 4.11c).

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.

4.6.4 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

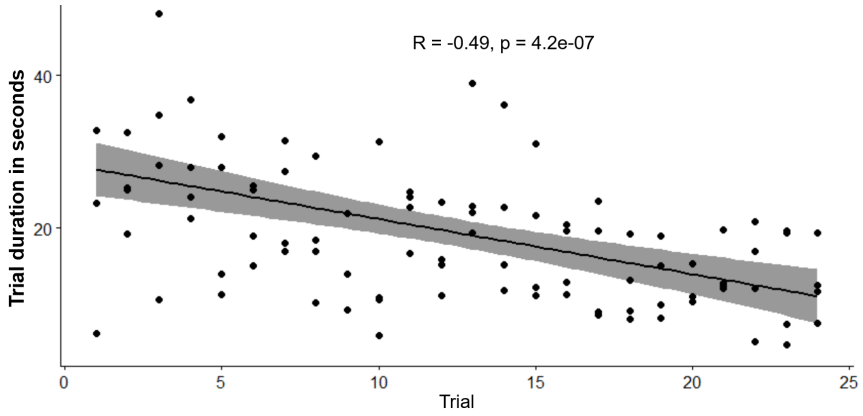


Fig. 4.12: Time to complete a trial over time with mode latitude/intensity & depth/audio in Study 2. The Fig. indicates the significant performance improvements caused by a learning effect that was only found in this mode combination.

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 4.4: Absolute errors in longitude and depth and the euclidean distance 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

4.6.5 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 4.5). 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).

Table 4.5: 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.

4.7 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 [88]. In their study, potential targets were always visible and hit rate was used as performance measure. Instead, 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 [202] as their setup is considerably different from ours by ways of resolution.

4.7.1 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 [88]. 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 a 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 [88] 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 [88] 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.

4.7.2 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.

4.7.3 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 [50, 158]. 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 [158] 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.

4.7.4 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 [221]. 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 [144]. Doing so, attention [171] and crossmodal issues [175] 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 [406], 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 [25], 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 [395]. 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 [384]. 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.

4.8 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 [88, 202]. 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 multi-sensory 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 [144] – 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 [202]). 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.

5 Comparing Non-Visual and Visual Guidance Methods

Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays

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

Abstract—Current augmented reality displays still have a very limited field of view compared to the human vision. In order to localize out-of-view objects, researchers have predominantly explored visual guidance approaches to visualize information in the limited (in-view) screen space. Unfortunately, visual conflicts like cluttering or occlusion of information often arise, which can lead to search performance issues and a decreased awareness about the physical environment. In this paper, we compare an innovative non-visual guidance approach based on audio-tactile cues with the state-of-the-art visual guidance technique EyeSee360 for localizing out-of-view objects in augmented reality displays with limited field of view. In our user study, we evaluate both guidance methods in terms of search performance and situation awareness. We show that although audio-tactile guidance is generally slower than the well-performing EyeSee360 in terms of search times, it is on a par regarding the hit rate. Even more so, the audio-tactile method provides a significant improvement in situation awareness compared to the visual approach.

Index Terms—Augmented Reality, view-management, guidance, audio-tactile cues, performance, situation awareness

1 Introduction

Locating virtual objects in augmented reality (AR) applications can be a challenging task. A major problem of many augmented reality displays is their relatively narrow field of view (FOV) that only covers a small part of human vision. The human visual system has a binocular FOV of about 210° horizontally and 150° vertically [40]. In comparison, the popular Microsoft HoloLens AR display has a FOV of 30°H and 17.5°V. Studies have shown that when the density of information increases, narrow FOV can negatively affect search performance [13]. Furthermore, conflicting visual cues can make it difficult to process and interpret stimuli [47] and can lead to a certain degree of sensory overload as human processing capacities are limited [56].

In this context, view-management techniques are dealing with the layout and appearance of augmentations [9]. However, designing effective view-management systems for narrow FOV displays is still an open issue in research. Depending on the application at hand, view management may need to handle both in-view and out-of-view information adequately. A major problem for narrow FOV displays is typically overlapping information (e.g., labels), where augmentations occlude each other and/or the reference object in the scene [23]. In-view labelling can aggravate the problem [48] as this method tries to place additional labels inside the limited FOV that refer to out-of-view objects. This can lead to visual conflicts that can cause visibility, legibility, depth ordering, scene distortion and object relationship issues [47]. With respect to solving problems related to out-of-view targets, researchers have focused on developing different so-called guidance approaches (see [12] and [79] for an overview). We can roughly differentiate between visual (e.g., [12, 32]) and non-visual guidance methods (e.g., [20, 42, 62]). Most research is directed towards visual methods [79]. Non-visual guidance methods typically look at reducing visual overload or conflicts by minimizing the number of stimuli in the visual sensory channel. To achieve this, sensory substitution - the transfer of information to a different sensory channel - can be used [58]. Sensory substitution is commonly used to overcome the limitations of blocked sensory channels, e.g., for people with visual disabilities [59]. In the context of dense information in narrow FOV displays, sensory substitution is believed to be a fruitful direction to improve user performance [62].

The effectiveness of the most common visual guidance techniques has been compared in a number of object search tasks [12]. Results indicated that EyeSee360 was performing very well against other well-established visual guidance methods [12, 33]. However, there is a lack of understanding how well non-visual guidance compare to their visual counterparts. To address this lack, the aim of this paper is to compare the performance of visual and non-visual guidance methods in the context of head-worn displays with narrow FOV. For this work, we used the currently widely used Microsoft HoloLens (first version with a diagonal FOV of about 35°) as a reference model. In our studies, we look into aspects like search performance, accuracy, cognitive load, and situational awareness (SA) of visual and non-visual guidance methods. We compare visual and non-visual guidance methods and investigate the above-mentioned aspects in three sub-studies. In contrast to previous research that mostly considered optimal laboratory conditions, we study real-world conditions more closely by examining the methods in a simulated real-world environment instead of relying on purely abstract use cases. We assumed that non-visual feedback can have a positive effect on both usability and performance in AR systems with a narrow FOV. We also expected that visual complexity can be reduced by transferring that visual information to another perceptual channel. This approach might be particularly useful in visual complex environments, as it can potentially lead to a reduction of visual workload [75] and, in direct relation, may increase situational awareness (SA).

1.1 Contributions

Through the results of our user studies, we present the following contributions that provide new insights into the effectiveness of non-visual guidance methods in comparison to a state-of-the-art visual technique. We do so in context of guidance in narrow FOV AR displays. The complete study was performed in virtual reality (VR), simulating an AR environment (for details, see Section 4.3). While comparing visual and non-visual guidance methods, we place a strong focus on SA. SA can be a fundamental factor in AR systems and considerations should be given for the usage of AR in real-world conditions [39]. Unfortunately, SA is frequently not taken into account sufficiently when addressing guidance in AR.

- In study part 1 we compared audio-tactile guidance with EyeSee360 during a simple object collection task in terms of general task performance. We showed that audio-tactile guidance could

- *Alexander Marquardt, Christina Trepkowski, Tom David Eibich, Jens Maiero, and Ernst Kruijff 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.*
- *Johannes Schöning is with University of Bremen. E-mail: schoening@uni-bremen.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.xxx/TVCG.201x.xxxxxx.

**These authors contributed equally to this work.*

This chapter focuses on the evaluation of the effectiveness of the multisensory guidance cues under sensory constraints. We compared our non-visual guidance approach developed in the previous Chapter with the state-of-the-art visual guidance technique EyeSee360. Both guidance approaches were examined under the influence of further sensory constraints, namely depth perception, disparity planes, sensory thresholds, scene structure, background features, sensory noise, and the FOV.

This work was extensively informed by the methodological framework and insights presented in Chapter 4. In three study parts (n = 16), we evaluated both guidance methods in terms of search performance and SA in narrow FOV AR displays. In part 1 of the study, we compared audio-tactile guidance with EyeSee360 for a simple object collection task. We showed that audio-tactile guidance, although slower, could compete with EyeSee360 in terms of hit rate. For part 2 of the study, the difficulty was increased by adding sensory noise and background features to the same task as in part 1. Increased task difficulty did not have a significant influence on search performance for both guidance methods. However, a subtle noticeability test in this substudy indicated that audio-tactile guidance tended to provide higher SA during search compared to visual guidance. Task difficulty was increased again in study part 3 by adding a visual secondary task. We showed that SA was significantly higher with audio-tactile guidance whereas task performance was not affected by the secondary task in either approach. The findings of this paper suggest that the use of audio-tactile guidance cues is beneficial in contexts that require improved SA and limited visual clutter.

The material in this chapter originally appeared in: Marquardt, A.*, Trepkowski, C.*, Eibich, T. D., Maiero, J., Kruijff, E., & Schöning, J. (2020). Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics, 26(12)*, 3389-3401. DOI: 10.1109/TVCG.2020.3023605. *Equal contribution.

5.1 Introduction

Locating virtual objects in augmented reality (AR) applications can be a challenging task. A major problem of many augmented reality displays is their relatively narrow field of view (FOV) that only covers a small part of human vision. The human visual system has a binocular FOV of about 210° horizontally and 150° vertically [196]. In comparison, the popular Microsoft HoloLens AR display has a FOV of 30°H and 17.5°V. Studies have shown that when the density of information increases, narrow FOV can negatively affect search performance [406]. Furthermore, conflicting visual cues can make it difficult to process and interpret stimuli [221] and can lead to a certain degree of sensory overload as human processing capacities are limited [247].

In this context, view-management techniques are dealing with the layout and appearance of augmentations [30]. However, designing effective view-management systems for narrow FOV displays is still an open issue in research. Depending on the application at hand, view management may need to handle both in-view and out-of-view information adequately. A major problem for narrow FOV displays is typically overlapping information (e.g., labels), where augmentations occlude each other and/or the reference object in the scene [104]. In-view labelling can aggravate the problem [224] as this method tries to place additional labels inside the limited FOV that refer to out-of-view objects. This can lead to visual conflicts that can cause visibility, legibility, depth ordering, scene distortion and object relationship issues [221]. With respect to solving problems related to out-of-view targets, researchers have focused on developing different so-called guidance approaches (see [44] and [356] for an overview). We can roughly differentiate between visual (e.g., [44, 146]) and non-visual guidance methods (e.g., [88, 202, 267]). Most research is directed towards visual methods [356]. Non-visual guidance methods typically look at reducing visual overload or conflicts by minimizing the number of stimuli in the visual sensory channel. To achieve this, sensory substitution - the transfer of information to a different sensory channel - can be used [252]. Sensory substitution is commonly used to overcome the limitations of blocked sensory channels, e.g., for people with visual disabilities [262]. In the context of dense information in narrow FOV displays, sensory substitution is believed to be a fruitful direction to improve user performance [267].

The effectiveness of the most common visual guidance techniques has been compared in a number of object search tasks [44]. Results indicated that EyeSee360 was performing very well against other well-established visual guidance methods [44, 147]. However,

there is a lack of understanding how well non-visual guidance compare to their visual counterparts. To address this lack, the aim of this paper is to compare the performance of visual and non-visual guidance methods in the context of head-worn displays with narrow FOV. For this work, we used the currently widely used Microsoft HoloLens (first version with a diagonal FOV of about 35°) as a reference model. In our studies, we look into aspects like search performance, accuracy, cognitive load, and situational awareness (SA) of visual and non-visual guidance methods. We compare visual and non-visual guidance methods and investigate the above-mentioned aspects in three sub-studies. In contrast to previous research that mostly considered optimal laboratory conditions, we study real-world conditions more closely by examining the methods in a simulated real-world environment instead of relying on purely abstract use cases. We assumed that non-visual feedback can have a positive effect on both usability and performance in AR systems with a narrow FOV. We also expected that visual complexity can be reduced by transferring that visual information to another perceptual channel. This approach might be particularly useful in visual complex environments, as it can potentially lead to a reduction of visual workload [343] and, in direct relation, may increase situational awareness (SA).

5.2 Contributions

Through the results of our user studies, we present the following contributions that provide new insights into the effectiveness of non-visual guidance methods in comparison to a state-of-the-art visual technique. We do so in context of guidance in narrow FOV AR displays. The complete study was performed in virtual reality (VR), simulating an AR environment (for details, see Section 5.5 “System and Implementation”). While comparing visual and non-visual guidance methods, we place a strong focus on SA. SA can be a fundamental factor in AR systems and considerations should be given for the usage of AR in real-world conditions [194]. Unfortunately, SA is frequently not taken into account sufficiently when addressing guidance in AR.

- In study part 1 we compared audio-tactile guidance with EyeSee360 during a simple object collection task in terms of general task performance. We showed that audio-tactile guidance could compete and even slightly exceed EyeSee360 regarding the hit rate. However, search times were considerably shorter for EyeSee360.
- In study part 2 we increased the difficulty by adding visual noise and optical flow to the same task as performed in study 1. Furthermore, a small noticeability test was

added to have a first indicator for SA. We showed that an increased task difficulty likely does not have an influence on search performance for both guidance methods. However, the noticeability test indicated already a notably higher SA for the audio-tactile mode.

- In study part 3 the task difficulty was increased again by adding a secondary task. The performance of the secondary task was also used to measure SA. We showed that SA was significantly higher with audio-tactile guidance while performance values of the object collection task (search times, hit rate) for both modes were not affected by the secondary task.

Summarizing, it has not been shown yet how non-visual guidance cues can compete with current visual guidance techniques. We do so by discussing performance measurements while solving an object collection task under different degrees of difficulty. Furthermore, we show how to improve SA in case audio-tactile guidance is used for the localization of out-of-view objects in AR.

5.3 Related Work

Our studies touch upon several fields of research, namely view management, visual and non-visual guidance methods, and situational awareness in AR, which we describe below.

5.3.1 View Management

Designing and optimizing the layout of information in view management methods have been researched over a longer period of time [30]. Studies so far have mainly focused on label placement for size and position [16, 30], depth-placed ordering [324, 325] and the appearance of labels (e.g., foreground-background issues [138] or the legibility of text [129, 239]). While in recent times some research has been done on view management for wide FOV displays [208, 224], not many researchers have focused on narrow FOV displays yet, except, e.g., [345, 406].

5.3.2 Narrow Field of View

Current-generation AR devices still suffer from a limited FOV. Limiting the FOV typically leads to various problems like perceptual and visuomotor performance decrements for real and virtual environments [25]. Even though most studies that focus on FOV limitations

were performed on virtual reality (VR) systems, it can be assumed that insights also apply to AR applications to a certain degree. Another intensively discussed issue is the consistent underestimation of distances for head mounted displays (HMD) with limited FOV in VR scenes [454] and for AR applications [394]. Dense information spaces in narrow FOV have also been shown to affect search performance negatively [406], while a decreased FOV can lead to a significant change in visual scan pattern and head movement, which may in turn also affect search performance [83, 387]. With respect to spatial awareness, it has been shown that FOV restrictions are degrading the abilities of developing spatial knowledge and navigation [5, 435]. Finally, a restricted FOV can result in decreased search performance [11] as well as selection performance [109].

5.3.3 Visual Guidance

With respect to visual guidance, effects like the pop-out effect [154] or attention guiding techniques [347, 433] have found some reasonable application. Less obtrusive methods like subliminal cueing [327] and saliency modulation in AR [421] have also been discussed. Furthermore, head-up displays (HUDs) are also widely used for guidance, e.g., in the aircraft sector, for basic navigation, flight information [17, 298] and pathway guidance [125]. Other common examples for guidance with visual aids are specific pointers to targets like arrows and attention tunnels [371], 3D arrows [147], radars and halos [44] or EyeSee360 [144]. The last method showed superior performance compared to five other visual guidance techniques in different scenarios regarding completion time, usability (SUS score) and workload [44].

5.3.4 Non-Visual Guidance

Non-visual guidance can be implemented in various ways. In terms of vibro-tactile cues, they can be used to direct navigation [242, 410], for 3D selection tasks [10, 265], for supporting pose and motion guidance [22, 266], and visual search tasks [235, 243]. In [267], we reported on different audio-tactile approaches that guide the user in 3D space. The used setup was specifically designed for AR displays with narrow FOV and is inspired by the ring-based tactile guidance systems of Oliveira et al. [88]. Similar head mounted tactile setups have been explored in a two dimensional manner (e.g., Haptic Radar [65], ProximityHat [34]) or as high-resolution tactor grid [202]. Alternative haptic feedback devices exist that can provide directional feedback towards the head, e.g., by using a

robot-arm attached to a HMD [443], however their applicability may be limited in AR systems.

Regarding auditory cues, research has looked at supporting visual search [292, 414] and navigation [200]. Studies showed that spatial auditory cues can improve search performance by up to around 25% [279]. Regarding visual search tasks, cross-modal effects have been researched for audio-tactile cues [212, 301] or conflicts between visual and auditory cues [217]. Sonification strategies also use auditory cues to inform or guide the user. It typically modulates sound attributes like pitch and loudness with respect to the presence of the auditory reference [98]. This metaphor can also be found in modern parking car systems, where the distance information is provided through a decreasing time interval between impulse tones [315].

5.3.5 Situational Awareness

Situational awareness describes the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [105]. AR technologies found a broad application in improving SA for diverse areas. The AR tool InfoSPOT [182] for example helps facility managers accessing required building information. SA is enhanced by overlaying device information on the view of the real environment. AR is also widely used in the aviation sector for pilots [121], military operations[75, 249], and driver assistance systems in cars [314, 246] to provide the user additional information, e.g., about incoming threads.

To measure SA, various techniques can be used (see [361] for an overview). SAGAT - a freeze probe technique - is one of the most common approaches to validate SA. On the other hand, measuring task-depending characteristics of the operator's performance is the probably simplest way to examine the impact of SA. Performance measures are non-intrusive as they are produced through the natural flow of the task and is used to indirectly measure SA [361].

5.3.6 Research Questions

The user study reported in this paper compares guidance performance of non-visual to visual cues under three different degrees of difficulty. In each study part users used guidance cues to identify a target among distractor objects. Task difficulty (from now on referred to as "task load") can be typically modulated by adding noise or including a secondary

task next to the main task [87]. During our experiment, task load is increased by adding visual noise, namely through a dynamic environment, and a secondary task. While the background was kept static and neutral in the first study part, the second and third study part were set in a vivid virtual city environment causing rich visual background noise and optical flow. In order to increase task load again in the third study part, users further had to perform a secondary task next to the guided search task. This allows us to examine the user's SA more closely [105].

These studies addressed our research questions, formulated as follows:

RQ_1: How well do non-visual guidance methods perform compared to visual guidance methods for a search task on different levels of task load (inferred by a static/dynamic environment and secondary task)?

H_1: We hypothesize that EyeSee360 will outperform audio-tactile guidance in low task load (static environment) conditions. On the other hand, we expect in the high task load (dynamic environment and secondary task) conditions a higher performance for the audio-tactile method because of the reduced workload and less visual clutter compared to EyeSee360.

RQ_2: Is there an effect of guidance method on situation awareness when a secondary task is included?

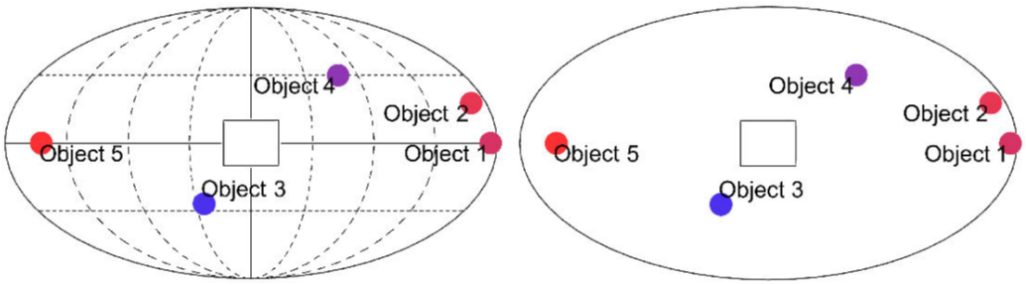
H_2: We hypothesize that the usage of EyeSee360 contributes to a lower SA compared to audio-tactile guidance. We expect this behavior because of a higher mental workload due to a higher density of visual information compressed inside a small FOV.

5.4 User Study

For visual guidance we used the EyeSee360 technique [144]. EyeSee360 was created for visualizing out-of-view objects in 360° around the user, depending on the user's orientation, and was improved over time in terms of reducing visual clutter and mental workload [145, 148]. Following, we will describe both methods in more detail. To provide non-visual cues, we used a modified version of our audio-tactile guidance interface reported in [267] and encoded latitude by audio and depth by vibration cues. Previously, we tested this cue combination against other non-visual audio-tactile feedback encodings and showed that it provides a superior performance regarding guidance accuracy and search time [267].

5.4.1 EyeSee360

The original EyeSee360 technique (see Fig. 5.1a) maps the 3D space to a 2D ellipse with a smaller rectangle in the central point. Colored dots (called proxies) are positioned in this 2D map inside the inner rectangle to indicate target locations of objects in the 3D space inside the user's FOV, while out-of-view objects are displayed inside the ellipse but outside the rectangle (see Fig. 5.1). The inner rectangle is sized so as not to occlude the user's focus. The horizontal line corresponds to the eye level of the user, the distance to this line indicates elevation level of the target. A proxy above this line indicates that the target is above eye level, a proxy beneath the line means that the target is located below. Distance to the vertical line indicates the longitudinal position of the target, which can be on the left or right side of the user. To illustrate the distance of the object, the proxy can take a color of a gradient from red (target is close) to blue (target is far away), as can be seen in Fig. 5.1. This heatmap-inspired coding is intended to make the interpretation of distances as intuitive as possible. The original version of EyeSee360 included helplines in addition to the horizontal and vertical line (Fig. 5.1a). However, we decided to use the improved variant with zero helplines only (Fig. 5.1b) as it has been shown to cause less distraction and resulted in a better search performance compared to the variant with helplines [148].



(a) All helplines. (b) No helplines.

Fig. 5.1: Out-of-view visualization with EyeSee360. (a) shows the initial method presented in [144]. (b) shows an improved variant of this method without helplines [148].

5.4.2 Audio-Tactile Guidance

Audio-tactile guidance encodes spatial information on longitude, latitude, and depth to guide the user to a position in the 3D space. Here, we briefly describe these encodings, for more detail, please refer to [267], as we basically replicated the methods reported therein. In the aforementioned paper, we investigated different approaches of non-visual guidance

in terms of performance, accuracy and information localization. These metaphors are partially adapted from Oliveira et al. [88]. For the purpose of this paper, we used the best-performing metaphor as reported in [267].

The user is informed about the relative position of the target on the longitude by the position of the vibrating tactor in the vibro-tactile setup (see Fig. 5.2 upper, system description in Section 5.5). If the target angular position was located horizontally between two tactor positions, both motors vibrated. Motor intensity of both motors was set in relation to the angular distance of the target. This was done to achieve an interpolation effect to indicate that a target lied in between the physical motor setup, similar to the phantom effect described in [183]. Once activated, the corresponding motors were running at a frequency from about 50 Hz up to 200 Hz, depending on the current angular distance of the target. These values were chosen as we previously showed that this feedback was clearly perceptible without being considered as disturbing [267]. If the head is turned towards the indicated direction (Fig. 5.2 lower), the vibration "wanders" with the head rotation until the feedback at the center of the forehead informs the user that the target is located directly in front of his view direction. In case the target angle temporarily lies above 90°or below -90°of the current head rotation, the corresponding outermost vibration motor keeps vibrating until the user rotates the head closer to the target direction.

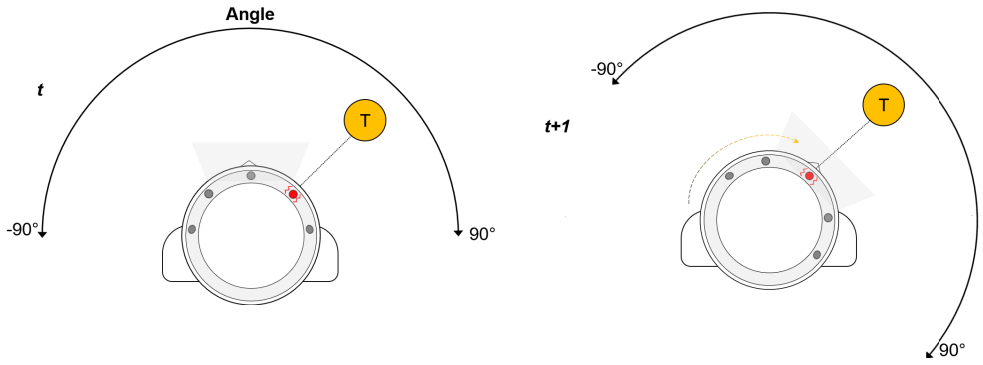


Fig. 5.2: Longitudinal encoding (top view). Initially (*time t*), the tactor position indicates the target direction. At time *t+1* the vibration feedback "wanders" with head rotation.

The latitude was provided by auditory feedback that used a modulating function with a quadratic growth, as this function has been demonstrated to be the best working one in conjunction with latitudinal encoding [88]. The modulating function adjusts the pitch and the volume of the sound source depending on the difference between the user viewing angle and the target elevation level as shown in Fig. 5.3. If the viewing angle is far from

target elevation, the auditory feedback is low for both volume and pitch, starting from about 300 Hz. As soon as the viewing angle gets closer to target elevation, pitch and volume increased to indicate the rapprochement on the latitudinal plane. Pitch and volume is the highest at about 1300 Hz if the viewing angle corresponds right to target elevation level to inform the user that the correct elevation angle is spotted. The mid-range spectrum from 300 ~ 1300 Hz was chosen as the human auditory system is particularly attuned in this range and frequency discrimination works sufficiently good. Higher frequencies however can sometimes be perceived as annoying or even painful over time [64].

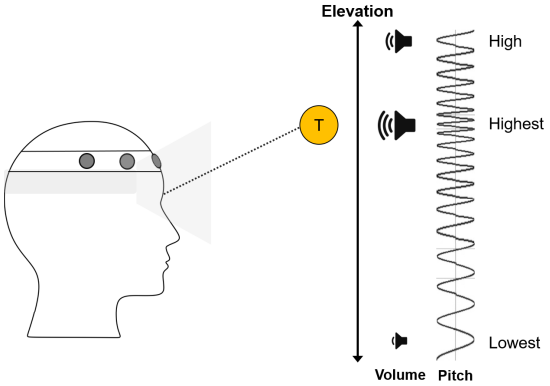


Fig. 5.3: Latitudinal encoding by auditory cues. Pitch and volume gets adjusted by the viewing angle of the user.

To provide information about target depth (distance), we used the implementation for target localization *with absolute depth feedback* [267]. This method uses the currently selected motor from longitudinal encoding (see Fig. 5.2) and applies a variable on/off pattern for activating the vibration motors - hereafter referred to as *pulse feedback* (see Fig. 5.4). Pulse Feedback is inspired by commonly used car parking metaphors that encode distance information through a decreasing time interval between impulse tones [315], but in a vibro-tactile manner. This makes pulse feedback easy to understand for most people since it is a commonly used real-world metaphor. One pulse is described as the time the motor is turned on and off again for a specific interval. These on/off times have always the same length and are set to periods from 100ms up to 500ms. A long pulse of 500ms would indicate that the target lies very far away from the user while a very short pulse of 100ms length signals that the target is positioned right in front of the user. The maximum pulse speed of 100ms was chosen to comply with the physical restrictions of the vibration motors, e.g., overcoming motor inertia and braking time without provoking interferences [265].

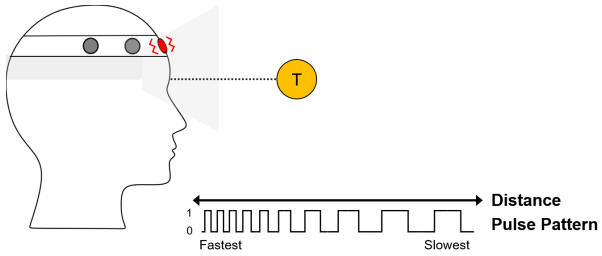
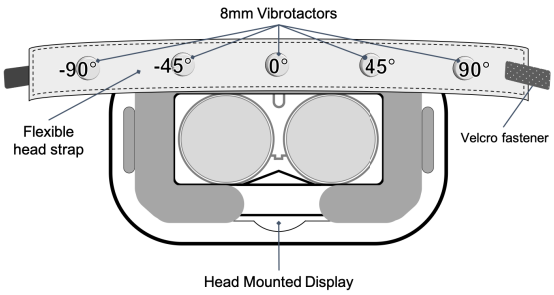


Fig. 5.4: Depth encoding by pulse feedback. Pulse duration gets adjusted by target depth.

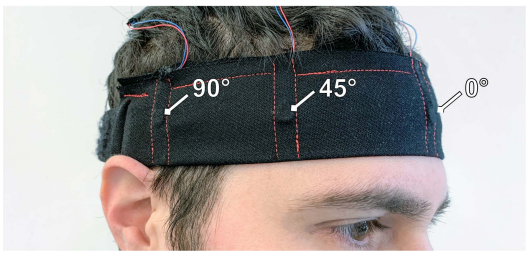
5.5 System and Implementation

In this work, we compare visual and audio-tactile guidance for AR applications. However, for the user studies we used virtual environments to ensure the same preconditions (e.g., lightning, visual and auditory noise) and to allow an overall comparability between the various study parts [342]. As such, we follow a similar approach as reported in [194]. For this, the FOV of the Microsoft HoloLens as current state-of-the-art AR headset is simulated in VR. This was achieved by placing a virtual display of about 35°(diagonal) size 3cm in front of the user's eyes. This display used a semitransparent glass-like material in order to gain the impression of using an actual AR device (compare to [347]). Virtual augmentations are only visible for the user inside that simulated AR FOV, as can be seen on Fig. 5.6 and Fig. 5.9a and 5.9b. To provide tactile cues, we created an extension that is usable in combination with various AR/VR HMDs. In contrast to our previous system [267], it consists of a headband which is made out of stretchable, comfortable-to-wear cotton instead of a solution integrated in the headset. In this headband, 5 vibrotactors are placed along the temples and the forehead in 45° intervals (see scheme in Fig. 5.5a. We used Precision Microdrives 8mm vibration motors (2mm type, model number 308-107). These motors were placed into sewn pockets, so both sides of the motors are protected by fabric, as can be seen in Fig. 5.5b. By this design, we avoid direct skin contact and uncomfortable pressure against the forehead while still maintaining clearly noticeable vibration feedback, even when it is worn below a HMD.

The system was implemented using Unity 2019.2. We used the HTC VIVE Pro VR headset including VIVE Pro controller as VR platform. Participants performed the study in a laboratory room in seated position on a rotating chair which is adjusted to a comfortable position beforehand. Spatial audio was enabled by the Steam Audio Spatializer plugin, using the integrated earphones of the HMD. The vibrotactors were controlled by a Raspberry Pi 3 Model B+ running a Python-based version of Open Sound Control to communicate with the Unity App.



(a) Scheme of the vibro-tactile interface in combination with a HMD.



(b) Wearing the interface. Vibrotactors are hidden in the sewn pockets.

Fig. 5.5: Custom made head strap attached with 5 vibrotactors placed in 45°intervals. The interface can be worn below a HMD.

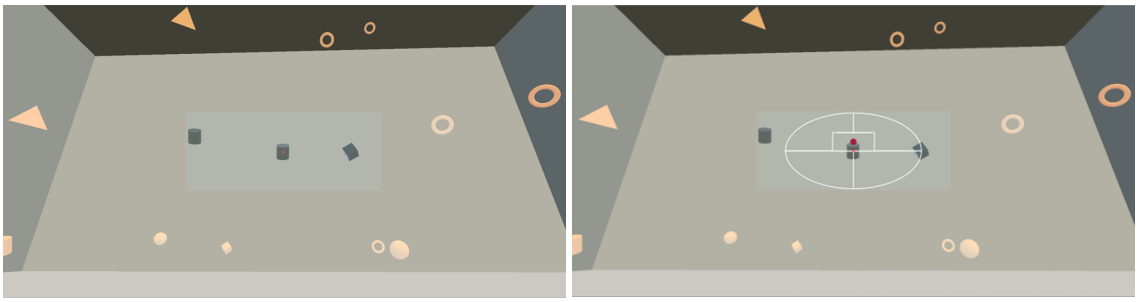
To model the visual noise conditions (see details in the next subsection), for study part 2 and 3, virtual pedestrians were created with a random appearance by the UMA 2 package (Unity Multipurpose Avatar). For the car traffic, from a pool of 8 different looking cars, models were distributed in the scene. To simulate simple crowd and car traffic movement, NavMesh Agent behavior managed a continuous movement over random predefined paths in the scene. Furthermore, ambient city sound effects are used to enhance immersion and to create additional auditory noise.

5.5.1 Study Design

Both guidance methods were compared in VR in three study parts to examine how they perform under different levels of task load in a fully controlled environment.

In each study part and trial the user had to identify a target among distractor objects that could not be differentiated by their appearance or position alone. All objects took a random shape of one of five primitives (sphere, cylinder, cube, pyramid, ring) with equal size. The primitives ultimately represented locations and objects within an urban environment. Therefore, they were colored in various shades of colors that are supposed to appear predominantly in urban surroundings. For this, we analyzed a static image of the city scene in study part 2 and 3 and extracted four independent color clusters (see Fig. 5.7a and 5.7b) by using k-means clustering. For the experiments, the primitives of the scene were randomly given one of the color of the resulting clusters (see Fig. 5.7c).

For each trial, 40 objects (1 target and 39 distractors) were distributed in the scene. To spatially distribute the targets around the user and to prevent them from overlapping each other, a virtual spherical grid is placed on the user´s position, similar to [267]. The spherical grid contains rows and columns, describing the angular distances to the user.



(a) Study part 1 using audio-tactile guidance. (b) Study part 1 using EyeSee360.

Fig. 5.6: Guidance methods in comparison without any visual distractors in study part 1. The center region shows the simulated AR display with the according guidance method. Out-of-view objects were visualized in semi-transparent orange color and were not visible outside the simulated AR display during the experiment.

We used three rows as elevation angles (on 0° respectively eye level, 22.5°, 45°) and ten columns along 180° (when looking straight ahead, the user was facing 90°). Initially, the objects are all placed in the center of each used row/column combination. Afterwards, each object was given a minor random horizontal and vertical offset and set to a random distance of 15-30 meters to the user to create different depth levels. We did not include further initial target elevation angles below 0° due to physical limitations [88, 267] and to prevent that an object would be occluded by the ground level. During the studies only the distributed primitives without the spherical grid were visible to the user. This setup also ensured that items were not occluded by each other or by any objects of the environment. Although there was no city environment in study part 1, we kept the depth range the same between studies for comparability reasons. The user was sitting on a swivel chair throughout the study and was not supposed to stand up or walk. To search the objects the user rotated the head (±90° left/right and up to 45° up) or turned the body on the stool. The user was always shown a crosshair in both guidance modes in the center of the display to select the virtual target. To select a target, the user had to orient the head towards the target object and place the crosshair over it and could then press the trigger of the controller for confirmation.

In *study part 1* there was no background noise and no secondary task. It was set in a 3D space with uniformly gray floor, walls and ceiling (see Fig. 5.6). A light source was also included in the scene to create an impression of depth. In *study part 2 and 3* the target had to be found under conditions with background noise. Objects (distractors and target) were generated in the same way as in study part 1, but were located in a busy city environment. In the VR setting the user was standing in front of a broad street and was

(a) Static input image. (b) Resulting cluster partitions.

Cluster	Pixels	RGB
1.	40.25%	35 35 33
2.	38.80%	64 65 63
3.	14.20%	131 138 139
4.	6.74%	30 77 41

(c) Color clusters values.

Fig. 5.7: A static image (a) of the environment used for study part 2 and 3 is taken as input for color analysis. Four cluster partitions (b) are extracted from the input and the resulting values (c) are used to color the target and distractor objects for the user study.

facing the opposite roadside. To create areas with increased optical flow at the 0° elevation level, cars were moving fast from left to right on the street and vice versa, while people walked around the pedestrian walkway in front of the user. To mimic real-world conditions, targets and distractors could not be occluded by buildings, but could be partially (and briefly be) occluded by pedestrians and cars. Horizontal optical flow between 22.5° and 45° was realized by recurrent wind gusts that transported small (visible) particles.

We added further distractors and minor optical flow for study 2 and 3 in the form of flying birds. The birds appeared in irregular intervals (every 12-17 seconds in study part 2 and 15-20 seconds in study part 3) on a path between the target elevation levels on 11.25° and 33.75°. The chosen path of the bird was depended on the current target elevation. If the target elevation was set on 0°, the bird was flying on 11.25°. If the target was on 45°, the bird was flying on 33.75°. Finally, if the target was placed on 22.5°, one of the two paths was chosen randomly. We did this to ensure the user always had the possibility to notice the bird during the search task. One bird was always present at the same time, visible for about 12 seconds. It followed a sinusoidal trajectory around the user from left-to-right

or right-to-left on the selected elevation (see Fig. 5.9c). If the user selected a possible target object while a bird was already flying in the scene, the bird adapted its elevation according to the new target object. Next to creating additional optical flow to the scene, the flying birds address two issues related to SA, namely noticeability and performance in a dual task condition. We focused on general perception (noticeability) in the first half of study part 2. For this, we let the birds fly more frequent and at a closer distance to the user at about 12 meters to make them clearly recognizable. Participants started every study part either with the visual or with the audio-tactile guidance method. Afterwards, they repeated the same object collection task with the other method. To receive an impression about the general perception of the environment, we asked every user after finishing the first mode of study part 2 which movable object was noticed in the scene. Prior to the study, participants were not explicitly advised to pay attention to their environment. The general perception was achieved in case the user indicated that he noticed the bird. With regards to measuring dual task performance we also included a secondary task next to the object collection task in study part 3. Here, we let the birds fly less frequent (every 15-20 seconds) and further away at about 20 meters to make them less obvious, yet still well visible for the user. The performance of the secondary task was primarily measured by the number of correctly detected birds during the regular object collection task with each of the two guidance methods. We also measured how often and long the bird was visible in the total FOV of the HMD (not the simulated AR FOV).

5.5.2 Procedure

Participants were recruited via a university mailing list and received a 10 Euro voucher as a reward for participation. We employed a 2x3 within-subject design to examine the effect of factors guidance feedback (visual versus audio-tactile) and task load (no noise, noise, noise and secondary task) on search time performance and hits/errors. Study parts 1 to 3 were always completed in ascending order as difficulty increased from study part 1 to 3. It was intended users got used to the guidance feedback when first noise and secondly an additional task was added to the search task. Users had to complete 30 training trials in total, ten before performing each study part. The task during the training trials was identical to the task of the performance session, except that the user had no time limit to find the targets in order to understand how the guidance methods work. Within each study part, guidance feedback was tested block-wise: First all trials with one guidance method were completed, then all trials with the other guidance method followed: (Mode A

→Mode B) or vice versa (Mode B→Mode A). Therefore, for three study parts, there are 2^3 possible feedback orders to perform the complete experiment that were balanced across participants. At the beginning of each study part a fixation point was shown to ensure the correct starting position of the user. As soon as the trial started, the guidance feedback was provided depending on the current condition to inform the user about the target location. The user could select the target by placing the cursor on an object and pressing the trigger to finish the trial. A red "x" was used as cursor, placed in the center of the simulated AR FOV (see Fig. 5.6 and Fig. 5.9a and 5.9b). This shape and color was used to be clearly visible in both visual and non-visual guidance mode. After confirming the selected target by pressing the trigger, the next trial started automatically, making it a continuous object collection task. The procedure for study part 1 and 2 was the same, which only differed with respect to the background.

In addition to the object collection task, the user was supposed to do a secondary task in study part 3. A bird in either red, black or blue color appeared in the scene, flying from one side of the street around the user to the other side (shown in Fig. 5.9c). The secondary task was about to react as quickly as possible by pressing a button as soon as the bird was spotted. The bird had to be visible inside the user's total HMD FOV while the button was pressed to be counted as hit. This enabled the user to select a target in the search task and to indicate the discovery of the bird at the same time. The three main study parts took 30 minutes (10 minutes for each part - 5 minutes with each guidance method). Including introduction, training and filling out the questionnaire, the whole study took 45 minutes.

5.6 Results

16 users (4 females), aged between 19 and 60 years ($M = 29.1, SD = 9.2$) took part in our study. The majority of participants played video games daily (50%) or weekly (31.3%) and indicated that the gaming console and the computer were their most often used mediums (37.5 % each) followed by the smartphone (18.8%). Regarding the experience with AR glasses 43.8% stated that they were using them sometimes.

A 2x3 repeated measures ANOVA was used to analyze the effect of task load (no noise, with noise, with noise and secondary task) and mode (EyeSee360, audio-tactile) on hit rate (hits/trials), each absolute and signed row, column total errors and errors per trial, trial duration and total number of trials. Greenhouse-Geisser correction was applied

when necessary. Row error was computed as the difference between the row of the chosen object and the row of the actual target on the spherical grid (see Subsection 5.5.1 "Study Design"). Column error was computed analogously. We further used Pearson correlation to analyze the association of the target distance and performance measures. We assumed that identifying and selecting targets that are far away and thus look smaller, could be more difficult and made a separate analysis accordingly.

5.6.1 Performance, Noise and Guidance Mode

In the following, please note that when we refer to task load, this relates to the load inherent to the task itself, while we refer to workload as the cognitive demand on the user side. The factor task load did not affect hit rate as neither background noise nor the secondary task did lead to performance decrements here ($p = .103$). Hit rate was consistently high with mean values ranging from 0.93 to 0.96. Guidance mode, on the other hand, significantly affected hit rate. Even though both modes facilitated hit rates above 0.9, mean hit rate of the audio-tactile guidance turned out be a significant 3% higher compared to the EyeSee360 technique, $F(1,15) = 8.45, p = .011, \eta_p^2 = .36$ (see Table 5.1). Trial duration was further affected by both, mode ($F(1,15) = 84.72, p < .001, \eta_p^2 = .85$) and the task ($F(1.1,17.1) = 6.3, p = .019, \eta_p^2 = .296$) and marginally by their interaction ($F(1.2,18) = 4.01, p = .054, \eta_p^2 = .211$).

Table 5.1: Mean values and standard errors of the hit rate which has been affected by the interface but not by task load.

Interface	Hit rate
EyeSee360	0.93 (0.02)*
audio-tactile	0.96 (0.01)*
Task load	
No noise	0.93 (0.02)
Noise	0.94 (0.01)
Noise + 2nd task	0.96 (0.01)

* p <.05

Main effects analysis showed that in each study part, trial duration was longer with the audio-tactile mode than with EyeSee360 ($p < .001$). Furthermore we found a trend that trial duration decreased slightly in the EyeSee360 condition from study part 1 (no noise) to 3 (noise and dual task) ($p = .062$), while with the audio-tactile guidance trial duration decreased from part 1 to 2 (noise) ($p = .038$) (see Fig. 5.8). The figure also shows that several values deviate upwards. We assume that these outliers can result from different

factors: 1) Selection difficulties, 2) target locations where the background color was more similar to the target and 3) targets which were particularly close to distractors. As we did not log these variables, it is not possible to clearly trace it back. However, some of these aspects will be considered in the upcoming discussion.

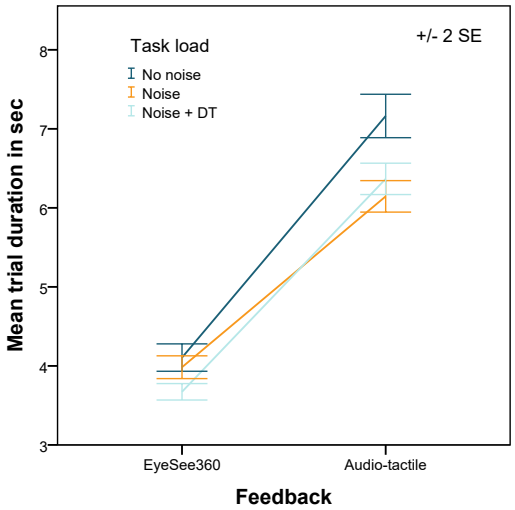


Fig. 5.8: Trial duration in seconds by task load and guidance method. Trial duration was significantly longer with the audio-tactile mode in each task at p level <.001.

5.6.2 Effect of Target Distance

We computed the euclidean distance from the user's viewpoint to the target position for each trial. Following, we analyzed the correlation between the distance, trial duration and hits in both guidance conditions. Only in the EyeSee360 condition there was a significant positive correlation between the distance and trial duration ($r(3662) = 0.135, p < .001$), and a negative correlation between distance and hits ($r(3662) = -0.047, p = .005$). That is, when using EyeSee360, the further away the target was, the longer the participants took and the fewer hits they made. Correlations were not significant in the audio-tactile condition. We also categorized data by near and far target distance and included distance as two-level factor in the ANOVA model. As targets were placed in a random depth between 15 to 30 meters, targets below 22.5 meters are classified as near distance, everything above as far distance targets. The repeated measures analysis shows a significant influence of distance (near/far) on trial duration, $F(1,15) = 31.7, p < .001, \eta_p^2 = .848$. Users generally needed a little more time when the target was located in the far area ($M = 6s, SE = 0.51$) compared to the near area ($M = 5.5s, SE = 0.46$). There was also a marginally significant interaction of guidance and distance on trial duration, $F(1,15) = 3.48, p = .082, \eta_p^2 = .188$: Main effects

Seite 1

(a) Study part 2 and 3 using EyeSee360.

(b) Study part 2 and 3 using audio-tactile guidance.

(c) Bird for SA measures used in study part 2 and 3.

Fig. 5.9: Busy city environment that is used for study part 2 and 3 to create visual noise and optical flow. The same object collection task is required to solve with EyeSee360 (a) and audio-tactile guidance (b). To measure SA in study part 3, the user has to react to a bird flying through the scene as secondary task (c). The dotted line visualizes an exemplary route of the bird. Note that out-of-view objects were visualized in semi-transparent orange color and were not visible outside the simulated AR display during the experiment.

analysis showed that only with EyeSee360 users needed more time for distant compared to near targets ($far: M = 4.8s(SE = 0.48)$, $near: M = 4s(SE = 0.32)$, $p = .001$). In the audio-tactile condition performance was similar for near ($M = 7.1s, SE = 0.61$) and far targets ($M = 7.3s, SE = 0.59$). Regarding hit rate, distance and the guidance method showed a marginally significant interaction effect, $F(1,15) = 4.05, p = .062, \eta_p^2 = .213$. Main effects analysis revealed that when comparing the performance between guidance methods for near and far distance targets separately, EyeSee360 and the audio-tactile technique differed only at the far target level: Hit rate was significantly higher with audio-tactile guidance ($M = 0.96, SE = .011$) than with EyeSee360 ($M = 0.92, SE = .017$), $p = .006$. That is, in case of far targets the audio-tactile guidance performed 4.2% better than EyeSee360. At the near distance level hit rates were also high for both feedback modes (EyeSee, M = 0.94, SE = 0.02, audio-tactile, M = 0.96, SE = 0.01) but did not differ significantly ($M = 0.92, SE = .017, p = .138$). When comparing the guidance methods at both distance levels, EyeSee360 had shorter search times at each level ($p < .001$): At the far target level, users needed 4.8 seconds on average ($SE = 0.48$) and were 34% faster than with

the audio-tactile mode ($7.3s, SE = 0.59$). At the near target level the EyeSee360 mode ($4s, SE = 0.32$) showed a 44% shorter mean search time than audio-tactile guidance ($7.1s, SE = 0.61$).

5.6.3 Secondary Task

After having finished the first block of trials with one mode in study part 2, users were asked which moving elements in the scene they had noticed. Users were not previously advised to pay special attention to the background. As the question could only be asked one time, a t-test for independent samples had to be performed to compare two groups of participants as half of them started with EyeSee360 and the other half with the audio-tactile mode. In the audio-tactile group, seven out of eight users noticed birds in the background ($M = 0.87, SD = 0.35$), but only two of eight with EyeSee360 ($M = 0,25, SD = 0.46$), $t(14) = 3.04, p = .009$. To further analyze how the mode affected the detection of the bird in the background we conducted t-tests for dependent variables in study part 3, where the secondary task was to press a button when the bird was noticed. In case the assumption of normality distribution was not met, the Wilcoxon signed rank test was used as non-parametric analysis. Mean values and standard errors are summarized in Table 5.2.

Table 5.2: Mean values and standard errors of bird performance measures for both guidance methods in study part 3.

Mode	Total time in FOV in s	Number of FOV entries	Number of correct detections	Misses
EyeSee360	2.5 (0.8)***	1.2 (0.1)*	13.7 (3.1)**	3.1 (3.2)**
Audio-tactile	1.8 (0.6)***	1.1 (0.1)*	16.4 (1.4)**	0.6 (1.1)**

* p <.05, ** p <.01, *** p <.001.

Users noticed the bird 28% faster when using the audio-tactile guidance mode than with EyeSee360. The total time the bird spent in the total HMD FOV until it was noticed was significantly lower in the audio-tactile condition ($t(15) = 5.28, p < .001$). Also, the time from the last FOV entry of the bird until it was found was significantly lower for the audio-tactile mode ($audio-tactile : M = 1.67, SE = 0.51, EyeSee360 : M = 2.16, SE = 0.17, t(15) = 4.64, p < .001$) as well as the average number of FOV entries of the bird per trial, $t(15) = 2.5, p = .024$. In addition, the mean number of detected birds was higher in the audio-tactile condition than with EyeSee360 , $Z = 3.06, p = .002$. The overall error

(misses and false-detections) was significantly higher for EyeSee360 ($M = 3.38, SE = 3.24$) than with the audio-tactile mode ($M = 0.81, SE = 1.22$), which could be attributed to the misses. Their number was higher in the visual condition than in the audio-tactile one ($Z = 3.21, p = .001$) while the number of false-detections did not differ significantly between conditions. Mean values and standard errors are displayed in Table 5.2. We further analyzed potential correlations between the performance of the search task and the performance of the secondary task, namely time until the bird was found. Regarding the audio-tactile guidance method, there was no correlation between primary and secondary task performance measures. When being guided by EyeSee360 a higher hit rate in the search task was associated with a faster detection of the bird after it entered the FOV ($r = -.656, p = .006$), the total time the bird spent in HMD FOV until it was noticed ($r = -.536, p = .032$) and bird detections ($r = .745, p = .001$).

Table 5.3: Significant differences between questionnaire ratings about distractors and task performance for study part 3.

	EyeSee360	Audio-tactile
Feeling disturbed by moving objects	4.9 (3.2)	4.1 (2.8)*
Fast secondary task performance	6.2 (2.3)	7.4 (2.3)*
Precise secondary task performance	5.8 (2.5)	7.1 (2.6)**
Concentration on secondary task	5.9 (2.4)	7.8 (2.4)*
Ease of judging the vertical position	9.4 (0.9)*	8.1 (2.2)

* p <.05, ** p <.01

5.6.4 Questionnaire Ratings

With regards to cognitive measures, a 2 x 3 repeated measures ANOVA was used to analyze the effect of task load and guidance method on workload through overall (raw) NASA TLX rating scores and on subscales. The overall NASA TLX score ranged from 0 to 100, ratings on a subscale from 1 to 21. Task load showed a significant effect on the overall NASA TLX score, $F(2,30) = 12.11, p < .001, \eta_p^2 = .447$. It was significantly lower in study part 1 compared to part 2 ($p = .001$) and to part 3 ($p = .001$). Regarding the analysis of subscales, task load affected mental demand ($F(1.43,217.67) = 6.5, p = .011, \eta_p^2 = .3$), marginally physical demand ($F(2,30) = 3.18, p = .056, \eta_p^2 = .175$) and performance ($F(2,30) = 6.19, p = .006, \eta_p^2 = .292$). Post-hoc comparisons revealed significantly higher ratings for mental demand for study part 3 compared to part 1 ($p = .019$) and higher ratings on the performance subscale in study part 1 than in 3 ($p = .012$). The effort subscale was not affected by neither task load nor guidance method. In contrast, frustration subscale

was affected by both, task load ($F(2,30) = 6.18, p = .006, \eta_p^2 = .292$) and guidance method ($F(1,15) = 4.34, p = .055, \eta_p^2 = .23$). Frustration was higher in study part 3 compared to part 1 ($p = .033$) (see Fig. 5.10) and higher with EyeSee360 ($M = 7.4, SE = 1.1$) than with the audio-tactile interface ($M = 5.8, SE = 1.1$) across all study parts. There was no interaction effect between study part and mode, however mean values and standard errors by both factors are displayed in Table 5.4.

We further compared usability ratings regarding distractors and task performance factors between EyeSee360 and the audio-tactile mode, see Table 5.3. In study part 3 but not in study part 2, users felt more disturbed by moving objects while performing the search task with EyeSee360 than with the audio-tactile guidance, $t(15) = -2.36, p = .032$. They further indicated they thought they had performed the secondary task faster ($t(15) = 2.40, p = .03$) and more precisely ($t(15) = 3.47, p = .003$) with the audio-tactile mode. Participants were also better able to concentrate on the secondary task ($t(15) = 3.21, p = .006$) and on the main task with audio-tactile guidance ($t(15) = 2.3, p = .036$). However, judging the vertical position was perceived to be easier with EyeSee360 ($t(15) = -2.44, p = .028$). Other usability ratings as the ease of performing the task, ease of learning, performing the main task fast and precisely, judging the horizontal position and distance, fatigue did not differ significantly between guidance modes.

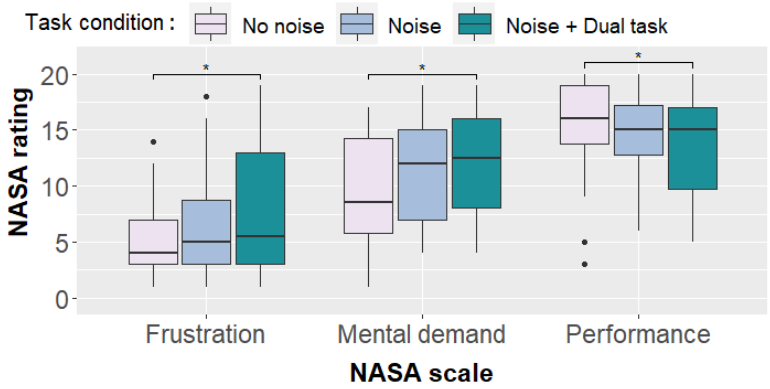


Fig. 5.10: NASA TLX scores across both guidance modes for the frustration, mental demand and performance subscale show significant differences between task load conditions, * = p <.05.

Regarding the overall usability of the system, users provided high ratings, indicating that they coped well with the task and the setup (see Table 5.5). Finally, users were asked post hoc which of the two guidance methods they would prefer using in VR/AR technologies and which method is potentially better to pay more attention on the surroundings on a 7-point scale (1 = audio-tactile, 7 = EyeSee360). With regard to the first point, user ratings

Table 5.4: Mean values and standard deviations by study part and guidance mode for NASA TLX subscales frustration, mental demand and performance.

Study part	Scale	Audio-tactile	EyeSee
1	F	4.8 (3.8)	6.3 (3.9)
	MD	9.3 (5.2)	10.4 (4.5)
	P	15.8 (3.1)	15.3 (5.1)
2	F	5.8 (5.2)	7.4 (4.6)
	MD	11.4 (5.3)	11.6 (4.9)
	P	14.2 (4.5)	14.9 (3.4)
3	F	6.7 (5.8)	8.5 (5.7)
	MD	11.9 (4.8)	12.5 (4.6)
	P	13.4 (4.4)	13.3 (4.8)

F = Frustration, MD = Mental demand, P = Performance

($M = 3.13 SD = 1.49$) indicated a slight tendency for the usage of audio-tactile feedback for the purpose of guidance in AR. On the latter point, ratings ($M = 2.56, SD = 1.77$) show a clear trend that users have the feeling to be more aware of their surroundings when using audio-tactile cues.

Table 5.5: Mean level of agreement with comfort and usability statements for the overall system on 11-point Likert items and standard deviations.

Statement	Mean Rating (SD)
Easy to detect targets	7,38 (2,47)
Sitting comfort	8,50 (2,24)
Interface (HMD+head strap) comfort	8,88 (1,76)
Task was easy to understand	10,13 (0,78)
Concentration on task	9,38 (1,17)
Easy to recognize targets	8,75 (1,52)
Improvement over time	9,25 (2,05)
Fun of use	9,88 (1,27)

5.7 Discussion

RQ_1: How well do non-visual guidance methods perform compared to visual guidance methods for a search task on different levels of task load?

In H_1 we expected that the performance of the guidance method would be related to the degree of subjective *workload* as experienced by the user. The hypothesis implied that EyeSee360 would outperform audio-tactile guidance on a low level of visual task load, but performance differences would be decreasing as soon as the task load would increase. The performance of the audio-tactile guidance on the other hand was expected not to decrease in

the high task load conditions. The self-evaluated mental workload generally increased with a higher task load during the experiment as intended, whereas users indicated that they did not put more effort into solving the task. Interestingly, there was no significant difference of the mental workload ratings across the guidance modes, which is in contradiction to our first hypothesis H_1. This was assumed since visual guidance methods usually compress a high level of information in a limited FOV (compare [144, 147]). However, this outcome has probably been reduced by recent improvements of the EyeSee360 method [145].

Surprisingly, our results show that the task load did not have an effect on task performance of both methods. Regarding hit rate, the audio-tactile guidance was on a par with EyeSee360 across study parts, which also indicates a comparable performance to other visual guidance techniques [146]. The overall hit rate of the audio-tactile mode was 3% better then EyeSee360 which was a small but significant difference in mean values (EyeSee360: 0.93% vs. audio-tactile 0.96%). However, the search duration per trial was considerably longer for audio-tactile guidance in comparison to using EyeSee360. This may be explained by the fact that audio-tactile cues used for guidance are relatively difficult to interpret and therefore require additional training until it can be used at higher speeds (see [267]). In comparison, the $focus + context$ approach used in EyeSee360 allows it to locate out-of-view objects directly and mostly intuitive (e.g., the proxy encodes already in which direction the user has to move their head to locate the object) [144, 146]. Another possible explanation regarding the consistently good guidance performance of EyeSee360 in the noise conditions could be explained by the observations that participants were partially able to fade out the background and focus mainly on the projection plane of the EyeSee360 interface (also see [68]). Therefore, the increased noise level did not affect the visual guidance method as much as expected and users were able to concentrate on the search task in a straightforward manner.

However, using EyeSee360 led to a significantly higher frustration level compared to the audio-tactile technique. We assume that this effect was partially caused by the occasional selection issues. During the selection phase, users sometimes needed several attempts to place the crosshair reliably on the target object, required for selection. This happened mostly if the object was placed at a high distance. We assume this problem arises as a part of stereoscopic depth disparities [68, 229, 221] in which users need to focus on different focal planes: EyeSee360 in the foreground for target guidance and object selection in the background. This problem might be aggravated if both planes have a large distance to each other - as is the case the target object is placed far away from the user

and thus appears smaller. In this context we found out that users generally took longer if targets were placed in higher distances with both methods. However, the tendency to an interaction between distance and guidance on hit rate shows that slightly fewer hits were made in EyeSee360 condition when targets were distant. This is comparable to [149], which also reports about a reduced selection accuracy of EyeSee360 compared to other visual methods. However, improving target selection, e.g., integrating a combination of head- and eye-based approach [228] or using novel interaction devices like 3D pen [328], might lead to a higher accuracy and overall hit rate for both visual and non-visual guidance.

Another source of frustration is potentially visual clutter. In this context, sensory overload is a relevant topic as visual guidance methods usually compress information into a relatively small FOV (see [104, 221]). Even though EyeSee360 was optimized to somehow reduce mental workload and visual clutter [145, 148], these techniques might still suffer from a limited FOV. By transcoding visual information into audio-tactile cues we potentially reduce the visual complexity and the number of distractors within the FOV. This allows the AR system to use the free display-space for any other non-guidance related further information. In this regard, it would also be interesting to investigate the search behavior between visual and non-visual guidance in context of a limited FOV (compare [387, 83]). Also, information density could be an additional factor which might affects search performance [406]. Generally, further studies are required to find out whether search behavior and performance differ considerably between visual and non-visual guidance methods. Also, while it makes sense to use wider FOV to reduce cognitive load, previous studies have only dealt with relatively low information density so far [25]. In addition, considerations should be given to how these factors might affect real AR environments compared to a highly controlled simulated AR environment in VR. While it can be assumed that results can be applied to AR systems up to a certain degree, simulated AR still has clear challenges related to the fidelity of the real-world component in the system. For example, physical conditions like relative brightness and contrast between real and virtual objects or the level of opacity of the virtual objects might have an additional impact on the user performance [342].

Finally, attention mechanisms likely play an important role. Human attention is primarily attracted to visually salient stimuli. Visual selective attention allows human perception only to focus on a small area of the visual field at a given moment [462]. However, multisensory integration and crossmodal attention have a large impact on how

we perceive the world, potentially enhancing selection attention in AR tasks. Providing information over multiple sensory channels has the potential to enable sensory stimulus integration. For this, attention mechanisms are used to process and coordinate multiple stimuli across sensory modalities, which also affects the way of managing resources [95]. However, multiple stimuli require a correct synchronization, otherwise sensory integration does not take place as stimuli could be interpreted independently [384]. This topic should be more closely addressed in further studies comparing visual and non-visual guidance methods.

In conclusion, with respect to H_1 we can state that although the audio-tactile guidance is slower, it is able to provide a similar and even slightly better hit rate compared to a well-established visual guidance method like EyeSee360. That is, when fast search times are not prioritized, the audio-tactile method allows precise guidance while freeing up the visual channel for other non-guidance information. The audio-tactile feedback can also be interesting for visually impaired people like [200, 348], since the same information is substituted to another sensory channel [262] without degradation of hit rate performance. For this purpose, also depth cues can be particularly helpful. Regarding to common depth judgement issues in VR/AR (see [394, 454]) the presented tactile depth cues might be supportive for a more accurate depth estimation.

RQ_2: Is there an effect of guidance method on situation awareness when a secondary task is included?

As expected EyeSee360 performed reasonably well in terms of an abstract object collection task. But regarding SA, an effect was noticeable as soon as the task load increased from study part 2 to 3. Audio-tactile guidance performed significantly better with regards to general perception (study part 2, noticeability) and SA performance measures (study part 3, secondary task performance). This outcome confirms our second hypothesis H_2 that a higher SA is achieved by using audio-tactile guidance. This, however, is not related to a higher workload when using EyeSee360 as initially supposed. With respect to the general perception, it was easy for most users (7 of 8) to notice the bird if audio-tactile guidance was used in study part 2. In contrast, only 2 of 8 users were able to notice the bird while solving the collection task with EyeSee360. This performance difference may be attributed to the focal disparity, in which users tend to focus on the AR-plane to primarily follow the visual guidance cues while blurring out the background (compare [68]). By this behavior, small details and objects are simply being overlooked by the user.

Concerning the performance measures, a significant difference between both methods became apparent. Almost all SA measures were significantly better with audio-tactile guidance than with EyeSee360 in the secondary task. Subjective questionnaire ratings also showed that users felt they could perceive their surroundings significantly better using audio-tactile guidance. This indicates a higher SA in terms of environmental perception using the audio-tactile interface and can probably be attributed to the fact that the user did not have to deal with visually related issues (clutter, occlusion, selection issues). Therefore, users were able to handle the main and secondary task in more balanced manner compared to the visual mode. In addition, frustration and workload also showed a significant difference for EyeSee360 in study part 3. Users were possibly more stressed when solving the main and secondary task at the same time. Since human capabilities for processing information are limited, there might not be enough capacity to solve a secondary task sufficiently while using a visual guidance method [273]. This could be due to the fact that that users tend to allocate their resources to higher priority-task components as soon as the arousal increases. Even though participants were briefed that both target search and secondary task were equally important to solve, some users might have prioritized the target search subconsciously since it was a continuous task over the whole user study. Furthermore, the level of immersion might be higher in a simulated environment if a visualization method is used for the search task compared to a non-visual approach. This possibly results in a trade-off between the degree of immersion and situation awareness, like reported in [194]. Generally, the usage of AR can cause distraction from the real world since it requires intensive concentration [13]. For that reason, reducing visual stimuli in the visible area of the AR device and substituting them into other modalities in a more intuitive way seems like a reasonable attempt to increase SA during the use of this technology. However, this approach is highly dependent on the current user task and further considerations have to be taken in case it is still required to display additional visual information inside the FOV.

Finally, we suspected a possible correlation between the main and the secondary task, namely that users tend to focus more on one task while neglecting the other. However, a statistical relationship between those two variables was not ascertainable in case of audio-tactile guidance. For visual guidance, however, the study revealed a quite contrary effect. It turned out that if users performed the object collection task well, they also showed a reasonable performance in the secondary task. In conclusion, with respect to H_2 we can state that a higher SA can be achieved using audio-tactile guidance. However, this result

could not be directly associated with a higher mental workload, but due to other factors that need further exploration.

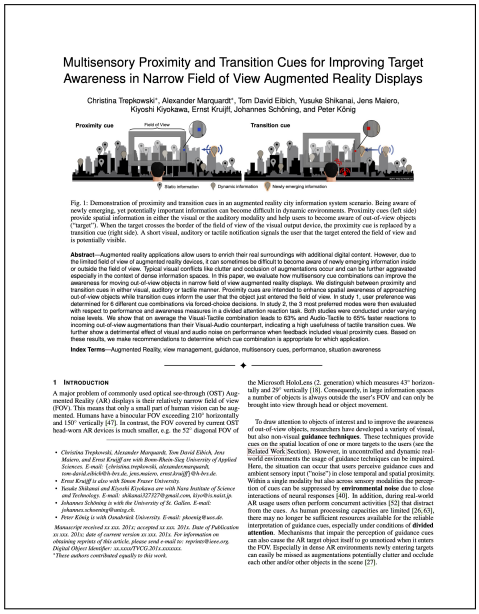
5.8 Conclusion

In this paper we compared EyeSee360, a state-of-the-art visual guidance method, with a non-visual guidance approach using audio-tactile stimuli. Doing so, we addressed head-mounted displays with narrow FOV. The main focus was on measuring performance, accuracy, cognitive load and SA during a object selection task in simulated AR. We used a vibration headband that consists of five vibration motors to create vibro-tactile feedback along the forehead and temples. By providing audio-tactile cues, it is possible to guide the user in the 3D space on the longitudinal, latitudinal, and depth plane. In particular, it can restrict negative effects like visual clutter or occlusion compared to common visual guidance approaches. As a result, we showed that users are more aware of their environment with audio-tactile in comparison with EyeSee360 with 16.5% more correctly detected background targets and 28% faster detection times, which implies a higher SA when using audio-tactile methods. However, during the main task, audio-tactile guidance performed slower than EyeSee360 regarding search times. As such, the choice of technique is context dependent - for example, if target search time is prioritized, EyeSee360 is preferable. Usage contexts that require improved SA and limited visual clutter can benefit from audio-tactile guidance. Despite the fact that the FOV in AR devices will increase in the future, it is still a challenging goal to build displays that could cover the entire human visual field [205]. Even if next generation devices included a larger FOV, they would still be limited. Therefore, problems associated with visual methods like cluttering, occlusion and a potentially high workload likely remain, especially with higher information density. Here, audio-tactile cues can help in order to address and improve these problems by substituting some of the visual information.

Future work includes an improvement of the physical setup. By extending the headband with more vibration motors (similar to [202]), a higher resolution and thus an increasing accuracy would be possible. This would allow us to investigate multiple target guidance in more complex situations more closely. Using a different motor driver technology like linear resonant actuators or piezoelectrics would also be reasonable in terms of usable bandwidth and acceleration characteristics to improve accuracy and performance. Another interesting venue of research is the combination of methods to assess characteristics like

performance and SA in more detail. In this constellation, audio-tactile cues could be responsible for the target point guidance while a visual method like EyeSee360 could be used to increase SA, e.g., warning of incoming objects similar to [194]. However, it still needs to be investigated how visual metaphors work best in combination with audio-tactile guidance without distracting or overloading the user with information. In this context it would also be worthwhile to consider the integration of transition signals to attract the user's attention as soon as certain information enters the FOV. Finally, eye tracking techniques can be used to enhance the effectiveness of both visual and non-visual guidance metaphors, support object selection [40] and could be used an additional indicator for situation awareness [413].

6 Multisensory Proximity and Transition Cues



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

Proximity cue Field of View Transition cue

Static information Dynamic information Newly emerging information

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.eibich4@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: shikanai3273270@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 coordinated equally to this work.*

This chapter investigates how multisensory cue combinations can enhance the perception of moving out-of-view objects in AR displays with a narrow FOV. The work presented in this chapter was informed by the awareness insights described in Chapters 2 and 5, as well as the methodological framework presented in Chapter 4. Here, we present novel multisensory cue combinations, consisting of proximity and transition cues in a visual, auditory, or tactile manner. The methods developed were investigated considering the influence of sensory constraints. For this purpose, we considered all intrinsic and extrinsic perceptual factors of sensory constraints relevant to this dissertation.

Overall, two user studies were conducted. In study 1 (n = 10), user preference was determined for six different cue combinations via forced-choice decisions. In study 2 (n = 14), the three most preferred modes were evaluated with respect to performance and awareness measures in a divided attention reaction task under varying noise levels. The combination of Visual-Tactile and Audio-Tactile proximity and transition cues resulted in faster responses to incoming out-of-view objects compared to visual-auditory cues. Overall, the results of this work demonstrate the usefulness of particular combined feedback modes for spatial and temporal perception and rapid reactions to incoming out-of-view targets, showing potential in raising SA. Finally, we have shown that modes that incorporate tactile cues can be particularly useful in noisy environments. Based on these findings, recommendations can be made regarding which cue combination is suitable for which application context.

The material in this chapter originally appeared in: Trepkowski, C.*, Marquardt, A.*, Eibich, T. D., Shikanai, Y., Maiero, J., Kiyokawa, K., Kruijff, E., Schöning, J., & König, P. (2021). Multisensory Proximity and Transition Cues for Improving Target Awareness in Narrow Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics, 28(2)*, 1342-1362. DOI: 10.1109/TVCG.2021.3116673. *Equal contribution.

6.1 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 [196]. 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[286]. 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 [167]. In addition, during real-world AR usage users often perform concurrent activities [208] that distract from the cues. As human processing capacities are limited [102, 247], 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 [104].

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. 6.1 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. 6.1 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.

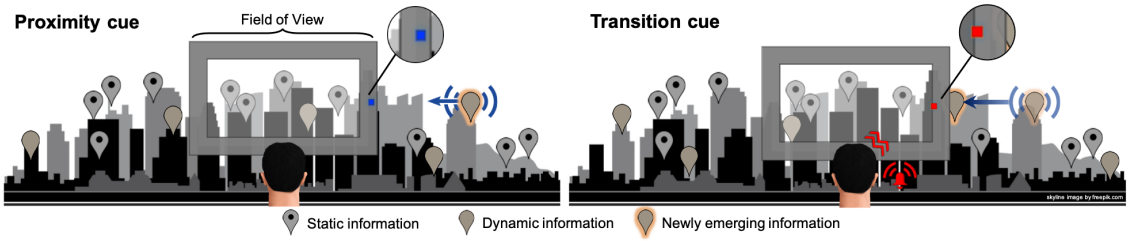


Fig. 6.1: Demonstration of proximity and transition cues in an augmented reality city information system scenario. Being aware of newly emerging, yet potentially important information can become difficult in dynamic environments. Proximity cues (left side) provide spatial information in either the visual 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.

6.2 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 [337]) 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: 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: 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.

6.3 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.

6.3.1 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 [403, 404] in contrast to other aspects (e.g., meaning of words) which have to be focused on [406]. Guiding features in the peripheral visual field can be beneficial [406] as they can significantly reduce visual search times [61, 448]. 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 [406]. 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 [340]. 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 [83, 184]. Finally, a change of visual scan patterns [83, 406], the degradation of navigation abilities and development of spatial knowledge [5, 435] 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 [276]. 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. [96]). Furthermore, the provision of redundant cues of different modalities can speed up responses [334]. In the following subsections, higher level visual and non-visual guidance methods are introduced that make use of aforementioned attentional mechanisms.

6.3.2 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 [55]. 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 [461] 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 [143]. 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 [153] – 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 [144]. 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[340]. 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 [54, 166]. Solutions like 3D Arrows, AroundPlot, 3D Radar, sidebARs, and Mirror Ball 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 [44].

Other methods trying to extend the limited FOV using a so-called fisheye view. For example, Aroundplot [189] maps 3D spherical coordinates to a 2D orthogonal fisheye, which addresses typical issues of location cue displays such as occlusion of information. Radial distortion [362] and EXMAR [178] 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.

6.3.3 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 subsection, we address both, purely non-visual methods and

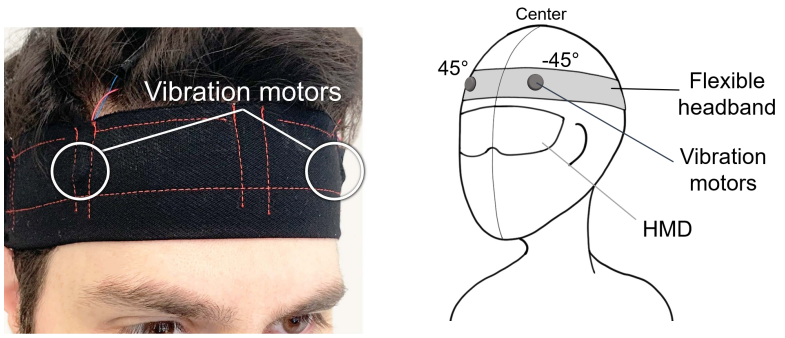
multisensory techniques that combine visual and non-visual cues. A meta-analysis [337] showed, that when added to a baseline task or presented simultaneously with visual cues, vibro-tactile cues enhance task performance. While vibro-tactile cues can be an effective replacement for visual alerts, they 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 [242], gloves [410] or a belt [218]. 3D selection methods using tactile cues are studied for non-directional feedback [10] and directional feedback [265]. Tactile cues also can be used to support visual search tasks [235] and target finding by device-based target scanning techniques [3] for mobile devices. Finally, researchers developed vibro-tactile head-mounted displays to provide guidance cues, e.g., in form of a ring-based tactile setup [88, 267] around the user's head or even higher resolution tactor grid resembling an EEG setup [202] to enhance awareness and support guidance.

Auditory cues also have been used to support visual search [292] and navigation [201]. 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 [279]. Another approach of using auditory cues is called sonification in which sound properties are used (predominantly pitch, but also loudness, duration/tempo, and timbre) [98] with respect to the presence of the auditory reference. Sonification metaphors are commonly used for speeding up visual search tasks (e.g.[255]) 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 [315]. This metaphor can also be applied for spatial data exploration and guidance for visually impaired people in AR [43, 348] or for improving accuracy in high precision tasks [352] without occluding the visual field. Furthermore, cross-modal effects have been studied, including audio-tactile effects [175, 301] and conflicts between audio and visual cues [217]. In general, multisensory stimulation has been shown to have benefits over unimodal feedback with regard to improved task performance [56, 185, 337] and higher user satisfaction [234]. Considering real-world usage conditions, multisensory feedback can also improve noticeability under noise [20] and reduce cognitive load [311, 375], 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 [268]. However, in comparison visual feedback is generally highly trained, easy to learn and can be interpreted fast [268]. Thus, choosing the appropriate method can be challenging and depends on the context of application.

6.4 System and Study Setup

In our studies we deployed the widely used Microsoft HoloLens (first generation). 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. 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 Section 6.6 "Discussion"). 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 [388].



(a) User wearing the tactile interface.

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

Fig. 6.2: Custom-made tactile headband with 2 Vibration motors are placed on −45° and 45° from the user's point of view.

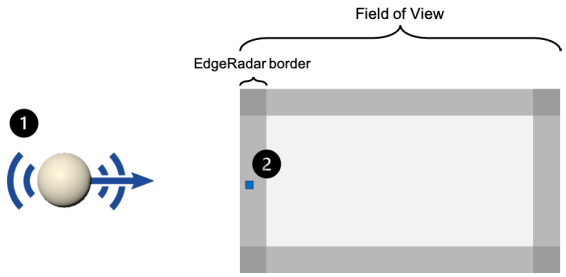
A modified version of [268] with 2 vibration motors (Precision Microdrives 308-107) provided accurate tactile cues on the user's forehead. We assume that adding further motors was not necessary to improve the feedback, since we focused on specific object movement trajectories in our studies (see the 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. 6.2a). 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 [89]. 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.

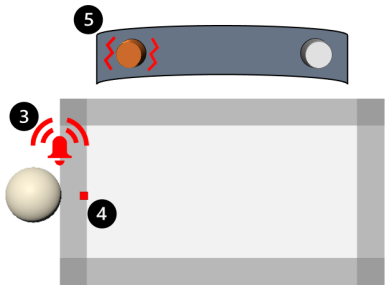
6.4.1 Feedback Modes

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. 6.3a) when an augmented information approaches and triggering a short transition cue (see Fig. 6.3b) 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 (~ 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 6.5.2). The provision of proximity 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. 6.1). 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

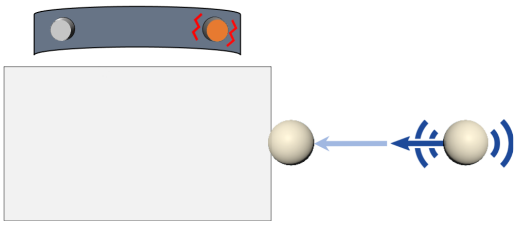
were not part of the conducted experiments. In the following, we will further describe our design choices and justifications for each feedback modality.



(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 tactor on the right side of the tactile headband triggers a short vibration burst.

Fig. 6.3: 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.*

6.4.2 Proximity Cue

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. 6.6), 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. 6.1.

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 [153], 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. 6.1 or Figs. 6.3a and 6.3b). 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 [153]. Although methods like 3D Radar have proven to be very useful in tracking off-screen object trajectories as well [146], the visualization still suffers from a certain overlapping and cluttering with on-screen augmentations, see [153]. 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 [146]. 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 [143]. 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. 6.3a) 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 [459] and frequency discrimination works sufficiently well [64] 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., [202]), 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 [360]. 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.

6.4.3 Transition Cue

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.

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 [317]. For tactile alerts values between 200 ms for short vibration respectively 600 ms for long vibration are recommended [360]. 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 [224]. 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 behavior [227]. 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 [291] 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 [267] 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.

6.4.4 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 [129, 221]. In particular, bright light intensity can limit projection in AR displays [221]. 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) [129, 221]. Background colors and textures of common objects in an urban setting can also vary and affect perception in AR [129]. As for audio noise, it is emitted in many locations and situations with different sources and sound pressure levels [74]. Only few situations in which vibro-tactile noise occurs can affect perception, e.g., low-frequency whole body vibration [20]. The presence of environmental noise can suppress the perception of cues due to close and direct neural interactions among sensory inputs [167]. Research addressed perceptual suppression within the visual [48], auditory [434] and vibro-tactile [423] modality but also cross-modal effects [167]. However, cues provided in one modality can also amplify the perception of cues in another modality through additive or facilitatory integration [167, 390]. So, multisensory stimulation can enhance noticeability [20] and improve task performance [56, 185, 337]. A predictive cue can also improve awareness and the perceptual sensitivity for a subsequent masked target [277]. 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 [155]. These advantages motivated the integration of multisensory modes for comparison.

6.4.5 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.

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

Fig. 6.4: 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.*

6.4.5.1 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. 6.4a). 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 [221]. 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 [305]. In terms of ambient auditory background noise, an average loudness of the laboratory environment of about 34.5 dB was measured (that corresponds to the acoustics of a library) on the basis of an audio recording made during this condition (see Fig. 6.5c). Also, the contrast ratio for the visual proximity cue (proxy symbol) and the background was calculated following WCAG 2.1 guidelines [428] 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 50 dB at ear level. This roughly resembles the WCAG 2.1 audio guidelines, which recommends a 20 dB difference to background sounds to make audio content clearly perceivable to the user.

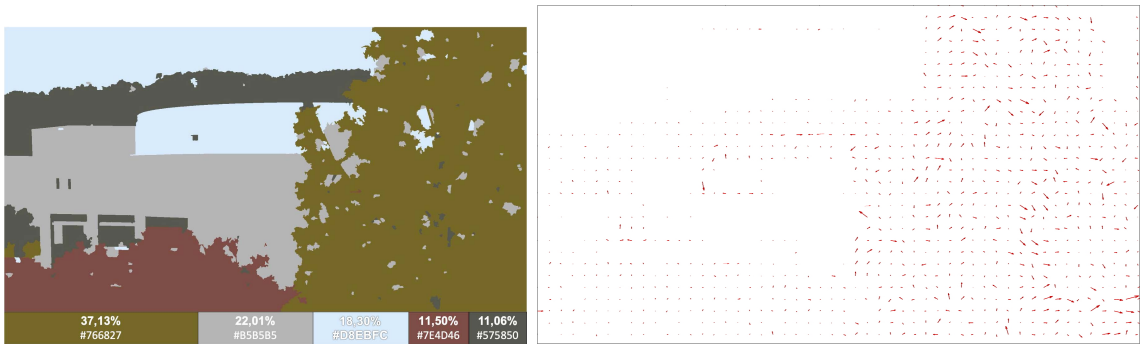
6.4.5.2 Increased-Noise Condition. The increased-noise condition included a vivid outdoor background (see Fig. 6.4b) with a heterogeneous color palette (see Fig. 6.5a). It contained lighting, some optical flow and a higher level of ambient auditory noise. Users were seated about 1.5 meters in front of a window. The outdoor environment viewed from the window made up the entire background of the AR display as can be seen in Fig. 6.4b. 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 one 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 [129]. Furthermore, we analyzed the background scenery in the increased-noise condition by using fuzzy c-means algorithms (as proposed in [236]) 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. 6.5a. Also, in contrast to the reduced-noise condition, motions do exist in the background of the environment, which potentially affects user performance [208]. 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. 6.5b). The user was exposed to an average auditory noise of about 44.6 dB (resembling a sound intensity similar to light traffic noise), which are about 10 dB more than the user experiences in the reduced-noise condition – therefore roughly doubling the loudness on a subjective level [74] (see Fig. 6.5c).

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 5.4 dB 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 6.4.2), the stimulus should still be reasonably well distinguishable despite the exposure of auditory noise.

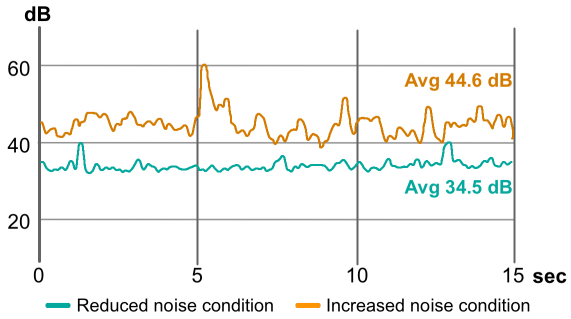
6.5 User Studies

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



(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 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. 6.5: 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. 6.4.

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 [120]. Regarding search behavior in AR,

there also seems to be a tendency for horizontal search patterns [406], which causes targets to enter the FOV more often through horizontal head movements [224].

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.

6.5.1 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 [342], 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 6.5.3). 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 (approximately six 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.

6.5.2 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 [234] where the multimodal interface yielded higher user satisfaction than the unimodal variant. We also expected a higher preference of multimodal modes due to spatial synergies in the attentional processing of information across sensory modalities [96]. Among the multimodal modes we further assumed that modes with tactile and audio transition cues would be preferred over modes with the visual variant as they have a higher temporal sensitivity [418]. 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.

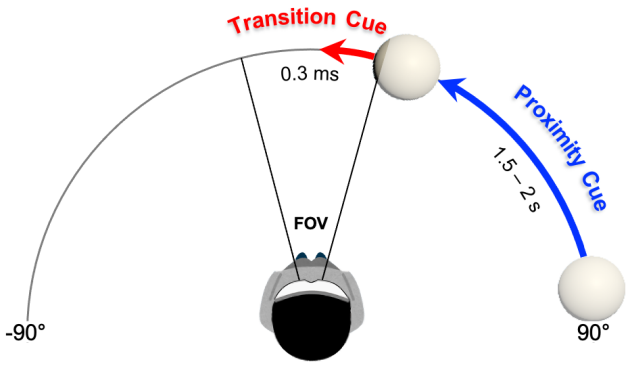


Fig. 6.6: Top view: During the awareness task, the target sphere follows a circle path around the user on the horizontal plane. The user receives the proximity cue as long as the sphere follows the path outside the FOV (blue line). As soon as the sphere touches the border of the FOV, the transition cue is triggered (red line).

6.5.2.1 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 Table A.3 in the Appendix for the exact wording of respective items (1-6).

6.5.2.2 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. 6.6). 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. 6.6) 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 [437], reference group effects [84], and non-linear distances between anchors [444]. 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 Subsection 6.5.2.1 "Method"). The second part of questionnaire comprised items on overall comfort and usability (see Tables A.6 and A.7 in the Appendix).

6.5.2.3 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. 6.7. Simple effects were examined by comparing different modes at each noise factor level. In the reduced-noise condition as expected (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 (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 6.1). As 10 participants took part in the study the maximum total score was 100 in each the noise with increased and with reduced noise.

Table 6.1: 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.

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

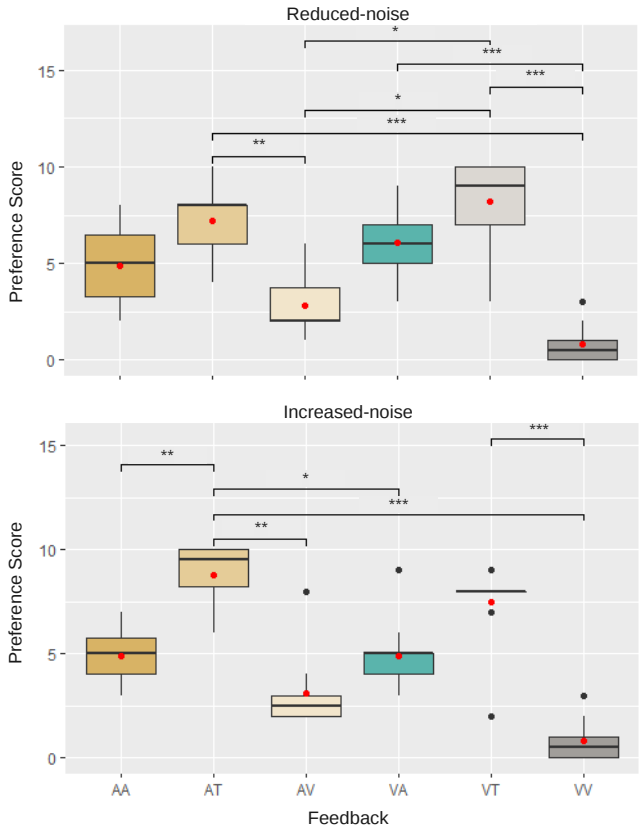


Fig. 6.7: Study 1, condition with reduced noise (top) with comparisons of preference scores of AT, VA, and VT each versus AA, AV, and VV feedback (see H_1) and Noise condition (bottom) with comparison of preference scores of AT and VT each versus VA, AA, AV and VV (see H_2). Only significant differences 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. 6.8 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 ($Z = 2.57$, $p = .03$, $r = .813$).

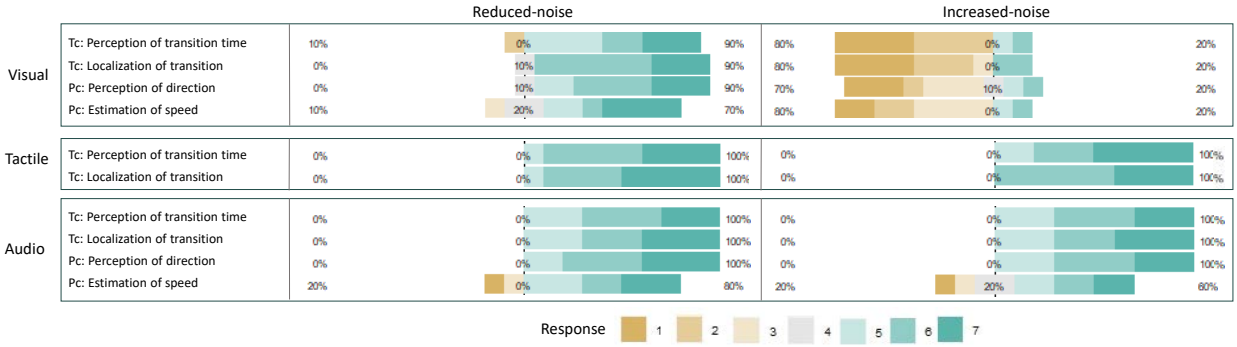


Fig. 6.8: Study 1: Users indicated their level of agreement on 7-point Likert type items (ranging from 1= "strongly disagree" to 7= "strongly agree") on how the respective cue helped them to perceive temporal and spatial events of an AR object under conditions with reduced noise (left) and with increased 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).

6.5.2.4 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 6.5.2.1). 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 [106]. 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 6.1). 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.

6.5.3 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?

Regarding RQ_2 we expected 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 [299]. Under increased noise, we expected a weaker detrimental effect on the auditory than on the visual proximity cue based on results in 6.5.2.3. 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.

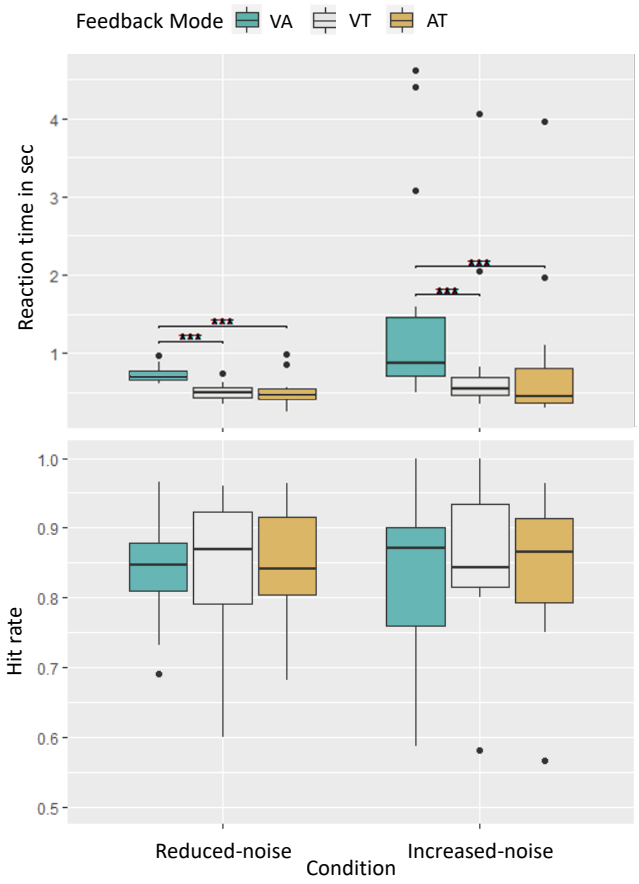


Fig. 6.9: Average reaction time in the awareness task (top) and hit rate in the focal attention task (bottom) for VA, VT, and AT modes by noise condition in Study 2. ***p <.001.

6.5.3.1 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 Table A.3 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).

6.5.3.2 Tasks and Procedure. Users were required to notice and react to an approaching AR target by using VA, VT, and AT (see previous Subsection 6.5.2.4). 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. 6.6). 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 [92] 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. 6.4), 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.

6.5.3.3 Results. The Aligned Rank Transform procedure was used [446] 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. 6.9). 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. 6.9). 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.

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. 6.5b) 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 ($t(13) = 2.03$, $p = .064$, $d = 0.54$). As expected, in the noise condition, 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. 6.11 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 [140] 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).

6.6 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.

6.6.1 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 [216]. To further enhance the effectiveness of the AA mode, it might be helpful to choose more distinguishable timbres for auditory proximity and transition cues [379]. Partly conform with H_2, 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

(i.e., [200, 279]) and has been considered to distribute content over visual and auditory modalities as vision can be highly loaded during certain (visual-only) based applications [379].

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 [227] 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 [168]. However, adjusting the characteristics of the transition cue, e.g., increasing its size, might increase the noticeability of peripheral augmentations [224]. 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. 6.8) 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 [224].

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 [360]. 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. [202]. Such sensory substitution devices are commonly used, e.g., for people with visual disabilities [262] and in terms of tactile cues considered to be a fruitful approach to improve user performance [267]. 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.

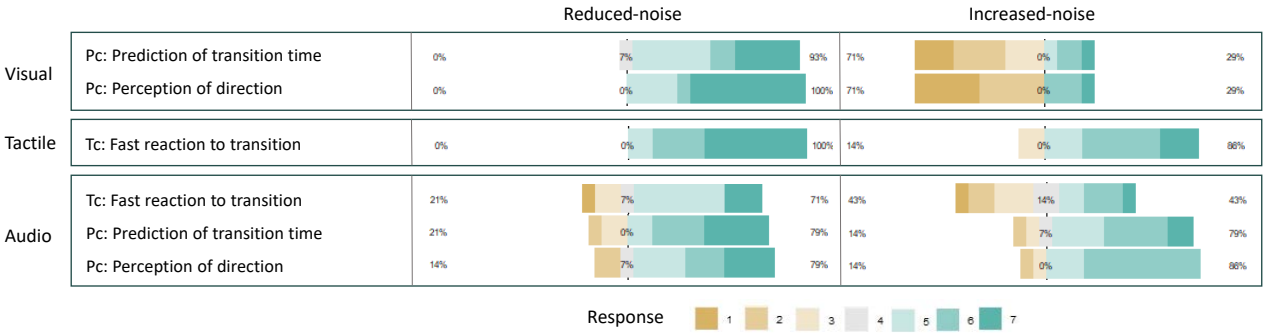


Fig. 6.10: Study 2: Users indicated their level of agreement on 7-point Likert type items (ranging from 1= "strongly disagree" to 7= "strongly agree") on how the respective cue helped them to perceive temporal and the spatial events (listed on the left) of an AR object under conditions with reduced noise (upper part) and increased noise (lower part). Pc = Proximity cue, Tc = Transition cue.

6.6.2 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 entering 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 [430].

However, different from our expectations in H_3, reaction time for VT and AT did not differ significantly in the awareness task. H_4 was also not confirmed as in the focal attention task no differences between all three modes were found. Findings differ from [268], 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 [280] although the vision of text, shape and color degrade towards the periphery [168]. It should also be noted that proximity cues have a supportive role in the awareness task by spatially cueing attention [334] and preparing the user for the upcoming transition. Thereby, attention can also be oriented covertly, thus without eye or head movement (e.g., [27, 332]).

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 [299] and in divided attention situations e.g., when indicating the direction to walk [355], or warning users [289, 293]. This underlines the noticeable yet unintrusive character of tactile feedback and the intuitive design of this particular tactile cue (compare to [88, 202, 267]).

Similar to [19], 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 [20]. As pointed out in [379], 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. 6.5) 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 [221], 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 [459]. Even though it requires more training than for visual cues to interpret it correctly [267]. 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 [432]. 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.

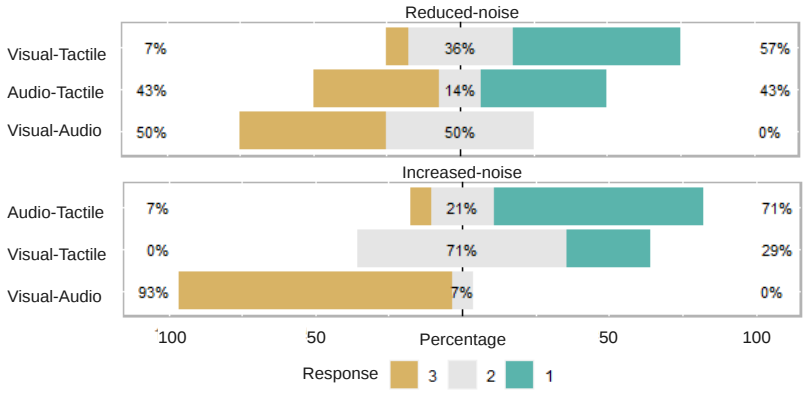


Fig. 6.11: Study 2, post-hoc questionnaire: Frequencies of the assignment of ranks (3=third, 2=second, 1=best) to a mode in the condition with reduced (upper part) and increased noise (lower part).

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 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, during 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 [363, 382] while tactile cues can be especially useful in noisy environments [289, 293].

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. 6.5a, 6.5b). 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 [302].

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 [141] to [146]). 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 [442], an increasing azimuth and elevation localization error behind the head [308] and a high listener specificity [436]. 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 [146]). 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 [429]. 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 [430]. 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 6.6.3.2 “Wider Field of View”), 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.

6.6.3 Reflection towards Field of View

6.6.3.1 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 [406]. 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 [406]. In addition, guided search can also counteract varying and unstructured visual search patterns that occur when the field of view narrows [406]. 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 [152] before, while 3D auditory cues have been shown to be intuitive and easy to use [265].

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 [311, 375] and capture spatial attention under dual task conditions [363, 382].

6.6.3.2 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 [205]. 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 [208]). In addition, with an extremely wide FOV in-view information becomes much less noticeable towards the periphery [208]. 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 [267]. 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 [102, 247]. To limit cue activation, the amount of information can be reduced by filtering and clustering techniques (e.g., by content, distance, amount, etc.) [399]. 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 [208]) 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 [44]. This method is useful for tracking moving out-of-view objects while avoiding clutter in the foveated vision of the user [146].

6.7 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 preselected 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.

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 [153] and auditory [464] 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 [419]. 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 [202]). In this manner, we will circumvent perceptual limitations at less sensitive regions of the head [89] and areas with high hair density [295]. 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 [88, 202].

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 [40], and to use it as an additional indicator for situation awareness [413]. 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.

7 Discussion and Conclusion

In this chapter, we discuss the research questions raised in Section 1.3 and conclude the work described in the foregoing chapters. Finally, we provide an outlook for future work.

7.1 Discussion of the Research Questions

In this section, we will provide detailed answers to research questions RQ_{1-3}. The discussion should shed light on the overarching research question RQ_{Main} to identify the potentials and limitations of multisensory guidance under sensory constraints.

7.1.1 RQ_1 :What is the effect of hand-based non-visual guidance on task performance in visually complex environments?

Sensory constraints to be addressed: Depth perception, sensory thresholds, scene structure.
Method: Hand-based non-visual guidance using audio-tactile proximity feedback and tactile patterns for motion guidance.

To address this research question, we developed two non-visual, hand-based guidance approaches that can be used for selection and manipulation task in visually complex AR/VR scenarios: (1) *audio-tactile proximity feedback* and (2) *tactile patterns for motion guidance*. Although visualization techniques exist to cope with visual ambiguities [8], results may vary with respect to spatial understanding. Our approaches are decoupled from visual feedback. Using this strategy, the performance of non-visual guidance under complex visual bindings could be investigated without being influenced by any particular visualization method.

Audio-tactile proximity feedback provides (a) high-resolution collision and friction cues over the full hand to improve touch compliance, and (b) audio-tactile proximity cue models to enhance spatial awareness and task performance. Proximity models were divided into scene-driven outside-in cues to support scene exploration and selection phases, and object-driven inside-out proximity cues used for manipulation phases. Specifically, audio-tactile proximity feedback was used to improve performance by reducing unwanted collisions, object penetrations, and errors (complete object pass-throughs). These problems

occur frequently in ambiguous, visually complex scenes [10, 42], causing depth estimation problems [10] and object occlusions [258].

Significant performance gains through multisensory feedback, for example, by combining vibrotactile and auditory cues, have rarely been demonstrated for typical interaction tasks such as 3D object selection or manipulation [10, 111]. Studies on single-point, non-directional feedback for selection tasks have shown that proximity-based multisensory feedback affects selection movements and user performance [10]. Scene-driven outside-in proximity feedback was used to assist the user's hand in exploring the scene under visual constraints, such as maneuvering around hidden obstacles. Our approach of providing directional cues on a higher density tactile grid enhanced spatial awareness in the selection phase due to the improved localization of contact points compared to glove-based approaches with lower resolution (see [85]). This results shows that the use of a higher-resolution tactor-grid distributed over the entire hand has a great impact on task performance compared to lower-resolution glove-based approaches. Further comparisons could be made with grasp-based tactile systems used for selection and manipulation tasks [344]. Here, users reported that the directional judgments of the vibration feedback were somewhat limited. In contrast, our approach demonstrated good discriminability in terms of contact point perception. While a contact point alone cannot identify an impact on vector, it allows an indication of the general direction of impact [203]. This effect could be attributed to the chosen distribution and density of the vibration motors used in our system, which encompassed the fingertips, inner palm, middle phalanges, and back of the hand. In particular, the back of the hand is normally used infrequently, but achieved equally good results as the index finger in terms of tactor localization and differentiation. By allowing contact points on different hand zones, direction could be better distinguished compared to multiple contact points within a single area, for example, along the palm [344]. This effect presents new opportunities for selection and manipulations tasks as potentially useful locations for contact-driven feedback. Finally, with regard to improving spatial awareness of the objects in a scene to guide hand motion, high-resolution feedback enabled a better understanding of where the hand had been touched or collided with. In general, we have shown that increased resolution can improve performance, which is in line what has been reported in [344] for handheld tactile devices.

Furthermore, audio-tactile inside-out proximity feedback significantly reduced the number of object collisions and errors compared to common collision-based object manipulation. Notably, good results were also obtained with unimodal auditory proximity

cues. This outcome could have been due to a higher cognitive load, consistent with studies [301, 344] showing that a combination of stimuli (e.g., visual-tactile) can sometimes lead to a lower increase in performance. However, complementary multisensory cues have been shown to reduce visual ambiguities [412]. The improved performance is likely based on the principle of maximum likelihood estimation [329], which states that a more reliable estimate (with the least variance) has a greater impact on the final perception. In this context, it has been shown that tactile modality can be modulated by auditory stimuli, indicating a dominance order [49]. Therefore, cognitive effects [329] may compensate the performance improvements to some extent. However, it is not uncommon for multisensory interfaces to increase cognitive load [425]. Further work revealed, that a switch from bimodal to unimodal tactile feedback could lead to a significant performance increase in touch related-tasks [373]. In this case, the change from audio-tactile to purely tactile cues resulted in a benefit that could not be achieved in the reverse order when a sound was added after a purely haptic experience [373]. This result suggests that other task-related factors might influence facilitation by multisensory cues. We also assume that switching the feedback models based on usage mode (outside-in for selection, inside-out for manipulation) might lead to confusions, potentially increasing cognitive load. One possible approach to address this issue would be to introduce a common model for proximity feedback for both phases.

Tactile patterns for motion guidance used a tactile arm sleeve in addition to the high-resolution tactile grid on the hand to deliver vibration cues to the outer and inner wrist. The tactile approach described in Chapter 2 informed the design of the tactile feedback presented in Chapter 3. For this, the prototypical setup as well as methods for tactile stimulation could be used as a template, which have been further developed for the purpose of tactile pattern generation (see Fig. 7.1).

Researchers have primarily focused on guiding arm motion or general poses (e.g., [23, 41, 211, 275, 463]). However, 3D selection and manipulation techniques often require fine motor control [226], which is not feasible with previous approaches. By applying higher-resolution vibrotactile patterns to the entire hand and wrist, the execution of specific biomechanical actions could be triggered. These actions includes detailed joint/muscle activations such as radial/ulnar deviation of the hand (yaw), pronation/supination of the hand (roll), hand/arm movement, and finger flexion/extension (e.g., to push buttons or grasp objects). This feedback is particularly useful for selection and manipulation tasks, such as assembly instructions [460], that require fine manipulations in visually

complex environments in which components may be mutually obscured [258]. Few other systems have explored the controls of finer hand and arm movements using electromuscular stimulation (EMS) [253]. However, EMS-based approaches require initial calibration, cause rapid muscle fatigue [253], and may even damage the stimulated receptors or muscles [303]. In contrast, our approach has extended previous tactile methods that support only general arm and pose guidance (e.g., [23, 41]) by fine-grained hand and arm motion guidance. Here, the granularity provided is comparably high to EMS-based methods [253, 396] but lacks the aforementioned drawbacks. Although the tactile patterns were generally well interpreted by most users, the guided motions were sometimes performed in the opposite direction. This result is comparable to other findings [380], in which users interpreted some patterns as either push or pull motions. This effect might be attributed to individual differences in the interpretation of directional patterns. We expect that the use of personalized patterns may lead to more reliable performance of correctly executed motions. Such individualized pattern would likely reduce the level of abstraction [275], resulting in faster learning of the intended motion. In this context, a potential learning effect was noticeable, as users were able to interpret motion patterns more quickly and reliably over time. Finally, the directionality of patterns can potentially be further enhanced by modulating feedback characteristics such as stimulus duration and the inter-stimulus onset interval [335].

Our results show that both non-visual, hand-based guidance approaches have extended current the state-of-the-art and are highly useful to improve task performance in visually complex environments. Furthermore, these results are also likely to be transferable to scenarios in which other sensory constraints exert an influence on user perception, for example when using an OST AR HMD with a narrow FOV. Because auditory cues alone have been shown to work well for proximity, a combination of the two guidance approaches, specifically audio-only proximity cues and tactile cues for motion and guidance, might be a useful extension. We assume that auditory and tactile cues are easy to separate because the cues refer to different actions. Overall, the contributions presented in Chapters 2 and 3 provided important insights that significantly informed the work described Chapter 4, as can be seen in Fig. 7.1. The methods and the findings on perception, cognition, and usability from Chapters 2 and 3 thus formed the conceptual basis for further research presented in this work. For this purpose, general insights on how spatial object localization can be represented in non-visual form were considered in the guidance approach described in Chapter 4. In addition, methodological concepts were further adapted, including the

modulation of auditory feedback (change of frequency and amplitude depending on the position of the target) from Chapter 2 or vibration patterns from Chapter 3, which partly inspired the distance encoding in Chapter 4. The methods were successively refined for head-based audio-tactile feedback and for application in search tasks in AR environments under sensory constraints.

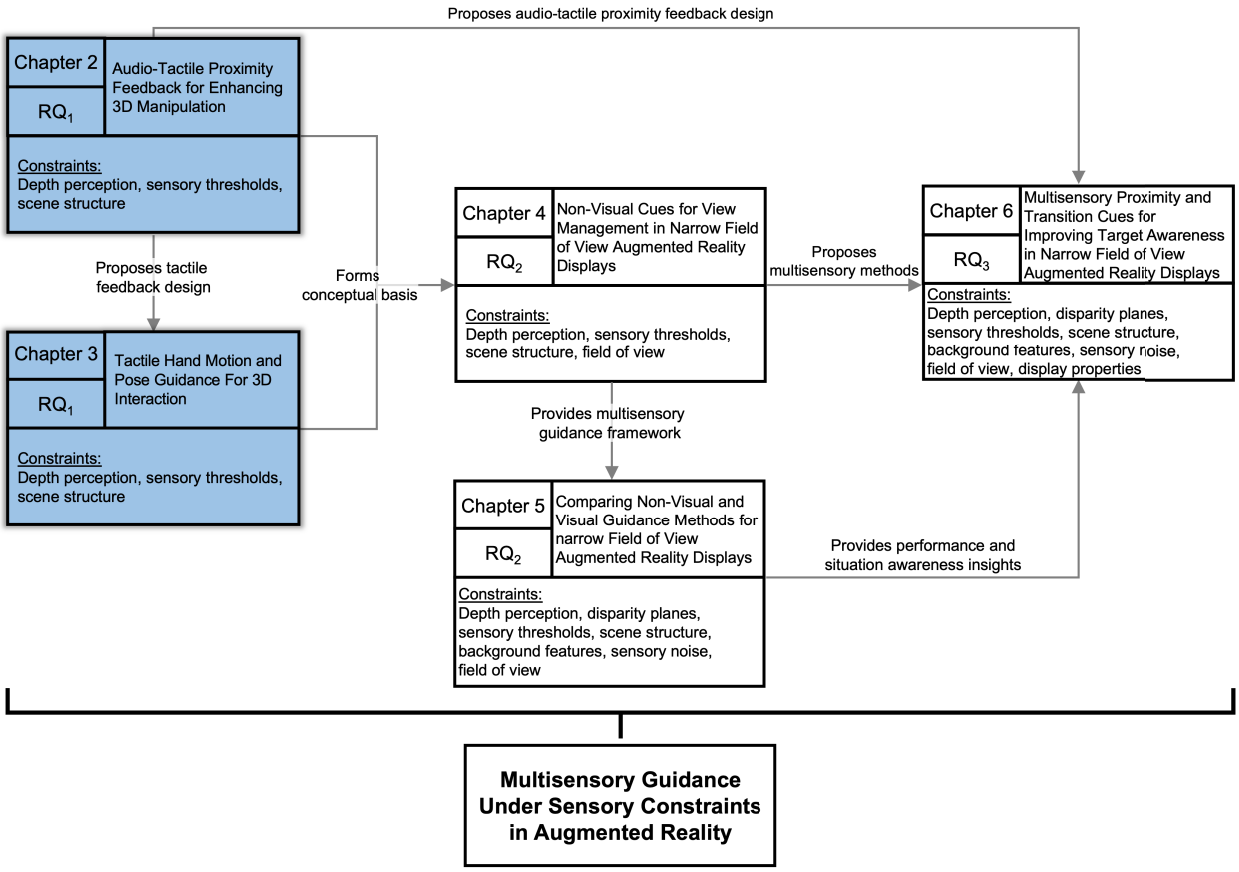


Fig. 7.1: Stages of this thesis: The highlighted Chapters 2 and 3 describe the contributions to RQ_1. The results of these chapters influenced the work described in Chapter 4. Furthermore, the results from Chapter 2 partially influenced the approaches presented in Chapter 6.

7.1.2 RQ_2 :How effective are head-based multisensory guidance cues to support search under sensory constraints?

Sensory constraints to be addressed: Depth perception, sensory thresholds, scene structure, background features, sensory noise, field of view.

Method: Head-based multisensory guidance using audio-tactile cues to support target localization on longitude, latitude, and depth.

To answer this research question, we developed a custom-made interface to provide audio-tactile cues at the user's head to support search in AR. We found that *non-visual guidance for narrow FOV AR* could potentially improve the affected perception caused by a variety of sensory constraints during search. Improvements are achieved by encoding the longitudinal and latitudinal position and distance information for a target object in 3D space into auditory and tactile guidance cues. Thus, the initial visually complex view could be partially untangled and a detailed localization of target positions with less ambiguous depth information could be provided. We then evaluated the search performance of multisensory guidance under the influence of additional sensory constraints by *comparing non-visual guidance and visual guidance methods*. The experiments investigated situations under sensory constraints, such as searching for non-visible or occluded target objects, or target searches in dense scene structures. In addition, all studies were conducted with a widely used OST AR HMD, the Microsoft HoloLens (1st gen) with a FOV of 30°H and 17.5°V, leading to the characteristic perceptual problems of a narrow FOV in AR [221].

For *non-visual guidance for narrow FOV AR*, we examined all possible combinations of auditory and tactile cues to encode longitudinal, latitudinal, and distance information for a target object in 3D space without visual aids. Sensory feedback was provided to the user continuously as a relative function of the current head orientation and target location. Specifically, we examined encoding by vibration intensity modulation, pulsed vibration, and auditory frequency and amplitude modulation. Current research has mainly focused on object localization through vibration cues [88, 202] and crossmodal effects in visual search [175, 301]. We have extended the current state-of-the-art by providing relative audio-tactile feedback to support search under sensory constraints not only for longitude and latitude but also for depth. In most cases, our approach could not be directly compared with the previous work in terms of search performance, because the technical setup, feedback methods, and search task differ considerably. However, we make a general comparison and discuss the differences below.

First, the localization capabilities for non-visible target objects that are spatially distributed in 3D space were investigated. This issue is particularly important when targets may be obscured by other information or the environmental structure itself, for example, in dense information scenes [221]. With respect to longitudinal cues, we were able to achieve similar high accuracy compared to well-performing vibrotactile guidance approaches [88] with maximum median deviations of only 2°. In terms of latitude accuracy, we obtained significant improvements by using combined audio-tactile feedback compared to previous

work [88] that used tactile cues only. This result is consistent with other findings, showing the potential of similar sonification strategies for 1D guidance tasks [315]. Potential crossmodal effects were found between different encodings. For example, overall accuracy was significantly higher with unimodal vibrotactile cues for latitude (vibration intensity modulation) and depth (pulsed vibration) compared to depth encoded by audio. This could again be explained by cognitive range effects [329], with tactile cues providing a more reliable estimate and affecting auditory perception in this particular encoding. Finally, the guidance mode latitude by audio and depth by vibrotactile pulse exhibited the highest accuracy in latitude estimation as well as the highest subjective preference. We assume that depth coding by vibrotactile pulses had a slight advantage over other tested depth encodings because it was based on a well-known sonification strategy from real-world applications [315] and therefore may have been more intuitive to users.

Next, we examined the search times for all tested modes in dense scene environments. As is typical for such scene structures [221], target and distractor objects were densely arranged in space and also distributed in depth, making localization of the target object difficult for previous methods [88, 202]. Here, the audio-tactile mode resulted in the shortest search times. Over time, the variability of search times for all modes decreased, indicating a possible training effect. This result suggests that with sufficient training, other modes can potentially achieve search times comparable to the best working mode, which was latitude by audio and depth by vibrotactile pulse.

Finally, we examined a simplified variant of non-visual feedback. Here, an absolute localization cue was provided without continuous guidance feedback, consisting only of longitudinal and depth cues. This approach considered real-world applications, such as city guidance systems, in which most information can be presented within a narrow latitude range; therefore, providing additional elevation information at a given time may not be necessary. Since the effectiveness of longitudinal vibration cues had already been demonstrated, it was of interest to investigate how well transiently provided depth cues perform for information localization in the absence of latitudinal cues. Of note, audio cues were more accurate than vibrotactile pulse for absolute depth perception. This effect might be explained by previous findings [50] that a lack of continuous vibrotactile stimulation affected localization abilities due to missing reference point. Notably, users were able to locate nearby targets more accurately than targets that were farther away using auditory depth cues. This effect can be explained by the feedback design of auditory cues. Targets were presented in a depth-dependent manner in the auditory frequency range from 300 Hz

(closest) to 1300 Hz (furthest). Thus, close targets were represented by lower frequencies than targets farther away. Human frequency discrimination performance is higher for low frequencies and becomes worse at higher frequencies [254, 331]. Therefore, higher frequencies should result in less accurate depth estimates for distant targets.

In the next step, we focused on *comparing non-visual guidance and visual guidance methods* to gain further insight into the effectiveness of non-visual guidance under sensory constraints. Visual guidance techniques in AR have become reasonably efficient in conveying the location of virtual objects in a scene [44]. Therefore, we compared the best non-visual guidance method identified in previous studies (hereafter referred to as audio-tactile guidance) with a state-of-the-art visual guidance method called EyeSee360. This method has been demonstrated to be particularly effective to aid search guidance in AR [44, 144]. As an extension of our preliminary work, we investigated the influence of additional sensory constraints on the search performance of visual and non-visual guidance approaches. By considering constraints such as sensory noise and background features, a more realistic search scenario could be investigated compared to experiments typically conducted under ideal laboratory conditions (e.g., [44, 88, 144]).

We investigated the search performance of both audio-tactile guidance and EyeSee360 for visual search task under different task load levels: In task load level 1, we examined baseline performance under the sensory constraints of a dense scene structure with a narrow FOV in a uniform environment without other visual distractors. Task load level 2 took place under more real-world-like conditions by also considering background colors, textures, and motions (in the form of a dynamic city environment) as well as ambient auditory noise. Here, targets and distractors were spatially densely distributed in the environment. Furthermore, objects were located at different depths, in front of objects with similar colors and textures, and temporarily occluded by parts of the dynamic environment (crowds, traffic). At task load level 3, a visual secondary task was added to the previous conditions. We hypothesized a trade-off between search performance and task load for visual guidance, with performance decreasing as task load increased. This was to be expected, as visual guidance leads to visual clutter and higher mental load, especially for devices with a narrow FOV (see [44, 221, 354]). Audio-tactile search performance was expected to remained independent of task load as the visual field remains unobstructed, thus avoiding the undesirable effects of visual guidance. Unexpectedly, task load showed no significant effect across guidance modes. This result might be attributed to recent efforts in clutter reduction strategies in EyeSee360 [145, 148]. It was also found that task

load did not affect search performance for both methods, leading to the conclusion that both audio-tactile and visual guidance are well suited to be used for tasks under higher visual load [322]. In terms of performance, it is worth noting that the hit rate of audio-tactile guidance was comparable to and even slightly exceeded that of EyeSee360 (3%), suggesting that audio-tactile guidance can be similarly accurate to other well-functioning visual guidance methods as well [44, 146]. However, search times for audio-tactile guidance were significantly slower compared to using EyeSee360. This can be explained by the fact that such focus+context approaches are easy to use and provide an immediate, intuitive sense of the object's position in relation to the head [37]. Compared to such approaches, the initial learning curve of audio-tactile guidance is steeper. The slower search times could be compensated by learning, as shown in previous work [267]. Furthermore, EyeSee360 was found to produce higher subjective frustration during use compared to the audio-tactile method. This effect was likely due to problems with stereoscopic disparities [221, 229] in which users were required to switch between different focal planes, namely visual guidance on the near plane and target locations at potentially far plane (see [36]). This change in vergence leads to visual fatigue [221] and might also affect target selection to some extent [119]. Focal disparities could also explain why Eyesee360 was able to achieve consistently good search performance despite higher visual loads due to sensory constraints (dense scene structure, vivid background colors and motions): Users might have partially faded out the potentially noisy background plane, allowing them to focus solely on the projection plane of EyeSee360 (compare to [68]). However, this behavior could raise new issues related to maintaining awareness of the environment when relying on visual guidance in AR [453]. To further investigate this issue, a visual secondary task was included in task load level 3. The secondary task was used to examine to what extent the level of SA could be influenced by the use of guidance methods under sensory constraints in AR. We showed that audio-tactile guidance performed significantly better in terms of secondary task targets hits (16.5% more correct hits) and reaction times (28% faster). This result could be explained by the consideration that visual clutter, further enhanced by visualization methods, can lead to attention tunneling, which diverts attention from other tasks [121, 192]. Although strategies have been developed to reduce visual clutter (e.g., [145, 148]), it remains a noticeable problem, especially on devices with a narrow FOV. Accordingly, it can be assumed that other visualization methods used for guidance can cause similar problems related to clutter (see [44]). Audio-tactile guidance offers the advantage of an unobstructed view without potentially cluttered visualizations but with

comparable reliability in finding the target. Moreover, focal disparities could remain a problem. By focusing on visual guidance cues, which are typically provided at the near disparity plane, details at more distant disparity planes may be easily overlooked by users. The use of audio-tactile guidance can reduce binocular disparities by eliminating the need for an additional visualization plane for target search. Substitution of visual localization information with auditory and tactile cues can minimize the potential for visual distraction [300].

In conclusion, we have shown that non-visual guidance is an effective method to support search under sensory constraints in AR. Compared to other current vibrotactile guidance approaches, audio-tactile guidance achieved significant improvements in latitudinal accuracy and search times while also providing precise depth cues. This feedback has shown to be highly useful when perception is affected by sensory constraints and can help resolve ambiguities caused by a dense scene structure, impaired depth perception, and a narrow FOV. In addition to the main approach, namely relative guidance using multisensory cues, absolute guidance cues were also investigated. The interpretation of absolute feedback has been shown to be more difficult and less accurate compared to guidance by relative feedback. However, both approaches may be used for different application scenarios. Relative guidance is more useful when precise and unambiguous target location or selection is required; for example, this could be the case for picking processes in larger warehouses (e.g., [371]). Absolute guidance, on the other hand, could be used in scenarios in which an estimate of the location is already sufficient to direct attention in that direction. For example, the use of absolute guidance could be envisioned in driving warning systems. In such scenarios, sensory warnings could direct attention to the general direction of the threat, for example, an approaching collision (see [289]). In this case, rapid but rather inaccurate indication of direction might be sufficient to direct the driver's attention to an imminent collision. Furthermore, it has been shown that non-visual guidance is competitive with well-performing visual guidance regarding target search hit rate. Although search times were not as rapid compared to visual methods, audio-tactile guidance produced significant improvements in dual task scenarios. Here, a higher number of correct selections and faster response times in the SA-related secondary task were achieved using audio-tactile guidance. This result suggests that audio-tactile guidance is more effective in settings in which SA needs to be maintained, for example in safety-critical domains [453]. Another important factor in terms of search performance is training. Although users were able to use the novel audio-tactile metaphors effectively

after some time, they were not able to compete with highly intuitive visual techniques in terms of search times. The question arises whether head-based non-visual feedback can lead to competitive search times compared to visual guidance after prolonged training or long-term use. However, it should be noted that the type of visual guidance technique affects the way users search for information in AR [44]. Therefore, the findings may only be partially transferable to other visual methods.

The results and findings from Chapters 4 and 5 informed further research, as can be seen in Fig. 7.2. The contributions described in Chapter 4 significantly informed the work presented in Chapter 5 by providing an evaluated, head-based, audio-tactile feedback system for guidance in AR. This allowed further investigation of this method in terms of sensory constraints, performance analysis, and impact on SA as described in Chapter 5. The multisensory feedback design from Chapter 4 was also used as the basis for designing multisensory proximity and transition cues detailed in Chapter 6. For this purpose, the feedback mechanisms were further adapted to be used as directional warning cues in AR under the influence of sensory constraints. In addition, the results presented in Chapter 5 provided valuable insights into the performance and SA capabilities of visual and non-visual guidance approaches, which proved useful for exploring the multisensory alerting methods presented in Chapter 6.

7.1.3 RQ_3 :What is the effect of multisensory guidance on situation awareness during search under sensory constraints?

Sensory constraints to be addressed: Depth perception, disparity planes, sensory thresholds, scene structure, background features, sensory noise, field of view, display properties.
Method: Head-based multisensory guidance in the form of proximity and transition cues to improve situation awareness.

As partially revealed in the discussion of RQ_2, multisensory guidance cues have the potential to enhance SA for AR applications. By providing combined *multisensory proximity and transition cues*, we have extended existing alerting methods (see [337] for a review). We have been able to achieve improved spatial and temporal perception of moving out-of-view objects in dynamic, information-rich scenes. This was achieved by providing information regarding the relationship between location, motion, and time of augmented information using multisensory guidance. Proximity cues provide spatial feedback on moving out-of-view objects while transition cues inform the user when an AR target has

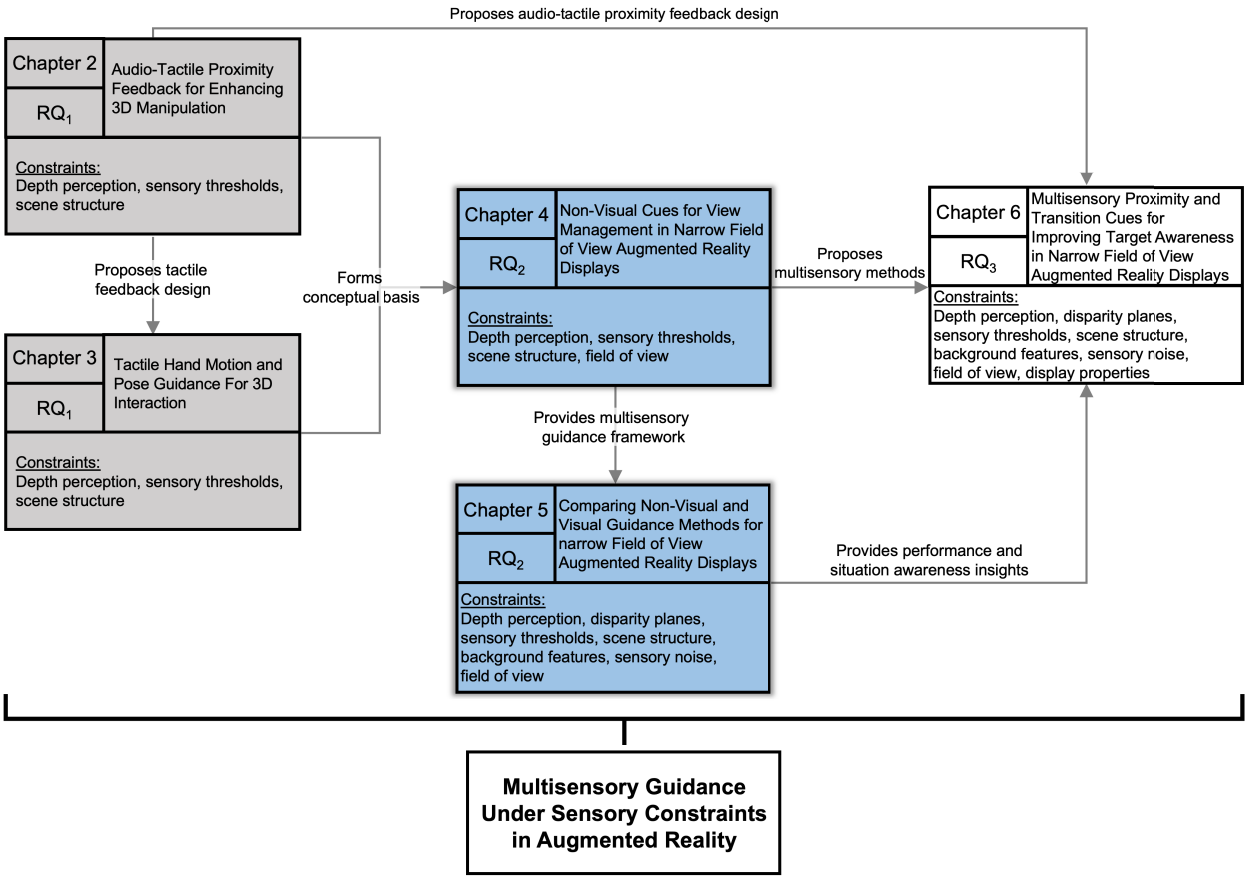


Fig. 7.2: Stages of this thesis: The highlighted Chapters 4 and 5 contribute to answering RQ_2. The results presented in these chapters affected the work presented in the following Chapters 5 and 6.

entered (or exited) the FOV. Such a sequential presentation of different parameters is novel compared to other works that have addressed improvements for SA in AR applications, typically used in safety-critical systems (e.g., in aviation [121] or in traffic as a vehicle driver [314] or pedestrian [194]), and in tasks that require interaction with the environment (e.g., [31, 182]).

In our studies, proximity and transition feedback consisted of visual, auditory, or tactile cues. The sensory cues were combined as a mode and presented sequentially to the user. The following notation is used to encode a mode: "Proximity modality–Transition modality". All modes were tested under different noise conditions, specifically taking into account sensory constraints in the form of sensory noise and background features; noise conditions were (1) laboratory "reduced noise" in the absence of visual distractions in the background and under office-like brightness and acoustical conditions and (2) outdoor-like "increased noise" providing vivid background colors and motions as well as increased levels of lighting and soundscape. In the context of increased noise exposure, higher ambient

lighting can also partially affect other sensory constraints such as display properties by degrading the contrast ratio of the display and causing reflections [221]. All main studies were examined with an OST AR HMD (Microsoft HoloLens, first generation) with the associated constraint of a narrow FOV. The general application scenario considered dense information scenes in which newly emerging information could be easily overlooked by users [104]. The main task of the study, hereafter referred to as awareness-task, consisted of (1) estimating the spatial position (angular distance, depth, direction, and speed) of an off-screen target based on proximity cues and (2) responding to the transition cue as soon as it entered the FOV.

The subjective preference for modes under both reduced noise and increased noise conditions was examined to provide insights regarding general usability. Bimodal proximity and transition combinations received generally higher user ratings than unimodal modes. This result is in line with previous work [234] showing that multisensory interfaces achieved higher user satisfaction, which is particularly useful for increased visual load [379]. However, somewhat contradictory, the audio-only proximity and transition mode was also rated reasonably well. This finding could be attributed to the fact that sonification strategies affect guidance efficiency in terms of speed, precision, or avoidance of target overshoot without a specific learning process (see [315, 348]). Another factor contributing to this effect could be cognitive load. It has been shown that bimodal cueing might require additional resources for cognitive processing compared to cue presentation using a single modality [301]. Furthermore, we assume that bimodal modes containing tactile cues would be preferred once sensory constraints (partially represented by the increased noise condition) further affect perception. In this case, information can be provided on an unconstrained channel when the visual or auditory modality is impaired. This effect was partially confirmed, as the Audio-Tactile was preferred over the Visual-Tactile mode at increased noise levels. For the Visual-Tactile mode, this result was likely due to poor visibility of the visualization methods, for example, when exposed to stronger lighting or when projections are presented against a distracting background [117, 129, 221]. This finding matches the ratings of the visual-only proximity and transition feedback, which received the lowest scores for both reduced and increased noise conditions. Of note, the Visual-Auditory mode also received adequate ratings, although perception of both sensory channels was likely affected by constraints of sensory noise. This indicates that even with impaired vision, well-perceptible auditory transition cues were considered helpful, supporting other findings on visual-auditory feedback (e.g., [200, 279]). Subjective measures

also suggest that modes with auditory proximity feedback can be highly useful for spatial localization when visual perception in AR is impaired by sensory constraints. From the subjective assessments, we can conclude the following: All variants of proximity and transition feedback were found to be suitable for estimating the direction and velocity of out-of-view targets and for informing about object transitions into the FOV as long as noise conditions were low. However, when higher levels of sensory constraints were present, caused by, for example, rich background features and increased sensory noise, users rated the visual-only mode as significantly less suitable. Finally, the combination of auditory proximity and tactile transition feedback was found to be particularly useful in conditions of both reduced and increased noise and received the highest overall preference scores.

The three best rated modes (Visual-Audio, Visual-Tactile, and Audio-Tactile) were then further investigated for their ability to improve SA in a divided attention task (focal attention task vs. awareness task). The focal attention task followed a go/no-go procedure in which participants were required to respond to a specific visual stimulus in the central visual field but were not allowed to respond to an alternative. The results showed that the Audio-Tactile mode produced the fastest response times in the awareness task compared to the other modes at both noise conditions. This is in line with other work [170] stating that auditory and tactile cues are inherently more alarming and lead to faster responses than visual cues. In direct comparison to visual proximity, our findings confirm previous work showing higher localization accuracy for moving audio sources than for visual spatial information (compare to [141, 146]). However, the Audio-Tactile mode did not differ significantly from the Visual-Tactile mode in reduced noise conditions, which performed second best. This result again demonstrates the usefulness and efficiency of visual guidance methods used for proximity feedback as long as perception is not or only slightly affected by sensory constraints. Further results showed that modes with tactile feedback allowed faster response times compared to auditory transitions, consistent with other findings comparing tactile and auditory cues [289, 293, 299]. With regard to the constraint of auditory noise conditions, it is recommended that tactile warnings should only be used in situations in which auditory warnings become ineffective, for example when auditory workload becomes too high [58]. We have shown that even under the influence of higher ambient auditory noise, auditory cues are still supportive next to tactile cues (compare to [52]). This effect suggests that combining tactile feedback is beneficial in distributing attentional processing in visuospatial tasks [430]. Moreover, there were no significant

differences in performance on the focal attention task across modes. This result is in contrast to those discussed regarding RQ_2, which showed that audio-tactile methods facilitated the concurrent visual task over visual guidance. However, this finding may be attributed to the use of a different visualization method (EyeSee360 [144]), which likely created more visual clutter and led to stereoscopic disparity issues compared to the method used for the visual proximity (EdgeRadar [153]). The EdgeRadar method has the advantage of providing cues in the peripheral view, which has proven to work relatively well for motion detection [280]. However, for increased noise, Visual-Tactile and Visual-Audio resulted in significant slowdowns in awareness task response times. Since both modes incorporated visual cues, this result suggests that visual-related constraints (e.g., illumination, background colors, textures, and motion, and thus to some extent display properties) have a stronger detrimental effect on AR perception than ambient auditory noise. In addition, performance degradations can probably be explained by subjective judgments that a higher mental demand is required when visual and auditory cues must be perceived and interpreted under sensory constraints (see [221]). However, the good performance of auditory cues even under ambient auditory noise could be attributed to the provision of signals in sensitive frequency ranges [459].

The results show that the combination of proximity and transition feedback had positive effects on SA. This was achieved by improving perception of moving out-of-view information and the transition of this information into the FOV. Furthermore, audio-tactile proximity and transition cues have been demonstrated to be particularly useful when perception has been impaired by sensory constraints. These insights are new compared to existing methods [337], which employ multisensory alerting methods but are focused primarily on providing redundant cues to support visual information. In addition, our results have extended existing visual approaches specifically addressing the improvement of SA in AR (e.g., [31, 194, 314]). In terms of SA improvements, we have shown that bimodal cues are generally preferred and considered more useful by users than unimodal cues. In general, all tested modes (Visual-Audio, Visual-Tactile, Audio-Tactile) demonstrated their usefulness in increasing SA when perception was not considerably impaired by sensory constraints. However, in the presence of stronger noise, the Audio-Tactile mode outperformed the Visual-Auditory and Visual-Tactile mode. Thus, SA seems to be more affected by sensory constraints when visual approaches are used as compared to tactile and auditory cues. This observation suggests that visual-only methods are not adequate to maintain SA when perception is affected by sensory constraints. However, performance may vary by

the choice of visualization method [44]. Various aspects such as display area (focal vs. peripheral) [280], mental demand, and visual load (see [44, 144, 153]) must be considered, as attentional resources of the visual system are limited [46, 232]. Further comparisons can be made with other approaches that use visual out-of-view guidance methods in AR (using EyeSee360 and 3D Radar) to improve SA and avoid pedestrian-vehicle accidents in a dual-task scenario [194]. The results related to RQ_2 have already suggested that non-visual methods can be beneficial for maintaining SA when perception is affected by intrinsic and extrinsic constraints [268]. This effect is likely to become more apparent the more (combined) sensory constraints exert an influence on perception [378]. Our results suggest that audio-tactile proximity and transition cues can maintain higher SA than visual-only methods when the task is performed under more influential sensory constraints. Another hypothesis concerns the engagement of the AR task with respect to SA. Previous studies have demonstrated that a shortage of SA can be observed as the immersion of the AR task increases [194]. We expect multisensory proximity and transition cues to counteract this trade-off by distributing attentional processing across other modalities with audio-tactile cues [430]. Finally, it is worth mentioning that multisensory proximity and transition cues have been studied to improve SA at level 1, as this level is mainly required for guidance and navigational tasks [274]. The extent to which multisensory guidance is applicable at higher levels of SA and how sensory constraints affect this perception remains to be investigated.

Overall, the findings presented in Chapters 2-6 provide insight into the capabilities of multisensory guidance under sensory constraints in AR and serve as a basis for discussion in answering RQ_{Main}, as can be seen in Fig. 7.3. The main research question regarding the potentials and limitations of multisensory guidance for search under sensory constraints in AR is discussed in the following subsection.

7.1.4 RQ_{Main} :What are the potentials and limitations of multisensory guidance to support search under sensory constraints in AR?

Based on the above discussion, we can address the overarching research question RQ_{Main} by examining the potentials and limitations of multisensory guidance under sensory constraints. We have shown that even under the influence of multiple sensory constraints that affect user perception, multisensory guidance is able to enhance task performance by reducing errors, providing high localization accuracy in search, and achieving higher levels of SA in AR. This improvement indicates that multisensory guidance has the potential

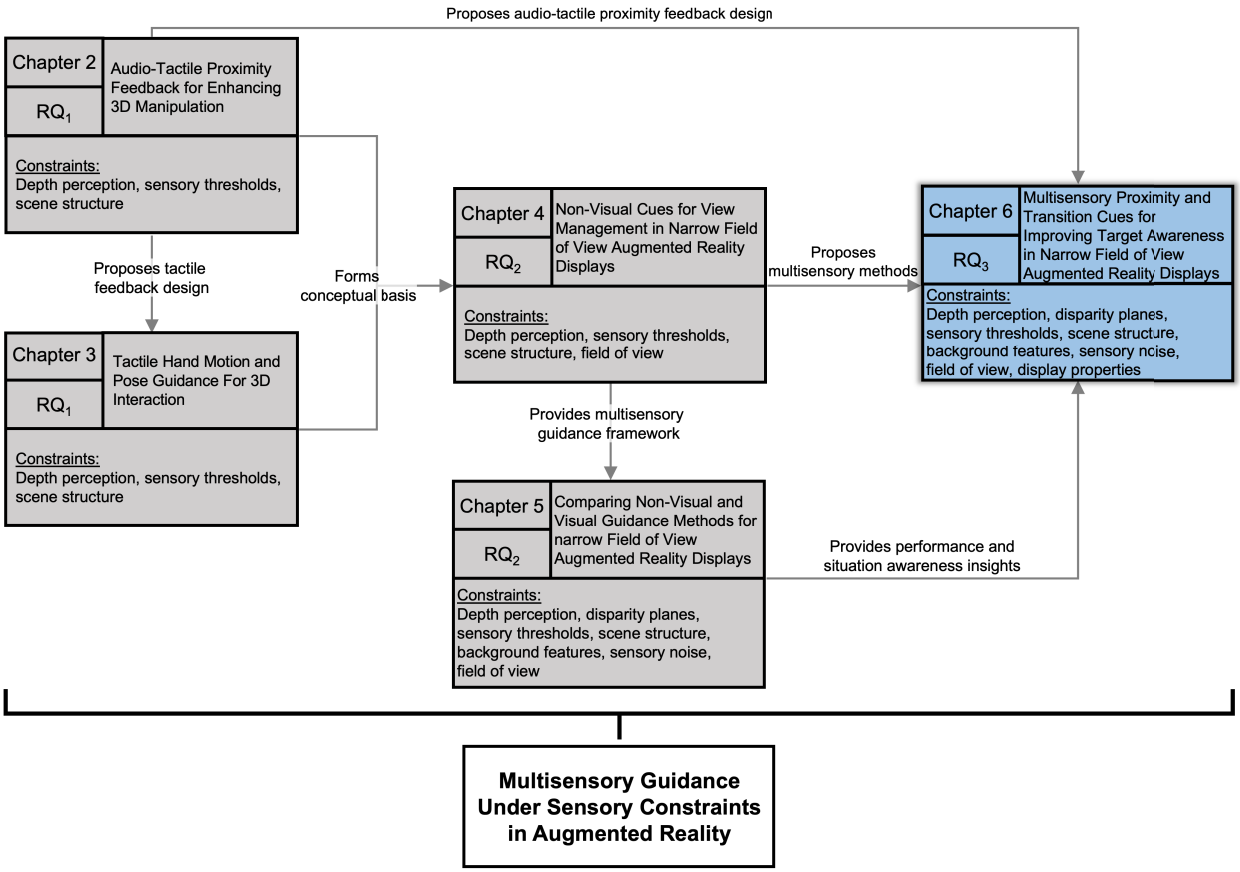


Fig. 7.3: Stages of this thesis: The highlighted Chapter 6 contributes to answering RQ_3. Chapters 2-6 have provided valuable insight into the capabilities of multisensory guidance under sensory constraints, serving as a basis for the discussion of RQ_{Main}.

to reduce detrimental effects on search performance caused by sensory constraints to some extent. Below, we discuss the effect of multisensory guidance under the influence of each intrinsic and extrinsic sensory constraint and describe the benefits and potential open issues of our approach. However, it is important to note that multiple sensory constraints can occur simultaneously when AR technologies are used; this is especially the case in uncontrolled outdoor scenarios [18]. The resulting impairments of combined sensory constraints may influence each other, which could also affect the effectiveness of multisensory guidance with regard to a particular constraint to some degree. After discussing each sensory constraint, the potentials and limitations of multisensory guidance are identified and summarized with respect to the corresponding constraint.

7.1.4.1 Intrinsic Factors. In this subsection, we discuss the influence of multisensory guidance on perception, which is constrained by intrinsic factors.

Sensory Thresholds. Multisensory guidance provides directional cues mainly in the auditory and tactile modality. It is important to consider human discrimination performance and the sensitivity of the corresponding modalities, as designers can consciously decide in which area high resolution is required.

Vibrotactile feedback imposes anatomical constraints [295]. This condition implies that locations with a high hair density should be avoided for vibrotactile feedback, as such areas can weaken the perceived stimuli. However, tactile frequencies should be chosen according to temporal and spatial processing characteristics of mechanoreceptors [76]. In terms of head-based tactile feedback, the forehead and temples have been found to be the most sensitive regions for presenting vibrotactile stimuli [89, 296]. Vibrotactors should be placed at least 15 mm apart to be clearly distinguished and signals do not interfere with each other [89, 296]. We used frequencies up to 200 Hz at the highest, which was moderately higher than the recommended 150 Hz stated in other work [89]. However, users reported high usability of our system without any particularly unpleasant feelings regarding vibrotactile cues. By temporarily using higher bandwidth, we expected to provide higher resolution for localization tasks, such as finding a target in depth using tactile cues. Finally, vibration patterns were clearly perceived by users. Therefore, they could be used effectively to indicate directions and distances even though they were sometimes interpreted inversely to the intended direction (e.g., backward instead of forward, see [380]). This observation suggests that vibration patterns need to be made more explicit in the future to resolve ambiguities [335].

To further exploit the perception of tactile thresholds, it seems reasonable to increase the spatial resolution of the existing tactile system, similar to other approaches that have investigated tactile stimulation through a high-density grid over the entire head [202]. It has been found that our current tactile setup (1D array consisting of 5 vibration motors placed in a line on the temples and forehead) is limited in terms of the perception of directionality of tactile cues. With this arrangement, it is currently only possible to effectively provide tactile directional cues in the horizontal plane without the need to modulate additional tactile parameters, as demonstrated in previous work [88]. Expanding the spatial resolution of the tactile setup would open up new possibilities to provide directional feedback through vibrotactile means, for example in vertical or diagonal directions. In addition, the extended tactile layout would allow the presentation of more complex tactile patterns that could be useful for directing attention in AR. However, previous studies have typically recommended the use of lower density arrays at the head consisting of up to five vibrotactors [91, 134].

The inclusion of other parts of the head, such as the back of the head [296], or facial areas such as the cheeks [133] may be considered. The occipital area of the head (lower-rear part of the skull) has demonstrated equal levels of sensitivity to vibration stimulation as the forehead. Adding stimuli on the back of the head to the current vibrotactile array could help provide more detailed spatial information by allowing stimuli to be displayed in a circular head array. In addition, tactile cues could be used for situation-specific warning signals, such as when information is approaching from behind the user. Regarding facial tactile stimulation, the tactile presentation of directional cues to the face is still an unexplored field [188]. While the face offers high spatial resolution in tactile perception, it also has complex, curved surfaces that can be easily deformed, making contact-based tactile representation difficult [188]. Furthermore, Pacinian corpuscles are generally absent in the facial area [133]. These receptors are responsible for detecting relatively high frequency vibrations near 250 Hz, which are commonly used for vibrotactile feedback. Thus, Meissner corpuscles, which are most sensitive at low frequency vibrations between 10 Hz and 50 Hz must be addressed for facial regions [133, 135]. Although these rapidly adapting receptors [135] respond strongly at the onset of the stimulus, the response diminishes rapidly over time. Thus, such stimuli ultimately provide little information regarding the duration of a static, long-lasting stimulus [133]. In addition, the pressure and tactile stimulation of the actuators on the user's face can cause increasing discomfort [319]. Alternatively, it would be conceivable to include tactile cues on the upper body or on the hands and arms (compare to the work presented in Chapters 2 and 3) in addition to the head-based tactile guidance feedback.

In terms of haptic technology, the developed system relies on the use of eccentric rotating mass (ERM) motors to provide vibrotactile stimuli. We chose this technology because ERM motors are small and easy to control while providing reasonably strong vibration [458]. However, compared to ERM, linear resonance actuator (LRA) technology offers several advantages, such as faster response times (delay and rise times) and decoupling of amplitude and frequency, which would enable more complex and detailed tactile mapping [339]. Thus, we expect that the use of LRA technology for vibrotactile feedback would support detailed haptic patterns and faster reaction times for guidance. This improved feedback could also be beneficial for the provision of tactile cues under exposure of potential tactile environmental noise. Piezoelectric actuators represent another haptic technology that is widely used. However, piezoelectric elements are expensive and require high voltages to operate [132], precluding a potential mobile, battery-powered setup. Due

to the rapid development of wearable actuator technology, there is now a growing variety of devices that can be worn, embedded in fabric, accessorized, or even tattooed directly onto the skin (see [73] for a review). Further research is needed to determine the extent to which such technologies can be used to optimize multisensory feedback, especially if feedback is to be provided at the user's head and under the influence of potential sensory constraints.

Regarding auditory cues, our feedback was always well perceived by users in the localization task for the selected frequencies between 300 Hz and 1300 Hz. However, some results have suggested that higher frequencies may lead to less accurate localization estimates (see [323]). This effect of reduced localization accuracy was also observed in the results of our absolute guidance approach described in Chapter 4. Thus, using frequencies up to 1000 Hz [323] may be sufficient for most localization tasks. In addition, it has been shown that sounds under 256 Hz to above 1024 Hz are perceived as more annoying [230], which could be used as a reference for long-term use of auditory guidance cues. We expect to improve localization accuracy further by using a personalized HRTF (P-HRTF) instead of a generalized HRTF (G-HRTF), as HRTFs can vary significantly from person to person [284]. Systems using a G-HRTF potentially lead to more localization errors, lateralization artifacts, and unconvincing spatial impressions [285]. It has shown that systems with P-HRTFs lead to more accurate localization by listeners [441] than those that rely on G-HRTFs, especially in terms of front-back confusions [436]. However, measuring the HRTFs of each potential user of a spatial auditory display is tedious and therefore may not always be feasible [436]. Alternatively, a longer training period with a non-individualized HRTF could also increase localization performance [284].

Crossmodal effects may also play a crucial role in the perception of multisensory guidance cues. We observed that certain multisensory feedback combinations were less accurate compared to unisensory feedback. This effect suggests an interaction between stimulus perception across modalities (see [301, 382]) causing cognitive range effects [329], which were more prominent in situations with fewer sensory constraints. In this context, crossmodal effects should be considered when designing cues that rely on sensory thresholds, as sensory cues may interfere with each other and thus affect performance (see [175]). Moreover, sensory overload is a critical factor in the provision of multisensory stimuli. Sensory overload can be caused by simultaneous overstimulation of multiple sensory systems, as human capacities are limited [247]. For example, a previous study showed that auditory and tactile stimuli may even have a negative effect in accuracy tasks,

as both stimuli were perceived as distracting and disruptive, leading to sensory overload [80]. Thus, the type of information presentation must be congruent with the type of processing characteristics required for task solution. Likewise, the allocated processing time and the amount of sensory information should be based on the limited processing capacity and the possibility of information overload [263].

In the following paragraphs, we summarize the potentials and limitations of multisensory guidance on sensory thresholds. The **potential of multisensory guidance on sensory thresholds** lies in the provision of head-based audio-tactile feedback. Vibrotactile cues at the temples and forehead as well as auditory cues are easy to perceive and well discriminable spatially and temporally. The sensory stimuli provided were subjectively rated as useful for guidance in AR; directional stimuli and patterns were mostly comprehensible to users and did not lead to sensory overload throughout the conducted studies. **The limitation of multisensory guidance on sensory thresholds** can be attributed to the low resolution of the vibrotactile setup, which limited the directional cues to predominantly horizontal directions on the tactile headband. In addition, the performance of the tactile modality in perceiving differences and absolute thresholds could likely be further improved through the use of new haptic technologies. The spatial accuracy of auditory cues is likely to be limited by the use of a G-HRTF, suggesting the use of P-HRTF in future systems. Finally, adverse effects of head-worn audio-tactile cues in long-term use, such as fatigue and discomfort, are not yet known and need to be further investigated.

Depth Perception. The incorrect interpretation of depth cues is still a common perceptual problem in AR [221]. A correct depth estimate can be useful to find the correct target in the midst of numerous distractors, especially in dense information spaces (see [267, 268]). While it is more common for visual methods to convey distances, for example, through color-coding or size-scaling [44], non-visual methods do not always provide depth cues [88] or do so only within close ranges of up to three meters (e.g., [34, 202]). However, visual guidance methods may create additional clutter, which can be further exacerbated by encoding distance, for example by scaling the arrow length in the 3D arrows method [44]. X-ray visualization techniques can support depth perception to locate occluded objects [142]; however, x-ray techniques introduce new depth ordering issues by mixing the spatial relationships between virtual and real objects in both direction and distance [221]. Compared to other approaches that are focused on short distance guidance (e.g., [34, 202, 393]), multisensory guidance can provide accurate depth cues

even for longer distances (tested up to a 30 meters target distance [268]). Combined longitudinal and latitudinal directional information helps to distinguish conflicting depth information by providing clearer localization information, which can be especially helpful when targets are partially occluded by distractors in scenes with high information density. By substituting visual information with multisensory cues, depth ordering problems can be partially reduced, which can improve layout and design in highly cluttered environments [221]. In particular, depth cues of multisensory guidance can be helpful in reducing typical problems with depth underestimation [221], as shown in our studies [267]. Moreover, the presented audio-tactile proximity feedback in this work has been shown to be useful for conveying distances to objects in the close surroundings [265]. Although multisensory guidance reduces ambiguities in depth localization, multiple objects that are nearly on the same depth plane will likely remain difficult to distinguish (see [221]) unless they differ sufficiently in at least one other dimension. Furthermore, the temporal resolution of our tactile system is potentially limited by the characteristics of the haptic technology and vibrotactile metaphors used for depth information. This condition leads to a discriminative limitation in terms of tactile frequency and pulse interval perception [124], which would presumably make it difficult to determine the smallest differences in depth with our approach. However, further studies are needed to investigate the potential of improving difference thresholds for depth perception through multisensory guidance.

Finally, in addition to the main approach of relative guidance, we have introduced experimental absolute guidance cues, including an absolute variant of depth feedback, as described in Chapter 4. Absolute depth feedback provides non-continuous distance feedback that is independent of head rotation. It has been shown that absolute depth feedback is less accurate than relative depth cues. However, we assume that absolute depth feedback could still be useful for navigation scenarios in which on-demand feedback that provides an approximate depth estimate (compare to [151, 177]), may already be sufficient.

The potential and limitations of multisensory guidance on depth perception are described below. **The potential of multisensory guidance on depth perception** lies in the provision of explicit non-visual depth cues to support the correct estimation of the target position in depth, even for more distant locations. This helps to reduce depth ambiguities and to improve depth ordering. In addition, typical depth-related problems such as underestimations in AR can be mitigated. **The limitation of multisensory guidance in depth perception** lies in the granularity of the depth cues provided. It is unlikely that minimal

depth variations can be reliably perceived by current tactile depth feedback. This, in turn, may affect depth ordering and selection performance when the target is at nearly the same depth as multiple distractors, for example, in scenes with high information density.

Disparity Planes. Disparity planes are related to different depth disparities at which the augmented content is observed [221]. Visual guidance cues are usually placed at a different disparity level than the augmented content to be searched, which can slow down search performance [36]. Moreover, the presentation of visual guidance information at the near disparity plane can lead to selection problems and higher frustration levels as users are forced to switch their vergence between different focal disparities [268]. This effect likely results in reduced SA, as users (possibly unintentionally) mask out more distant depth disparities and thus perceive their environment more unconsciously. Results of our studies suggest that disparity problems are promoted by near-plane visualizations. This particular problem can likely be reduced by multisensory guidance. The use of non-visual guidance information allows an unobstructed view without focusing on visualizations in the near disparity, which in turn reduces the need to switch between different focal disparities [180]; this has further proven to be beneficial in target selection and improving awareness of the environment [268]. This effect could also be attributed to potentially lower access costs resulting from less frequent switching between different planes as a result of multisensory guidance, which has a measurable impact on AR usability [36]. However, the disparity problem is likely to persist in stereoscopic systems, in which content can be displayed on multiple depth planes, as opposed to monoscopic systems, in which all content is displayed on a single depth plane [221]. One possible approach to further address disparity problems is to integrate eye-tracking techniques into multisensory guidance. Evidence has shown that eye vergence measurements can track the depth component of the 3D stereo gaze point [100]. These measurements could be used to adapt guidance cues as a function of current eye vergence and target object depth. For example, audio-tactile cues could be displayed for information at out-of-focus planes while the user is fixated on other stimuli at a nearby focal disparity, such as a visual cue. Although this approach seems promising, there are still problems in measuring eye vergence using eye-tracking techniques, such as high levels of noise [100] and inaccuracies [174] that require further attention. Finally, there are indications that selection problems caused by information on different focal disparities can be improved by using tactile depth cues [426]. These findings suggest that multisensory guidance could likely be used to target disparity-related issues. However, further research

is needed regarding the extent to which multisensory guidance can be adapted and possibly combined with other techniques to further reduce the impact of disparity problems.

The potential and limitations of multisensory guidance on depth perception can be summarized as follows: **The potential of multisensory guidance on disparity planes** refers to the ability to provide non-visual directional cues. Doing so reduces the need to switch vergence between different focal disparities, which can improve search and selection performance as well as the level of SA. **The limitation of multisensory guidance on disparity planes** suggests that disparity-related issues can be alleviated but not completely solved. Information at multiple focal depths will likely continue to affect how information is perceived in stereoscopic HMDs. Further research is needed to investigate how multisensory guidance can be combined with other approaches, such as eye-tracking techniques, to determine how the perception of information at different focal depths can be further improved.

7.1.4.2 Extrinsic Factors.

The effects of extrinsic sensory constraints on multisensory guidance are discussed below.

Scene Structure. Augmented environments can potentially contain a large number of virtual objects in different locations, making scene comprehension and navigation difficult [44, 221]. This can lead to a degradation of detection capacity over time when searching for infrequent targets embedded in more frequent, non-target distractions [159]. Using visual guidance methods to support target search in AR can add severe clutter to the view in visually complex scenes (compare to [44]). The advantage of multisensory guidance is that it can help untangle the cluttered environment by substituting visual information with non-visual cues. By using audio-tactile feedback, the user's view remains free of guidance-related cues, potentially reducing clutter and occlusion problems in visually complex environments. In this context, isolated auditory [279] and vibrotactile [88, 202] approaches have already shown adequate performance in supporting visual search without increasing visual complexity within the user's view. However, previous work has shown that combinatorial cues provide the best performance gains compared to auditory or tactile cues alone [159]. We have demonstrated that our multisensory guidance approach, which uses audio-tactile cues, contributes to effective target search in dense information spaces to find targets embedded amid more frequent non-target distractors. This approach could have further positive effects on search performance in dense information scenes

related to the FOV size of the AR device. In this regard, other studies have indicated that a high information density affects visual search, especially when using devices with a narrow FOV [406]. Furthermore, if information is spatially distributed over a large area, potential targets may not be captured by the AR device's FOV at any given time. This circumstance leads to augmentations located outside the FOV and remaining invisible to the user [44]. Additional methods are required to inform the user about the presence of this concrete information. Similar to visual [44] and other non-visual (e.g., [88, 202]) guidance approaches, multisensory guidance can effectively guide the user to locations that are located outside the FOV. However, multisensory guidance can also be used to draw attention to new information within and outside the FOV to increase SA [268, 407]. This approach is particularly useful in dense scene structures in which new information can be easily overlooked [104]. Here, multisensory guidance has been shown to be significantly better at maintaining SA in dual-task scenarios when compared directly to visual guidance [194, 267].

Assuming that multiple targets are to be found in dense information spaces, a current limitation of multisensory guidance is that only sequential search can be performed. However, other vibrotactile approaches are also limited to sequential search (see [88, 202]), likely due to a limited spatial resolution of vibrotactile cues. Although visual guidance methods are often capable of displaying multiple targets in parallel [44], the visual representation of targets (usually through the use of proxy symbols) often poses a problem by creating additional clutter and ambiguities. One possible approach to enable searching for multiple targets with multisensory guidance could be to present different object locations across different modalities. For the purpose, current audio-tactile guidance metaphors need to be reconsidered to allow the representation of multiple targets. This concerns the study of constraints on the perception of multiple visual, auditory, and tactile targets at the same time. For example, visual studies suggest that no more than about 60 items can be arrayed in the central 30 degrees of the visual field while still allowing attentional access to each individually [181]. On the other hand, researchers have shown that users are able to spatially separate up to four tonal signals [464]. The ability to discriminate multiple target objects in tactile systems is likely to depend strongly on tactile layout. We can assume that tactile grids with higher densities are able to discriminate more targets than systems with lower resolution. However, it can be expected that the close proximity of many tactile targets would lead to an overload of the tactile channel [51], implying that fewer targets should be presented in lower resolution tactile systems.

However, in addition to spatial resolution, tactile feedback parameters can be modulated to support discrimination of multiple targets in the tactile modality. Such approaches may include the modulation of duration, frequency, or amplitude [50] or the use of vibration patterns (e.g., [266, 335]). However, it should be noted that additional cognitive problems may arise when multiple targets are presented in a multisensory manner. Sensory overload may occur due to simultaneous overstimulation of multiple sensory systems [247]. Furthermore, due to the limited nature of the working memory, only a certain amount of information can be held at one time. Thus, information may be lost due to distraction or potentially inappropriate presentation [377]. Finally, when presenting multiple targets across sensory modalities, it should be considered that performance is based on the principle of maximum likelihood estimation [329]. As a result, stimuli with a more reliable estimate will have a greater influence on the final perception. This condition implies that probably not every sensory representation of a target (visual vs. auditory vs. tactile targets) is processed equally but rather is subject to a dominance order [49].

The potentials and limitations of multisensory guidance on scene structure are described below. **The potential of multisensory guidance on scene structure** is to substitute visual information by auditory and tactile cues; the provision of audio-tactile cues can partially reduce clutter and occlusion problems in visually complex scenes. In addition, multisensory guidance allows effective target guidance for information that is outside the FOV of the AR device due to its density and distribution in complex scenes. **The limitation of multisensory guidance on scene structure** is the ability to track only one source of information at a time. In dense scene environments, however, it might be necessary to search for multiple information sources at the same time. This raises the question of how to display multiple sources of information in an effective multisensory manner.

Background Features. Background features refer to color and texture as well as motion in the background. The color and texture of the target background can affect the legibility of the augmented content, leading to perceptual distortions and interpretation problems that can affect user performance [129, 221]. In addition, background motion can cause temporal occlusions as well as distractions from the actual search task [118, 267]. However, we could not find significant effects of background features on search performance for both visual guidance and multisensory guidance, compared to conditions without background features. Both approaches showed a consistently high hit rate even when targets were affected by background colors and motions [267]. This result shows

that search using visual guidance methods has become reasonably efficient [44] even under the influence of background features. However, it is expected that the effects of color and texture would become more apparent under high illumination [130] or changing lighting conditions [129, 221]. This effect became apparent in the study described in Chapter 6, in which modes containing visual cues were significantly more impaired than non-visual mode combinations under constrained conditions with higher illumination and background features. We assume that a higher degree of background features (nearly identical color and texture as the target, a high optical flow) also enhances the impairment of other sensory constraints and thus exerts a stronger influence on the perception of augmentations. Even at higher levels of background features, multisensory guidance would likely remain unaffected by perceptual impairments such as reduced visibility and legibility [130, 221], clutter, and occlusion by motions [117, 118], as non-visual cues convey information to determine the target location.

Furthermore, background distractions caused for example by motion [118], are considered to be on different disparities. Information located at different depth planes, such as an augmentation in the foreground and a distraction in the background, likely promotes a shift in vergence [221] (see Subsection 1.2.1.2 “Disparity Planes”) and thus could also affect task performance and SA [198]. This switching is likely to be further promoted by the presentation of visual guidance methods at nearby focal levels, as indicated by the results presented in Chapter 5. Regarding the effects of focal switching due to visual methods, we have already shown that multisensory guidance appears promising for maintaining SA during AR search by reducing visual information provided at different depths [267]. This effect can potentially save access cost and thus support the usability of the AR system [36]. We further hypothesize that multisensory guidance can support users in maintaining their attentional capacity for target search, even when visual distractions are present in the background [382]. However, studies have shown that physically salient distractions that are otherwise irrelevant to the current task can interrupt top-down control and capture attention (called “attentional capture”) [115]. Therefore, it is likely that distractions from background motion that are particularly salient may continue to affect search performance. Problems may also occur when the target is moving amidst background motions [118]. In this case, it would likely be difficult to localize the target with multisensory guidance unless the target movements are asymmetric to the background motions [357]. These detection capabilities are attributed to special motion detectors in the visual system that act

as high-pass filters in the velocity domain [353]. However, further research is needed to determine exactly how background motion affects perception in AR.

We summarize the potentials and limitations of multisensory guidance for scene structure as follows: **The potential of multisensory guidance on background features** is that the provision of non-visual guidance information is likely to remain unaffected by background colors, patterns, and motions. This condition promotes search performance and maintenance of SA compared to the visualization of guidance information that can potentially be influenced by background features. **The limitation of multisensory guidance on background features** addresses the issue that the presence of salient distractions, for example caused by motion in the background, may still attract the attention of the user to that location and thus affect search performance. In addition, it is expected that detection of moving targets in the midst of a moving background would prove difficult as long as distractors are moving in the same direction.

Sensory Noise. A bright environment can affect legibility, making augmentations partially or completely invisible if the illuminance exceeds the capabilities of the display [110, 221]. Despite continuous improvement in display technologies, strong environmental illumination remains a critical aspect for displaying AR content in OST devices [221]. Compared to visual guidance methods, multisensory guidance can be used to support the search task regardless of current lighting conditions because sensory cues are provided in audio-tactile modalities. In general, we have shown that multisensory methods involving the tactile modality proved particularly useful when users were exposed to sensory noise [268, 407]. It can be assumed that even purely tactile methods may already work reasonably well when vision is impaired in AR. This assumption is supported, for example, by findings regarding tactile guidance methods using sensory substitution techniques to assist visually impaired people during navigation (see [86]). However, studies have shown that combined audio-tactile cues result in higher search accuracy than tactile cues alone [175], showing promise for aiding search when vision is impaired by sensory noise. Furthermore, results described in Chapter 6 demonstrate that audio-tactile methods facilitate faster response times compared to other cross-modal combinations when vision is partially impaired due to increased light intensity.

The auditory guidance cues in this work were designed to be perceived easily by users even under conditions of increased ambient auditory noise. However, the degree to which auditory localization abilities are affected depends on the frequencies and intensity of

auditory noise [4]. Thus, auditory guidance cues may become inaudible if the ambient auditory noise level exceeds a certain threshold. In this context, studies have shown that audio feedback performance degrades when user are exposed to 94 dB or more of environmental auditory noise [172]. This level of auditory noise intensity could be used as an indicator to replace potentially inaudible audio cues with a more appropriate sensory presentation in a less affected modality. However, the audibility of a sound depends on both sound pressure level and frequency [135]. Therefore, the absolute auditory threshold for the audio component of multisensory guidance needs to be determined in further studies.

It seems obvious to use noise-cancelling headphones for the provision of auditory cues to overcome impairments caused by auditory noise. Studies have shown that using noise-cancelling headphones can improve task performance for both single- and dual-task situations in noisy environments [290]. This can be beneficial in reducing distractions caused by noise or engagement in tasks other than the target task. However, the exclusion of the environment by removing external sound also leads to filtering of potential important auditory information, such as alarm signals and traffic sounds. Thus, all noise sources in the environment are suppressed in the same way regardless of their aesthetics and relevance. While this strategy allows users to hear the desired audio source, it can significantly impair SA [240] and can lead to considerable safety risks [66] in mobile situations, such as walking outdoors [156]. A potential approach to prevent the decrease of SA is to obtain "smart noise-cancelling" [304]. Sounds from the environment are analyzed so that only selected frequencies, such as certain alarm sounds (e.g., [367]), are transmitted to the user [304]. However, noise-cancelling is ineffective for sound frequencies above 1000 Hz, leading to a leakage of certain sounds. Another problem is that objects can flexibly change their appearance at any time, which makes filtering in highly dynamic environments difficult [304]. An alternative to noise-cancellation methods is so-called Hear-Through AR [245]. Here, a bone-conduction headset is used to deliver augmented sounds through the bony structure of the skull while environmental sound is perceived through the unoccluded ear canals. Bone-conduction headsets have been shown to enable reliable spatial separation, but they probably cannot be used to lateralize signals to apparent locations at the extreme left or right [431]. In addition, the use of bone-conduction headsets might require special or individualized HRFT to work well for localization tasks [245, 427].

In our work, the tactile modality remained largely unaffected by sensory noise. However, there are (typically outdoor) scenarios in which tactile noise can exert an influence on

the perception of tactile stimuli. Tactile noise can be caused by background gravitational acceleration, for example inside a moving vehicle [59] or train [172]. Such exposure to ambient tactile noise can potentially affect tactile perception thresholds and lead to frequency interactions with externally provided vibration cues [296]. Other work has found that tactile feedback performance decreases at environmental vibration levels of approximately 9.18 g/s or higher [172]. This finding suggests that multisensory guidance cues should be provided on less affected sensory channels when tactile environmental disturbances exceed this noise level. However, the effect of tactile noise on multisensory guidance needs to be investigated more closely in future work.

Finally, outdoor conditions can be highly dynamic not only in terms of lighting [18, 221] but also regarding ambient auditory and tactile noise levels [4, 33, 172]. Such situations can include, for example, temporal changes in information density, illumination levels, and auditory noise, which can affect search performance in different ways. This effect suggests that future systems should be able to adaptively switch the transmission of sensory stimuli to the most appropriate (i.e., the least impaired) sensory channel. However, an adaptive sensory switching system might create new cognitive challenges. Previous research has shown that switching from one modality to another during perceptual processing is associated with processing costs [318]. This condition leads to the so-called "modality switch effect", which states that reaction times increase when the stimulus is preceded by a stimulus of a different modality [137, 381]. The problem could be further exacerbated by the need of frequent modality switching under conditions of highly dynamic sensory noise. Thus, switching behavior could affect cognitive performance, as processing capacities could be exceeded [82], for example, by switching cues too quickly. This issue creates the need for intelligent switching algorithms to keep the cognitive load of modality switching as low as possible. Integration of machine learning approaches [6] would be conceivable for this purpose and could enable modality switching depending on incoming sensory noise, user behavior, and user preference.

The potentials and limitations of multisensory guidance on sensory noise can be described as follows: **The potential of multisensory guidance on sensory noise** lies in the presentation of guidance information in the audio-tactile modality. In particular, audio-tactile guidance cues that we have developed have been shown to be less affected by sensory noise compared to other sensory combinations. This advantage facilitates effective search guidance and the maintenance of SA, even under the influence of higher levels of illumination and auditory ambient noise. **The limitation of multisensory guidance**

on sensory noise concerns the degree of sensory interference. Excessive auditory noise can still lead to impairments of auditory guidance, including inaudibility of auditory cues. In addition, tactile interferences have not yet been investigated, which leaves an open question regarding how robust tactile guidance cues are under conditions of tactile noise. Furthermore, it is likely that conditions of highly dynamic noise could have a stronger influence on the perception of audio-tactile cues. This assumption suggests that a dynamic switching process of multisensory guidance cues in the form of visual, auditory, and tactile stimuli would be desirable.

Field of View. A narrow FOV remains a common constraint even for many modern OST AR devices [210, 221]. Although the FOV of AR devices is likely to become wider in the future, it remains challenging to build displays that can fully cover the human visual field [205]. Therefore, associated problems such as clutter, occlusion, and a potentially higher workload when visual methods are used [44] are likely to remain even with future devices that have a larger FOV (see [208]). Although previous work has indicated that a wider FOV likely increases search and task performance (e.g., [83, 406]), it has not always led to better performance [409]. Furthermore, modern visual guidance approaches are usually placed in the foveated vision of the FOV (see [44, 146]), causing additional clutter. This condition could affect the performance of a task that requires attention in the central visual field [208]. The main advantage of multisensory guidance is that sensory cues can generally be presented regardless of the FOV size. By substituting visual information, visually related issues such as clutter and occlusion can be reduced in both narrow and wider FOV displays. In our studies, this effect has been demonstrated to be beneficial for task performance, including in situations that require a secondary task to be performed within a narrow FOV [268, 407]. In addition, multisensory guidance has been shown to be effective in directing attention to targets that are located outside the limited FOV. In this context, audio-tactile cues for out-of-view guidance have been shown to be as reliable as well-performing visual methods in terms of hit rate while providing higher selection performance. Multisensory guidance exhibits significantly longer search times than its visual counterparts [268]. The slowdown in search time could also be related to the FOV size and the associated need to search for spatially distributed, temporarily non-visible information outside the FOV. We assume that multisensory guidance would provide faster visual target detection times when targets are displayed within the FOV. This prediction would be in line with other crossmodal research that has observed faster response times

for visual target detection tasks in cluttered scenes [175, 414, 415]. However, the specific effects of multisensory guidance on FOV sizes and in-view versus out-of-view target guidance needs to be addressed in further studies.

It is hypothesized that advantages of multisensory guidance may be partially mitigated in the future by the availability of wide FOV displays and advances in visual methods. The latter include clutter reduction strategies [145, 148] that can reduce or compress visual guidance information to free up limited screen space to some extent. In addition, subtle visual methods that exploit the periphery have the potential to keep the focal area free for performing concurrent visual tasks. Examples for such methods include screen flickering approaches [347] or radial light displays around the eyes [150]. However, visual peripheral cues may force users to switch frequently between central and peripheral areas in scenarios that require maintaining attention in the central visual field, as indicated by the results presented in Chapter 6 (compare to [208, 407]). This effect could be exacerbated for larger FOV displays, potentially leading to performance degradation due to increased access costs [36]. We have shown that this effect can be potentially avoided by using non-visual guidance. Our results suggest that audio-tactile cues provided better perception and faster reaction times to out-of-view objects, while users were presumably more able to focus their visual attentional resources on the central attention-demanding task. This result is consistent with other work in which dual task conditions performed crossmodally resulted in increased task performance compared to concurrent tasks performed unimodally [365].

Of particular note are the transition cues presented in Chapter 6. Transition cues were designed to be presented during the transition of information into the FOV. Thus, in contrast to the general multisensory guidance approach presented in Chapters 4 and 5, transition cues depend on the FOV size of the AR device. With respect to wide FOV devices, we assume that information transitions are likely to occur more frequently due to wider FOV boundaries. This condition poses the risk of stimulus overload, which can have a negative impact on perception, attention, and cognitive performance [102, 247]. Even though multisensory guidance methods will likely be straightforward to apply to wide FOV devices in the future, sub-approaches such as multisensory proximity and transition cues may need to be adapted to systems with a wider FOV. This includes, for example, the integration of filtering or clustering methods to meet the emerging requirements.

The potentials and limitations of multisensory guidance on the field of view are described below. **The potential of multisensory guidance on the field of view** lies in the

substitution of visual information. This strategy can contribute to a reduction in visual complexity within the (currently still typically narrow) FOV, potentially reducing clutter and occlusion of information. This has been shown to be beneficial for performance in search tasks on narrow FOV displays and perception of targets that are out-of-view. **The limitation of multisensory guidance on the field of view** refers to significantly longer search times to find out-of-view information compared to visual approaches. In addition, it is conceivable that the advantages of multisensory guidance could be partially compensated by future development of visual guidance techniques and advances in display technologies, such as an increase in FOV size. Finally, if multisensory guidance techniques are offered based on FOV size (see the multisensory proximity and transition cues described in Chapter 6), future wide FOV displays may pose new problems related to perception and cognition of the sensory cues that are provided.

Display Properties. The quality of display devices is constantly being improved [221]. Therefore, it can be expected that new display technologies and approaches that provide improved brightness and contrast ratios will be available. A current example in this regard is the use of an additional transparent display to regulate the transmission of external light [204]. In addition, the integration of more effective backlighting and anti-reflective coatings will likely further reduce the susceptibility of displays to reflections [221]. However, it can be assumed that especially in typical outdoor conditions, for example under higher illumination levels [18, 221], the perception of augmentations would remain impaired to some extent (see [110]). The threshold at which visual aids become impaired will potentially increase, making the use of visual methods more flexible even under conditions of higher ambient lighting. Nevertheless, the use of multisensory guidance is considered a fruitful approach even in the context of more advanced display technologies. By replacing visual content with audio-tactile cues, multisensory guidance can help overcome visual impairments in search caused by display properties. Finally, improvements in display capabilities would open new possibilities for multisensory guidance composed of visual, auditory, and tactile cues that would remain more robust to combined sensory constraints.

The potentials and limitations of multisensory guidance on the display properties can be described as follows: **The potential of multisensory guidance on display properties** is that audio-tactile cues can be presented regardless of current display property impairments, such as low contrast ratios and reflections on the screen. Therefore, multisensory guidance

remains useful even when perception is impaired by the effects of display properties. Advances in display technologies, such as potentially higher contrast ratios, present new possibilities for combining visual, auditory, and tactile cues under constrained conditions. **The limitations of multisensory guidance on display properties** can be indirectly inferred, as these primarily affect visual perception. Display properties are likely have less influence on visual perception in the future due to technological improvements, suggesting that visual methods may be sufficient to support general search tasks in situations with moderate impairments of display capabilities due to sensory constraints.

7.2 Conclusion

The perception of augmentations can be affected by intrinsic and extrinsic factors of sensory constraints, leading to degradation in search performance in AR environments. In this work, we presented a novel multisensory guidance approach to effectively support search under the influence of sensory constraints in AR. We have shown that the detrimental effects of sensory constraints on perception can be partially mitigated and search performance can be improved. However, the use of multisensory guidance also raised new open questions that could not be fully addressed within the scope of this thesis and require further research. Overall, we have highlighted the potentials and limitations of our novel multisensory guidance approach in supporting search under sensory constraints. Despite future advances in AR technology, sensory constraints will continue to represent a significant factor to consider in the perception of augmentations. Therefore, our multisensory guidance approach will remain useful to support search in future AR systems.

In Chapter 2, we have examined the effect of audio-tactile proximity cues for hand touches and movement sequences in the context of exploration and manipulation actions in dense information spaces. Our findings provide positive indications that the two proposed proximity models can enhance spatial awareness by reducing the number of object collisions and errors. In addition, performance improvements were evident when a higher resolution tactor grid was used compared to lower resolution approaches. Our results provide useful insights regarding how audio-tactile proximity cues can improve 3D interaction when perception is impaired by a complex scene structure.

In Chapter 3, we have investigated how tactile patterns can improve hand motor planning and action coordination by guiding hand motions and postures. We have shown that individual tactile cues, as well as tactile cue patterns distributed over multiple vibrotactors

on the hand and forearm, can be well localized and interpreted by users. Furthermore, we have provided insights that vibration patterns support finer-grained 3D selection and manipulation tasks, which are particularly useful when information density of the scene increases.

In Chapter 4, we have described the design and evaluation of different variants of audio and vibration cues on the user's head to support target guidance under sensory constraints. Guidance cues were provided as a relative function of the head orientation and the spatial location of the target. These cues helped users to find digital information by guiding them to the information using audio-tactile cues without relying on additional visual aids. The results showed that users were able to judge the position of the information with our reference method precisely, with only minor deviations between the judged position and the real location. We found that audio-tactile guidance was superior to other sensory combinations tested in terms of search time and accuracy. In addition, we investigated the usefulness of absolute guidance cues. Although absolute guidance cues generally performed well, they were less accurate than relative guidance cues. Finally, the results demonstrated the usefulness of audio-tactile guidance to improve search under sensory constraints, such as when perception is impaired during search in dense scene structures with a narrow FOV AR device.

In Chapter 5, we have described the comparison of our audio-tactile reference method against a popular visual guidance method called EyeSee360. In this way, we aimed to investigate the differences in the effectiveness of both guidance methods in AR search under sensory constraints and their impact on SA. Under different levels of task load, we have shown that search under the influence of sensory constraints, such as a narrow FOV, auditory ambient noise, and background motions, can benefit from multisensory guidance. Although search times were slower, multisensory guidance performed comparably well to EyeSee360 in finding virtual information while generally maintaining a higher level of SA.

In Chapter 6, we addressed the condition that users often divide their attention between different tasks when interacting with AR. In such situations, relevant or newly emerging information can easily be missed, especially in information-rich and dynamic environments. To address this problem, we developed and evaluated combined visual, auditory, and tactile guidance cues for our multisensory head-mounted system, called proximity and transition feedback. Proximity feedback provided guidance cues regarding the location of augmented information that emerged outside the FOV. Transition feedback informed the user that

an augmented information had entered the visible area of the AR display. The results demonstrated that users were able to react promptly to newly emerging information using multisensory proximity and transition feedback, effectively supporting SA under perceptually constrained conditions in AR. In particular, the audio-tactile mode was found to be the most effective among all sensory combinations tested under conditions with high influence of sensory constraints.

In Chapter 7, we have addressed the research questions raised in Chapter 1 to identify the potential and limitations of our multisensory guidance approach to support search under sensory constraints in AR. We anticipate that the insights gained from this work will help researchers and practitioners to address sensory constraints in AR and make search guidance more efficient in future systems. Finally, we recognize great potential for multisensory guidance to be applicable in other fields, such as virtual reality, teleoperation, video games, and automotive safety systems.

7.3 Outlook and Future Work

Future work will consider refinements to the audio-tactile system to address the limitations of multisensory guidance under sensory constraints identified previously.

Potential extension possibilities include increasing the number of vibration elements in the tactile array and the use of new haptic technologies. Furthermore, it would be of interest to combine our audio-tactile approach with modern visual guidance methods to potentially improve versatility and efficiency for guided search. Such enhancements would also open new possibilities for presenting multiple spatially distributed targets through visual, auditory, and tactile means. In this context, the investigation of dynamic multisensory switching would also be a promising approach. The proposed system would measure the influence of currently occurring sensory constraints, such as incoming illumination and auditory noise, on perception in AR. Multisensory guidance cues could be switched accordingly and provided in the optimal modality – typically the modality that is currently least affected by constraints. In addition, it would be useful to explore the use of machine learning approaches to optimize the use of audio-tactile cues to support search guidance. By analyzing data regarding user behavior and performance, machine learning algorithms could be trained to adapt sensory guidance provided to individual users in real time, considering the current environmental conditions as well as the user's needs and preferences. This strategy could lead to more effective and efficient search guidance systems in AR.

Eye-tracking technologies also have the potential to improve multisensory guidance by providing additional information regarding the user's attentional focus and visual search behavior. Analysis of eye-tracking data could provide a better understanding of how users interact with AR environments under sensory constraints. The insights gained could help optimize the provision of multisensory guidance, and thus improve user behavior and performance under the influence of sensory constraints.

List of Figures

List of Tables

References

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

[2] C. Afonso and S. Beckhaus. How to not hit a virtual wall: Aural spatial awareness for collision avoidance in virtual environments. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, AM '11, pages 101–108. ACM, 2011.

[3] 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, pages 335–342. ACM, 2009.

[4] K. A. Alali, J. G. Casali, et al. The challenge of localizing vehicle backup alarms: effects of passive and electronic hearing protectors, ambient noise level, and backup alarm spectral content. *Noise and Health*, 13(51):99, 2011.

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

[6] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan. Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, 16(4):1996–2018, 2014.

[7] R. B. Ammons. Effects of knowledge of performance: A survey and tentative theoretical formulation. *The Journal of General Psychology*, 54(2):279–299, 1956.

[8] F. Argelaguet, A. Kulik, A. Kunert, C. Andujar, and B. Froehlich. See-through techniques for referential awareness in collaborative virtual reality. *International Journal of Human-Computer Studies*, 69(6):387–400, 2011.

[9] N. Ariza, P. Lubos, F. Steinicke, and G. Bruder. Ring-shaped haptic device with vibrotactile feedback patterns to support natural spatial interaction. In *ICAT - EGVE '15 Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*, 10 2015.

[10] 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)*, 2018.

[11] K. Arthur and F. Brooks Jr. *Effects of field of view on performance with head-mounted displays*. PhD thesis, University of North Carolina at Chapel Hill, 2000.

[12] M. Auvray. Multisensory and spatial processes in sensory substitution. *Restorative Neurology and Neuroscience*, 37(6):609–619, 2019.

[13] J. W. Ayers, E. C. Leas, M. Dredze, J.-P. Allem, J. G. Grabowski, and L. Hill. Pokémon GO—A New Distraction for Drivers and Pedestrians. *JAMA Internal Medicine*, 176(12):1865–1866, 12 2016.

[14] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001.

[15] R. Azuma, M. Billinghurst, and G. Klinker. Special section on mobile augmented reality, 2011.

[16] 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, pages 66–, Washington, DC, USA, 2003. IEEE Computer Society.

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

[18] R. T. Azuma et al. The challenge of making augmented reality work outdoors. *Mixed reality: Merging real and virtual worlds*, 1:379–390, 1999.

[19] 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.

[20] 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*, volume 56, pages 1431–1435. SAGE Publications Sage CA: Los Angeles, CA, 2012.

[21] R. Bane and T. Hollerer. Interactive tools for virtual x-ray vision in mobile augmented reality. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 231–239. IEEE, 2004.

[22] K. Bark, E. Hyman, F. Tan, E. Cha, S. Jax, L. Buxbaum, and K. Kuchenbecker. Effects of vibrotactile feedback on human learning of arm motions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(1):51–63, 2014.

[23] K. Bark, P. Khanna, R. Irwin, P. Kapur, S. A. Jax, L. Buxbaum, and K. Kuchenbecker. Lessons in using vibrotactile feedback to guide fast arm motions. In *World Haptics Conference (WHC), 2011 IEEE*, pages 355–360. IEEE, 2011.

[24] P. W. Battaglia, M. Di Luca, M. Ernst, P. R. Schrater, T. Machulla, and D. Kersten. Within- and cross-modal distance information disambiguate visual size-change perception. *PLOS Computational Biology*, 6(3):1–10, 03 2010.

[25] 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.

[26] M. Bear, B. Connors, and M. A. Paradiso. *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Jones & Bartlett Learning, 2020.

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

[28] M. Beck, M. Lohrenz, and J. Trafton. Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of experimental psychology. Applied*, 16:238–50, 09 2010.

[29] S. Beckhaus, F. Ritter, and T. Strothotte. Cubicalpath-dynamic potential fields for guided exploration in virtual environments. In *Proceedings the Eighth Pacific Conference on Computer Graphics and Applications*, pages 387–459, 2000.

[30] 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, pages 101–110. ACM, 2001.

[31] B. Bell, T. Höllerer, and S. Feiner. An annotated situation-awareness aid for augmented reality. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 213–216, 2002.

[32] H. Benko, C. Holz, M. Sinclair, and E. Ofek. Normaltouch and texturetouch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 717–728. ACM, 2016.

[33] B. Berglund, T. Lindvall, D. H. Schwela, W. H. Organization, et al. Guidelines for community noise, 1999.

[34] 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, pages 31–38. ACM, 2015.

[35] P. Bertelson. Ventriloquism: A case of crossmodal perceptual grouping. In *Advances in psychology*, volume 129, pages 347–362. Elsevier, 1999.

[36] N. Binetti, L. Wu, S. Chen, E. Kruijff, S. Julier, and D. P. Brumby. Using visual and auditory cues to locate out-of-view objects in head-mounted augmented reality. *Displays*, 69:102032, 2021.

[37] F. Biocca, A. Tang, C. Owen, and X. Fan. The omnidirectional attention funnel: A dynamic 3d cursor for mobile augmented reality systems. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, volume 1, pages 22c–22c, 2006.

[38] 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.

[39] J. Blake and H. B. Gurocak. Haptic glove with mr brakes for virtual reality. *IEEE/ASME Transactions on Mechatronics*, 14(5):606–615, 2009.

[40] 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*, pages 1–9, 2018.

[41] A. Bloomfield and N. Badler. Virtual training via vibrotactile arrays. *Presence: Teleoperators and Virtual Environments*, 17(2):103–120, 2008.

[42] A. Bloomfield, Y. Deng, J. Wampler, P. Rondot, D. Harth, M. McManus, and N. Badler. A taxonomy and comparison of haptic actions for disassembly tasks. In *Virtual Reality, 2003. Proceedings. IEEE*, pages 225–231. IEEE, 2003.

[43] 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, editors, *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 49–62, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[44] 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.

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

[46] T. Bøttern, A. Hansen, R. Höfer, and C. E. Pellengahr. Selective attention in augmented reality. *Aalborg University Copenhagen. Available online at: http://schnitzel.dk/rasmus/files/med7paper.pdf*, 2009.

[47] M. Bouzit, G. Burdea, G. Popescu, and R. Boian. The rutgers master ii-new design force-feedback glove. *IEEE/ASME Transactions on mechatronics*, 7(2):256–263, 2002.

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

[49] J.-P. Bresciani, M. O. Ernst, K. Drewing, G. Bouyer, V. Maury, and A. Kheddar. Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental brain research*, 162(2):172–180, 2005.

[50] 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*, pages 15–23. Australian Computer Society, Inc., 2004.

[51] S. A. Brewster, S. Wall, L. M. Brown, and E. Hoggan. Tactile displays. *The Engineering Handbook on Smart Technology for Aging, Disability and Independence*, pages 339–352, 2008.

[52] J. C. Brill, B. D. Lawson, and A. H. Rupert. Audiotactile aids for improving pilot situation awareness. In *18th International Symposium on Aviation Psychology*, page 13, 2015.

[53] G. Burdea. *Force and Touch Feedback for Virtual Reality*. John Wiley & Sons, Inc., 1996.

[54] 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*, pages 239–246, 2006.

[55] 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*, pages 93–102, 2012.

[56] 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*, pages 108–117, 2006.

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

[58] J. L. Campbell, C. M. Richard, J. L. Brown, and M. McCallum. Crash warning system interfaces: human factors insights and lessons learned. *DOT HS*, 810:697, 2007.

[59] Y. Cao, F. Van Der Sluis, M. Theune, R. op den Akker, and A. Nijholt. Evaluating informative auditory and tactile cues for in-vehicle information systems. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 102–109, 2010.

[60] M. Carrasco. Visual attention: The past 25 years. *Vision research*, 51(13):1484–1525, 2011.

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

[62] 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.

[63] A. Carton and L. E. Dunne. Tactile distance feedback for firefighters: design and preliminary evaluation of a sensory augmentation glove. In *Proceedings of the 4th augmented human international conference*, pages 58–64, 2013.

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

[65] A. Cassinelli, C. Reynolds, and M. Ishikawa. Augmenting spatial awareness with haptic radar. In *2006 10th IEEE International Symposium on Wearable Computers*, pages 61–64, Oct 2006.

[66] P. J. Catalano and S. M. Levin. Noise-induced hearing loss and portable radios with headphones. *International journal of pediatric otorhinolaryngology*, 9(1):59–67, 1985.

[67] 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*, volume ii, pages 659–669 vol.2, Jan 1992.

[68] P. Chakravarthy, D. Dunn, K. Akşit, and H. Fuchs. Focusar: Auto-focus augmented reality eyeglasses for both real and virtual. *IEEE transactions on visualization and computer graphics*, PP, 09 2018.

[69] L. Chan, R. Liang, M. Tsai, C. Cheng, K.and Su, M. Chen, W. Cheng, and B. Chen. Fingerpad: Private and subtle interaction using fingertips. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 255–260. ACM, 2013.

[70] E. Chancey, J. Brill, A. Sitz, U. Schmuntzsch, and J. Bliss. Vibrotactile stimuli parameters on detection reaction times. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1):1701–1705, 2014.

[71] W. Chang, W. Hwang, and Y. Ji. Haptic seat interfaces for driver information and warning systems. *International Journal of Human-Computer Interaction*, 27(12):1119–1132, 2011.

[72] C. Chen, Y. Chen, Y. Chung, and N. Yu. Motion guidance sleeve: Guiding the forearm rotation through external artificial muscles. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3272–3276. ACM, 2016.

[73] Y. Chen, Y. Yang, M. Li, E. Chen, W. Mu, R. Fisher, and R. Yin. Wearable actuators: An overview. *Textiles*, 1(2):283–321, 2021.

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

[75] M. Chmielewski, K. Sapiejewski, and M. Sobolewski. Application of augmented reality, mobile devices, and sensors for a combat entity quantitative assessment supporting decisions and situational awareness development. *Applied Sciences*, 9(21):4577, 2019.

[76] V. Chouvardas, A. Miliou, and M. Hatalis. Tactile displays: Overview and recent advances. *Displays*, 29(3):185–194, 2008.

[77] A. Cockburn and S. Brewster. Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48(9):1129–1150, 2005.

[78] A. Colley, J. Thebault-Spieker, A. Y. Lin, D. Degraen, B. Fischman, J. Häkkilä, K. Kuehl, V. Nisi, N. J. Nunes, N. Wenig, D. Wenig, B. Hecht, and J. Schöning. The geography of pokémon go: Beneficial and problematic effects on places and movement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1179–1192, New York, NY, USA, 2017. Association for Computing Machinery.

[79] C. C. Collins, F. A. Saunders, B. White, L. Scadden, et al. Vision substitution by tactile image projection. *Nature*, 221(5184):963–964, 1969.

[80] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PloS one*, 13(2):e0191846, 2018.

[81] A. B. H. Corinna Haupt. How axons see their way – axonal guidance in the visual system. *FBL*, 13(8):3136–3149, 2008.

[82] M. L. Courage, A. Bakhtiar, C. Fitzpatrick, S. Kenny, and K. Brandeau. Growing up multitasking: The costs and benefits for cognitive development. *Developmental Review*, 35:5–41, 2015.

[83] 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.

[84] 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.

[85] CyberGlove Systems Inc. Cybertouch ii. http://www.cyberglovesystems.com/cybertouch2/.

[86] D. Dakopoulos and N. G. Bourbakis. Wearable obstacle avoidance electronic travel aids for blind: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):25–35, 2009.

[87] D. Damos. *Multiple task performance*. CRC Press, 1991.

[88] 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.

[89] 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)*, pages 1–6. IEEE, 2016.

[90] A. Diederich and H. Colonius. Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, 66(8):1388–1404, 2004.

[91] M. K. Dobrzynski, S. Mejri, S. Wischmann, and D. Floreano. Quantifying information transfer through a head-attached vibrotactile display: principles for design and control. *IEEE Transactions on Biomedical Engineering*, 59(7):2011–2018, 2012.

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

[93] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In *Stereoscopic displays and virtual reality systems III*, volume 2653, pages 123–134. Spie, 1996.

[94] J. Driver and C. Spence. Cross–modal links in spatial attention. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1373):1319–1331, 1998.

[95] J. Driver and C. Spence. Crossmodal attention. *Current opinion in neurobiology*, 8(2):245–253, 1998.

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

[97] D. Drobny and J. O. Borchers. Learning basic dance choreographies with different augmented feedback modalities. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Extended Abstracts Volume, 2010*, pages 3793–3798, 2010.

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

[99] A. T. Duchowski, D. H. House, J. Gestring, R. I. Wang, K. Krejtz, I. Krejtz, R. Mantiuk, and B. Bazyluk. Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the acm symposium on applied perception*, pages 39–46, 2014.

[100] A. T. Duchowski, B. Pelfrey, D. H. House, and R. Wang. Measuring gaze depth with an eye tracker during stereoscopic display. In *Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization*, pages 15–22, 2011.

[101] A. Dufour. Importance of attentional mechanisms in audiovisual links. *Experimental Brain Research*, 126(2):215–222, 1999.

[102] 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.

[103] M. P. Eckstein. Visual search: A retrospective. *Journal of vision*, 11(5):14–14, 2011.

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

[105] M. Endsley and D. Garland. *Situation Awareness Analysis and Measurement*. CRC Press, 2000.

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

[107] M. R. Endsley. A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of sagat and spam. *Human factors*, 63(1):124–150, 2021.

[108] M. R. Endsley and E. S. Connors. Situation awareness: State of the art. In *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–4. IEEE, 2008.

[109] B. Ens, D. Ahlström, and P. Irani. Moving ahead with peephole pointing: Modelling object selection with head-worn display field of view limitations. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, pages 107–110. ACM, 2016.

[110] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Exploring the limitations of environment lighting on optical see-through head-mounted displays. In *Symposium on Spatial User Interaction*, New York, NY, USA, 2020. Association for Computing Machinery.

[111] L. Eriksson, A. Berglund, B. Willén, J. Svensson, M. Petterstedt, O. Carlander, B. Lindahl, and G. Allerbo. On visual, vibrotactile, and 3d audio directional cues for dismounted soldier waypoint navigation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(18):1282–1286, 2008.

[112] M. Ernst and M. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429, 2002.

[113] M. O. Ernst and H. H. Bülthoff. Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169, 2004.

[114] L. Fabiani, G. Burdea, N. Langrana, and D. Gomez. Human interface using the rutgers master ii force feedback interface. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pages 54–59, 1996.

[115] M. Failing and J. Theeuwes. Selection history: How reward modulates selectivity of visual attention. *Psychonomic bulletin & review*, 25(2):514–538, 2018.

[116] 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, pages 74–81, Washington, DC, USA, 1997. IEEE Computer Society.

[117] V. Ferrer, Y. Yang, A. Perdomo, and J. Quarles. Background motion, clutter, and the impact on virtual object motion perception in augmented reality. In *EGVE/EuroVR*, pages 57–64, 2013.

[118] V. Ferrer, Y. Yang, A. Perdomo, and J. Quarles. Consider your clutter: Perception of virtual object motion in ar. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–6. IEEE, 2013.

[119] N. J. Finlayson and P. M. Grove. Visual search is influenced by 3d spatial layout. *Attention, Perception, & Psychophysics*, 77(7):2322–2330, 2015.

[120] 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.

[121] 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*, volume 1, pages 3–9, 2005.

[122] 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. *SAE Transactions*, 105:1411–1418, 1996.

[123] T. Francart, A. Lenssen, and J. Wouters. Enhancement of interaural level differences improves sound localization in bimodal hearing. *The journal of the acoustical society of America*, 130(5):2817–2826, 2011.

[124] O. Franzén and J. Nordmark. Vibrotactile frequency discrimination. *Perception & Psychophysics*, 17(5):480–484, 1975.

[125] 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*, volume 2, pages 9–C. IEEE, 2003.

[126] J. P. Fritz and K. E. Barner. Stochastic models for haptic texture. In *Telemanipulator and Telepresence Technologies III*, volume 2901, pages 34–44. SPIE, 1996.

[127] 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, editors, *Augmented Environments for Computer-Assisted Interventions*, pages 24–35, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[128] J. L. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *2010 IEEE Virtual Reality Conference (VR)*, pages 79–86. IEEE, 2010.

[129] 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.

[130] J. L. Gabbard and J. Swan II. Usability engineering for augmented reality: Employing user-based studies to inform design. *IEEE Transactions on Visualization and Computer Graphics*, 14(3):513–525, 2008.

[131] P. Gallotti, A. Raposo, and L. Soares. v-glove: A 3d virtual touch interface. In *2011 XIII Symposium on Virtual Reality*, pages 242–251, 2011.

[132] S. Gao, S. Yan, H. Zhao, and A. Nathan. Haptic feedback. In *Touch-Based Human-Machine Interaction*, pages 91–108. Springer, 2021.

[133] H. Gil, H. Son, J. R. Kim, and I. Oakley. Whiskers: Exploring the use of ultrasonic haptic cues on the face. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[134] K. Gilliland and R. E. Schlegel. Tactile stimulation of the human head for information display. *Human Factors*, 36(4):700–717, 1994.

[135] E. Goldstein. *Sensation & Perception*. Cengage Learning, 2013.

[136] U. Gollner, T. Bieling, and G. Joost. Mobile lorm glove: Introducing a communication device for deaf-blind people. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 127–130. ACM, 2012.

[137] M. Gondan, K. Lange, F. Rösler, and B. Röder. The redundant target effect is affected by modality switch costs. *Psychonomic bulletin & review*, 11(2):307–313, 2004.

[138] 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, pages 177–186, Washington, DC, USA, 2012. IEEE Computer Society.

[139] P. Green and L. Wei-Haas. The rapid development of user interfaces: Experience with the wizard of oz method. *Proceedings of the Human Factors Society Annual Meeting*, 29(5):470–474, 1985.

[140] 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*, volume 59, pages 1727–1731. SAGE Publications Sage CA: Los Angeles, CA, 2015.

[141] 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.

[142] U. Gruenefeld, Y. Brück, and S. Boll. Behind the scenes: Comparing x-ray visualization techniques in head-mounted optical see-through augmented reality. In *19th International Conference on Mobile and Ubiquitous Multimedia*, pages 179–185, 2020.

[143] 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.

[144] 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*, SUI '17, pages 109–118. ACM, 2017.

[145] U. Gruenefeld, D. Hsiao, and W. Heuten. Eyeseex: Visualization of out-of-view objects on small field-of-view augmented and virtual reality devices. In *PerDis '18*, 2018.

[146] 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)*, pages 742–746. IEEE, 2019.

[147] U. Gruenefeld, D. Lange, L. Hammer, S. Boll, and W. Heuten. Flyingarrow: Pointing towards out-of-view objects on augmented reality devices. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, page 20. ACM, 2018.

[148] U. Gruenefeld, L. Prädel, and W. Heuten. Improving search time performance for locating out-of-view objects. *Mensch und Computer 2019-Tagungsband*, 2019.

[149] U. Gruenefeld, L. Prädel, and W. Heuten. Locating nearby physical objects in augmented reality. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–10, 2019.

[150] U. Gruenefeld, T. C. Stratmann, A. E. Ali, S. Boll, and W. Heuten. Radiallight: Exploring radial peripheral leds for directional cues in head-mounted displays. In

Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, pages 1–6, 2018.

[151] Y. Guo, L. Yang, B. Li, T. Liu, and Y. Liu. Rollcaller: User-friendly indoor navigation system using human-item spatial relation. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 2840–2848. IEEE, 2014.

[152] 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*, pages 787–796, 2008.

[153] 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*, pages 2399–2404, 2007.

[154] 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, pages 208–219. ACM, 2017.

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

[156] G. Haas, E. Stemasov, M. Rietzler, and E. Rukzio. Interactive auditory mediated reality: Towards user-defined personal soundscapes. In *Conference on Designing Interactive Systems*, pages 2035–2050, 2020.

[157] K. S. Hale and K. M. Stanney. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press, 2014.

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

[159] P. A. Hancock, J. E. Mercado, J. Merlo, and J. B. Van Erp. Improving target detection in visual search through the augmenting multi-sensory cues. *Ergonomics*, 56(5):729–738, 2013.

[160] S. Hanneton, M. Auvray, and B. Durette. The vibe: a versatile vision-to-audition sensory substitution device. *Applied Bionics and Biomechanics*, 7(4):269–276, 2010.

[161] J. Hartcher-O'Brien. *Multisensory integration of redundant and complementary cues*. PhD thesis, Oxford University, UK, 2012.

[162] J. Hartcher-O'Brien, M. Auvray, and V. Hayward. Perception of distance-to-obstacle through time-delayed tactile feedback. In *2015 IEEE World Haptics Conference (WHC)*, pages 7–12, 2015.

[163] C. Hatzfeld and T. Kern. *Engineering Haptic Devices: A Beginner's Guide*. Springer Series on Touch and Haptic Systems. Springer London, 2014.

[164] H. Heinze, G. R. Mangun, W. Burchert, H. Hinrichs, M. Scholz, T. Münte, A. Gös, M. Scherg, S. Johannes, H. Hundeshagen, et al. Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, 372(6506):543–546, 1994.

[165] 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.

[166] 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*, pages 675–678, 2010.

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

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

[169] C. Ho. *Multisensory aspects of the spatial cuing of driver attention*. PhD thesis, University of Oxford, 2006.

[170] C. Ho and C. Spence. *The Multisensory Driver: Implications for Ergonomic Car Interface Design*. Ashgate Publishing, Ltd., 2008.

[171] 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.

[172] E. Hoggan, A. Crossan, S. A. Brewster, and T. Kaaresoja. Audio or tactile feedback: Which modality when? In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2253–2256, 2009.

[173] B. Holbert. *Enhanced Targeting in a Haptic User Interface for the Physically Disabled Using a Force Feedback Mouse*. PhD thesis, University of Texas at Arlington, 2007.

[174] I. T. Hooge, R. S. Hessels, and M. Nyström. Do pupil-based binocular video eye trackers reliably measure vergence? *Vision Research*, 156:1–9, 2019.

[175] 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.

[176] J. P. Houston, H. Bee, and D. C. Rimm. *Invitation to psychology*. Academic Press, 2013.

[177] S. Hu, L. Su, S. Li, S. Wang, C. Pan, S. Gu, M. T. Al Amin, H. Liu, S. Nath, R. R. Choudhury, et al. Experiences with enav: A low-power vehicular navigation system. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 433–444, 2015.

[178] 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*, pages 235–236. IEEE, 2010.

[179] K. Iida. *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer Nature Singapore, 2019.

[180] W. Ijsselsteijn, H. De Ridder, and R. Hamberg. The effect of image disparity, convergence distance and focal length on perceived quality in stereoscopic displays. *IPO Annual progress Report*, 32:51–59, 1997.

[181] J. Intriligator and P. Cavanagh. The spatial resolution of visual attention. *Cognitive psychology*, 43(3):171–216, 2001.

[182] J. Irizarry, M. Gheisari, G. Williams, and B. Walker. Infospot: A mobile augmented reality method for accessing building information through a situation awareness approach. *Automation in construction*, 33:11–23, 2013.

[183] 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, pages 2019–2028. ACM, 2011.

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

[185] 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*, volume 44328, pages 263–272, 2011.

[186] H. Jin, Q. Chen, Z. Chen, Y. Hu, and J. Zhang. Multi-leapmotion sensor based demonstration for robotic refine tabletop object manipulation task. *CAAI Transactions on Intelligence Technology*, 1(1):104 – 113, 2016.

[187] S. Jin Cho, A. Ovcharenko, and U.-P. Chong. Front-back confusion resolution in 3d sound localization with hrtf databases. In *2006 International Forum on Strategic Technology*, pages 239 – 243, 11 2006.

[188] A. Jingu, M. Fujiwara, Y. Makino, and H. Shinoda. Tactile perception characteristics of lips stimulated by airborne ultrasound. In *2021 IEEE World Haptics Conference (WHC)*, pages 607–612. IEEE, 2021.

[189] 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.

[190] R. Johansson and J. Flanagan. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature reviews. Neuroscience*, 10(5):345, 2009.

[191] R. Johansson and A. Vallbo. Tactile sensibility in the human hand: relative and absolute densities of four types of mechanoreceptive units in glabrous skin. *The Journal of physiology*, 286(1):283–300, 1979.

[192] M. S. John and H. S. Smallman. Staying up to speed: Four design principles for maintaining and recovering situation awareness. *Journal of Cognitive Engineering and Decision Making*, 2(2):118–139, 2008.

[193] K. Johnson and S. Hsiao. Neural mechanisms of tactual form and texture perception. *Annual review of neuroscience*, 15(1):227–250, 1992.

[194] J. Jung, H. Lee, J. Choi, A. Nanda, U. Gruenefeld, T. Stratmann, and W. Heuten. Ensuring safety in augmented reality from trade-off between immersion and situation awareness. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 10 2018.

[195] K. A. Kaczmarek, J. Webster, P. Bach-y Rita, and W. Tompkins. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering*, 38(1):1–16, 1991.

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

[197] D. Karnopp. Computer simulation of stick-slip friction in mechanical dynamic systems. *J. Dyn. Syst. Meas. Control.*, 107(1):100–103, 1985.

[198] S. J. Kass, K. S. Cole, and C. J. Stanny. Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(4):321–329, 2007.

[199] T. Kassuba, M. Menz, B. Röder, and H. Siebner. Multisensory interactions between auditory and haptic object recognition. *Cerebral Cortex*, 23(5):1097–1107, 2013.

[200] 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.

[201] 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)*, pages 1–7, 01 2008.

[202] 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, pages 2533–2539. ACM, 2016.

[203] D. V. Keyson and A. J. Houtsma. Directional sensitivity to a tactile point stimulus moving across the fingerpad. *Perception & Psychophysics*, 57(5):738–744, 1995.

[204] J. Kim, S.-w. Oh, J. Choi, S. Park, and W. Kim. Optical see-through head-mounted display including transmittance-variable display for high visibility. *Journal of Information Display*, 23(2):121–127, 2022.

[205] 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.

[206] 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*, pages 133–142. ACM, 2009.

[207] 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. In *Proceedings of the Advances in Multimedia Information Processing, and 11th Pacific Rim Conference on Multimedia: Part II*, pages 682–693, 09 2010.

[208] 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)*, pages 177–186, Sept 2014.

[209] K. Kiyokawa. A wide field-of-view head mounted projective display using hyperbolic half-silvered mirrors. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 207–210, 12 2007.

[210] K. Kiyokawa. Trends and vision of head mounted display in augmented reality. In *2012 International Symposium on Ubiquitous Virtual Reality*, pages 14–17, 2012.

[211] M. Klapdohr, B. Wöldecke, D. Marinos, J. Herder, C. Geiger, and W. Vonolfen. Vibrotactile pitfalls: Arm guidance for moderators in virtual tv studios. In *Proceedings of the 13th International Conference on Humans and Computers*, HC '10, pages 72–80. University of Aizu Press, 2010.

[212] A. Klapetek, M. Ngo, and C. Spence. Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, 74(6):1154–1167, 2012.

[213] R. Klatzky and N. Giudice. *Sensory substitution of vision: Importance of perceptual and cognitive processing*, pages 162–191. CRC Press, 01 2012.

[214] R. M. Klein. On the control of visual orienting., 2004.

[215] R. M. Klein and D. I. Shore. Relations among modes of visual orienting. *Attention and performance XVIII: Control of cognitive processes*, pages 195–208, 2000.

[216] 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)*, pages 3192–3195. IEEE, 2020.

[217] 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.

[218] 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.

[219] K. Kozak, J. Pohl, W. Birk, J. Greenberg, B. Artz, M. Blommer, L. Cathey, and R. Curry. Evaluation of lane departure warnings for drowsy drivers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(22):2400–2404, 2006.

[220] G. Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(7):1912–1931, 2015.

[221] 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)*, pages 3–12. IEEE Computer Society, 2010.

[222] E. Kruijff, A. Marquardt, C. Trepkowski, R. Lindeman, A. Hinkenjann, J. Maiero, and B. Riecke. On your feet!: Enhancing vection in leaning-based interfaces through multisensory stimuli. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, pages 149–158. ACM, 2016.

[223] E. Kruijff, A. Marquardt, C. Trepkowski, J. Schild, and A. Hinkenjann. Designed emotions: Challenges and potential methodologies for improving multisensory cues to enhance user engagement in immersive systems. *Vis. Comput.*, 33(4):471–488, Apr. 2017.

[224] 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*, pages 1–1, 2018.

[225] E. Kruijff, K. Wesche, G.and Riege, G. Goebbels, M. Kunstman, and D. Schmalstieg. Tactylus, a pen-input device exploring audiotactile sensory binding. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '06, pages 312–315. ACM, 2006.

[226] R. Kumar, G. D. Hager, A. Barnes, P. Jensen, and R. H. Taylor. An augmentation system for fine manipulation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 956–965. Springer, 2000.

[227] 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.

[228] M. Kytö, B. Ens, T. Piumsomboon, G. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[229] M. Kytö, A. Mäkinen, T. Tossavainen, and P. Oittinen. Stereoscopic depth perception in video see-through augmented reality within action space. *Journal of Electronic Imaging*, 23(1):1 – 11, 2014.

[230] D. A. Laird and K. Coye. Psychological measurements of annoyance as related to pitch and loudness. *The Journal of the Acoustical Society of America*, 1(1):158–163, 1929.

[231] 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, editors, *Augmentation, Application, Innovation, Collaboration, OzCHI '13, Adelaide, Australia - November 25 - 29, 2013*, pages 545–554. ACM, 2013.

[232] N. Lavie and Y. Tsal. Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & psychophysics*, 56(2):183–197, 1994.

[233] J. LaViola, E. Kruijff, R. McMahan, D. Bowman, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Usability. Pearson Education, 2017.

[234] 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.

[235] 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, pages 445–452, New York, NY, USA, 2012. ACM.

[236] 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.

[237] T. Letowski and S. Letowski. *Localization error: Accuracy and precision of auditory localization*, volume 55, page 78. InTech, 2011.

[238] D. E. Levac and H. Sveistrup. Motor learning and virtual reality. In *Virtual reality for physical and motor rehabilitation*, pages 25–46. Springer, 2014.

[239] 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*, pages 224–230, Nov 2004.

[240] R. Lichenstein, D. C. Smith, J. L. Ambrose, and L. A. Moody. Headphone use and pedestrian injury and death in the united states: 2004–2011. *Injury prevention*, 18(5):287–290, 2012.

[241] J. Lieberman and C. Breazeal. Tikl: Development of a wearable vibrotactile feedback suit for improved human motor learning. *IEEE Transactions on Robotics*, 23(5):919–926, 2007.

[242] 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*, pages 146–149, 2004.

[243] R. Lindeman, Y. Yanagida, J. Sibert, and R. Lavine. Effective vibrotactile cueing in a visual search task. In *Proc. of Interact 2003*, pages 89–96, 2003.

[244] 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, pages 175–178. ACM, 2007.

[245] R. W. Lindeman, H. Noma, and P. G. De Barros. Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 173–176. IEEE, 2007.

[246] P. Lindemann, T. Lee, and G. Rigoll. Supporting driver situation awareness for autonomous urban driving with an augmented-reality windshield display. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 358–363, Oct 2018.

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

[248] L. Liu and R. van Liere. Designing 3d selection techniques using ballistic and corrective movements. In *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments*, JVRC'09, pages 1–8. Eurographics Association, 2009.

[249] M. Livingston, L. Rosenblum, D. Brown, G. Schmidt, S. Julier, Y. Baillot, E. Swan, Z. Ai, and P. Maassel. Military applications of augmented reality. In *Handbook of augmented reality*, pages 671–706. Springer, 2011.

[250] T. Lloyd-Esenkaya, V. Lloyd-Esenkaya, E. O'Neill, and M. J. Proulx. Multisensory inclusive design with sensory substitution. *Cognitive Research: Principles and Implications*, 5(1):1–15, 2020.

[251] L. H. Loiselle, M. F. Dorman, W. A. Yost, S. J. Cook, and R. H. Gifford. Using ild or itd cues for sound source localization and speech understanding in a complex listening environment by listeners with bilateral and with hearing-preservation cochlear implants. *Journal of Speech, Language, and Hearing Research*, 59(4):810–818, 2016.

[252] J. Loomis, R. Klatzky, and N. Giudice. Sensory substitution of vision: Importance of perceptual and cognitive processing. In *Assistive technology for blindness and low vision*, pages 179–210. CRC Press, 2018.

[253] P. Lopes, D. Yüksel, F. Guimbretière, and P. Baudisch. Muscle-plotter: An interactive system based on electrical muscle stimulation that produces spatial output. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 207–217, 2016.

[254] E. A. Lopez-Poveda. Chapter 10 - development of fundamental aspects of human auditory perception. In *Development of Auditory and Vestibular Systems*, pages 287–314. Academic Press, 2014.

[255] 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, pages 1093–1102. ACM, 2014.

[256] W. Lu, B.-L. H. Duh, and S. Feiner. Subtle cueing for visual search in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 161–166, 2012.

[257] W. Lu, H. B.-L. Duh, S. Feiner, and Q. Zhao. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):404–412, 2014.

[258] A. MacAllister, M. Hoover, S. Gilbert, J. Oliver, R. Radkowski, T. Garrett, J. Holub, E. Winer, S. Terry, and P. Davies. Comparing visual assembly aids for augmented reality work instructions. *Interserv Industr Train*, pages 1–14, 2017.

[259] E. Macaluso, U. Noppeney, D. Talsma, T. Vercillo, J. Hartcher-O'Brien, and R. Adam. The curious incident of attention in multisensory integration: bottom-up vs. top-down. *Multisensory Research*, 29(6-7):557–583, 2016.

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

[261] V. Maheshwari and R. Saraf. Tactile devices to sense touch on a par with a human finger. *Angewandte Chemie International Edition*, 47(41):7808–7826, 2008.

[262] 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.

[263] N. K. Malhotra. Information and sensory overload. information and sensory overload in psychology and marketing. *Psychology & Marketing*, 1(3-4):9–21, 1984.

[264] M. Mancas and V. P. Ferrera. How to measure attention? In *From human attention to computational attention*, pages 21–38. Springer, 2016.

[265] 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*, pages 1–10, 2018.

[266] A. Marquardt, J. Maiero, E. Kruijff, C. Trepkowski, A. Schwandt, A. Hinkenjann, J. Schoening, and W. Stuerzlinger. Tactile hand motion and pose guidance for 3d interaction. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '18. ACM, 2018.

[267] A. Marquardt, C. Trepkowski, T. 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)*, pages 190–201, Oct 2019.

[268] 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.

[269] J. Martinez, A. Garcia, M. Oliver, J. P. Molina, and P. Gonzalez. Identifying virtual 3d geometric shapes with a vibrotactile glove. *IEEE Computer Graphics and Applications*, 36(1):42–51, 2016.

[270] M. Marucci, G. Di Flumeri, G. Borghini, N. Sciaraffa, M. Scandola, E. F. Pavone, F. Babiloni, V. Betti, and P. Aricò. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports*, 11(1):1–15, 2021.

[271] T. H. Massie and J. K. Salisbury. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME Dynamic Systems and Control Division*, pages 295–301, 1994.

[272] V. Mateevitsi, B. Haggadone, J. Leigh, B. Kunzer, and R. Kenyon. Sensing the environment through spidersense. In *Proceedings of the 4th augmented human international conference*, pages 51–57. ACM, 2013.

[273] G. Matthews and I. Margetts. Self-report arousal and divided attention: A study of performance operating characteristics. *Human Performance*, 4(2):107–125, 1991.

[274] M. L. Matthews, D. J. Bryant, R. D. Webb, and J. L. Harbluk. Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems. *Transportation research record*, 1779(1):26–32, 2001.

[275] T. McDaniel, D. Villanueva, S. Krishna, and S. Panchanathan. Movement: A framework for systematically mapping vibrotactile stimulations to fundamental body movements. In *Haptic Audio-Visual Environments and Games (HAVE), 2010 IEEE International Symposium on*, pages 1–6. IEEE, 2010.

[276] 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.

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

[278] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, 1976.

[279] 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.

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

[281] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18(4):626–633, 2012.

[282] E. McSorley and J. Findlay. Visual search in depth. *Vision Research*, 41(25):3487 – 3496, 2001.

[283] P. B. Meijer. An experimental system for auditory image representations. *IEEE transactions on biomedical engineering*, 39(2):112–121, 1992.

[284] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. P. Ferreira, and J. A. Santos. On the improvement of localization accuracy with non-individualized hrtf-based sounds. *Journal of the Audio Engineering Society*, 60(10):821–830, 2012.

[285] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha. P-hrtf: Efficient personalized hrtf computation for high-fidelity spatial sound. In *2014 IEEE*

International Symposium on Mixed and Augmented Reality (ISMAR), pages 53–61. IEEE, 2014.

[286] Microsoft Corporation. Hololens 2 - overview, features, and specs | microsoft hololens. https://www.microsoft.com/en-us/hololens/hardware.

[287] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–292. International Society for Optics and Photonics, 1995.

[288] M. Mine, F. Brooks, Jr., and C. H. Sequin. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 19–26. ACM Press/Addison-Wesley Publishing Co., 1997.

[289] 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.

[290] B. R. Molesworth, M. Burgess, and D. Kwon. The use of noise cancelling headphones to improve concurrent task performance in a noisy environment. *Applied Acoustics*, 74(1):110–115, 2013.

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

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

[293] 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.

[294] R. Musil. Hmd geometry database. https://risa2000.github.io/hmdgdb/.

[295] 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.

[296] K. Myles and J. T. Kalb. Guidelines for head tactile communication. Technical report, Army Research Lab Aberdeen Proving Ground Md Human Research And Engineering ..., 2010.

[297] 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, 2006.

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

[299] A. W. Ng and A. H. Chan. Finger response times to visual, auditory and tactile modality stimuli. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 2, pages 1449–1454, 2012.

[300] V. Ng-Thow-Hing, K. Bark, L. Beckwith, C. Tran, R. Bhandari, and S. Sridhar. User-centered perspectives for automotive augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 13–22, 2013.

[301] 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.

[302] 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.

[303] K. Nosaka, A. Aldayel, M. Jubeau, and T. C. Chen. Muscle damage induced by electrical stimulation. *European Journal of Applied Physiology*, 111(10):2427, Aug 2011.

[304] L. Nyre, B. Tessem, B. Wendelbo, and S. H. Øvreås. The perceptual mechanics of noise-cancelling headphones. https://teklab.uib.no/artikler/the-perceptual-mechanics-of-noise-cancelling-headphones/.

[305] N. O. A. Observatory. Recommended light levels (illuminance) for outdoor and indoor venues, 2015. https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf.

[306] V. Occelli, C. Spence, and M. Zampini. Audiotactile interactions in temporal perception. *Psychonomic Bulletin & Review*, 18(3):429–454, Jun 2011.

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

[308] 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.

[309] S. Oney, N. Rodrigues, M. Becher, T. Ertl, G. Reina, M. Sedlmair, and D. Weiskopf. Evaluation of gaze depth estimation from eye tracking in augmented reality. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–5, 2020.

[310] T. Oron-Gilad, J. L. Downs, R. D. Gilson, and P. A. Hancock. Vibrotactile guidance cues for target acquisition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(5):993–1004, 2007.

[311] 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*, pages 129–136, 2004.

[312] Oxford English Dictionary. noise, n. https://www.oed.com/viewdictionaryentry/Entry/127655.

[313] D. Pai. Multisensory interaction: Real and virtual. In *Robotics Research. The Eleventh International Symposium*, pages 489–498. Springer, 2005.

[314] B. Park, C. Yoon, J. Lee, and K. Kim. Augmented reality based on driving situation awareness in vehicle. In *2015 17th International Conference on Advanced Communication Technology (ICACT)*, pages 593–595, July 2015.

[315] 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.

[316] R. Pastel. Measuring the difficulty of steering through corners. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1087–1096. ACM, 2006.

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

[318] D. Pecher, R. Zeelenberg, and L. W. Barsalou. Verifying different-modality properties for concepts produces switching costs. *Psychological science*, 14(2):119–124, 2003.

[319] R. L. Peiris, W. Peng, Z. Chen, L. Chan, and K. Minamizawa. Thermovr: Exploring integrated thermal haptic feedback with head mounted displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, page 5452–5456, New York, NY, USA, 2017. Association for Computing Machinery.

[320] D. R. Perrott, J. Cisneros, R. L. Mckinley, and W. R. D'Angelo. Aurally aided visual search under virtual and free-field listening conditions. *Human factors*, 38(4):702–715, 1996.

[321] D. R. Perrott, K. Saberi, K. Brown, and T. Z. Strybel. Auditory psychomotor coordination and visual search performance. *Perception & psychophysics*, 48(3):214–226, 1990.

[322] D. R. Perrott, T. Sadralodabai, K. Saberi, and T. Z. Strybel. Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors*, 33(4):389–400, 1991.

[323] D. R. Perrott and J. Tucker. Minimum audible movement angle as a function of signal frequency and the velocity of the source. *The Journal of the Acoustical Society of America*, 83(4):1522–1527, 1988.

[324] 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*, pages 169–176. IEEE Computer Society, 2008.

[325] 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, pages 143–152. IEEE Computer Society, 2008.

[326] B. Petzold, M. Zaeh, B. Faerber, B. Deml, H. Egermeier, J. Schilp, and S. Clarke. A study on visual, auditory, and haptic feedback for assembly tasks. *Presence: Teleoper. Virtual Environ.*, 13(1):16–21, Feb. 2004.

[327] 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, pages 1269–1274. ACM, 2013.

[328] D. Pham and W. Stuerzlinger. Is the pen mightier than the controller? a comparison of input devices for selection in virtual and augmented reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19. Association for Computing Machinery, 2019.

[329] T. G. Philippi, J. B. van Erp, and P. J. Werkhoven. Multisensory temporal numerosity judgment. *Brain research*, 1242:116–125, 2008.

[330] E. Piateski and L. Jones. Vibrotactile pattern recognition on the arm and torso. In *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint*, pages 90–95. IEEE, 2005.

[331] C. J. Plack. *The sense of hearing*. Routledge, 2018.

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

[333] M. I. Posner, M. J. Nissen, and R. M. Klein. Visual dominance: an information-processing account of its origins and significance. *Psychological review*, 83(2):157, 1976.

[334] 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.

[335] D. Prabhu, M. M. Hasan, L. Wise, C. MacMahon, and C. McCarthy. Vibrosleeve: A wearable vibro-tactile feedback device for arm guidance. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4909–4912. IEEE, 2020.

[336] M. Prachyabrued and C. W. Borst. Visual feedback for virtual grasping. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 19–26, 2014.

[337] 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.

[338] M. J. Proulx, D. J. Brown, A. Pasqualotto, and P. Meijer. Multisensory perceptual learning and sensory substitution. *Neuroscience & Biobehavioral Reviews*, 41:16–25, 2014.

[339] D. Pyo, T.-H. Yang, S. Ryu, and D.-S. Kwon. Novel linear impact-resonant actuator for mobile applications. *Sensors and Actuators A: Physical*, 233:460–471, 2015.

[340] 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.

[341] P. Quinlan. Visual feature integration theory: Past, present, and future. *Psychological bulletin*, 129:643–73, 10 2003.

[342] 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*, pages 287–288, 2009.

[343] A. K. Raj, J. D. Beach, M. E. Stuart, and L. A. Vassiliades. *Multimodal and Multisensory Displays for Perceptual Tasks*, page 299–324. Cambridge Handbooks in Psychology. Cambridge University Press, 2015.

[344] H. Regenbrecht, J. Hauber, R. Schoenfelder, and A. Maegerlein. Virtual reality aided assembly with directional vibro-tactile feedback. In *Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, GRAPHITE '05, pages 381–387. ACM, 2005.

[345] 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*, pages 93–102. IEEE, 2016.

[346] H. Renkewitz and T. Alexander. Perceptual issues of augmented and virtual environments. Technical report, FGAN-FKIE WACHTBERG (GERMANY), 2007.

[347] 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)*, pages 186–194, March 2017.

[348] 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)*, pages 319–324, Sep. 2012.

[349] P. Richard, G. Burdea, D. Gomez, and P. Coiffet. A comparison of haptic, visual and auditive force feedback for deformable virtual objects. In *Proceedings of the Internation Conference on Automation Technology (ICAT)*, volume 49, page 62, 1994.

[350] M. Risoud, J.-N. Hanson, F. Gauvrit, C. Renard, P.-E. Lemesre, N.-X. Bonne, and C. Vincent. Sound source localization. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 135(4):259–264, 2018.

[351] S. Roffler and R. Butler. Localization of tonal stimuli in the vertical plane. *The Journal of the Acoustical Society of America*, 43(6):1260–1266, 1968.

[352] 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.

[353] R. Rosenholtz. Search asymmetries? what search asymmetries? *Perception & Psychophysics*, 63(3):476–489, 2001.

[354] R. Rosenholtz, Y. Li, and L. Nakano. Measuring visual clutter. *Journal of vision*, 7(2):17–17, 2007.

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

[356] S. Rothe, D. Buschek, and H. Hußmann. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1):19, 2019.

[357] C. S. Royden, J. M. Wolfe, and N. Klempen. Visual search asymmetries in motion and optic flow fields. *Perception & psychophysics*, 63(3):436–444, 2001.

[358] E. Ruffaldi, A. Filippeschi, A. Frisoli, O. Sandoval, C. A. Avizzano, and M. Bergamasco. Vibrotactile perception assessment for a rowing training system. In *World Haptics 2009 - Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 350–355, March 2009.

[359] 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.

[360] 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*, New York, NY, USA, 2013. Association for Computing Machinery.

[361] P. Salmon, N. Stanton, G. Walker, and D. Green. Situation awareness measurement: A review of applicability for c4i environments. *Applied ergonomics*, 37(2):225–238, 2006.

[362] 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)*, pages 47–50. IEEE, 2010.

[363] 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.

[364] K. Sato, K. Minamizawa, N. Kawakami, and S. Tachi. Haptic telexistence. In *ACM SIGGRAPH 2007 Emerging Technologies*, SIGGRAPH '07, San Diego, California, 2007. ACM.

[365] V. E. Scerra and J. C. Brill. Effect of task modality on dual-task performance, response time, and ratings of operator workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1):1456–1460, 2012.

[366] T. Schinke, N. Henze, and S. Boll. Visualization of off-screen objects in mobile augmented reality. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '10, page 313–316, New York, NY, USA, 2010. Association for Computing Machinery.

[367] J. J. Schlesinger, E. Reynolds, B. Sweyer, and A. Pradham. Frequency-selective silencing device for digital filtering of audible medical alarm sounds to enhance icu patient recovery. *23th Meeting of the Intl. Conf. on Auditory Display (ICAD)*, 2017.

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

[369] R. Schmidt and C. Wrisberg. *Motor Learning and Performance*. Human Kinetics, 2004.

[370] C. Schönauer, K. Fukushi, A. Olwal, H. Kaufmann, and R. Raskar. Multimodal motion guidance: Techniques for adaptive and dynamic feedback. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, pages 133–140. ACM, 2012.

[371] 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, pages 91–94. IEEE Computer Society, 2008.

[372] C. Seim, N. Doering, Y. Zhang, W. Stuerzlinger, and T. Starner. Passive haptic training to improve speed and performance on a keypad. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):100:1–100:13, Sept. 2017.

[373] S.Hazenberg and R. van Lier. Touching and hearing unseen objects: Multisensory effects on scene recognition. *i-Perception*, 7(4), 2016.

[374] 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, editors, *Haptic and Audio Interaction Design*, pages 51–60, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[375] 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*, pages 2651–2656, 2007.

[376] T. Shibata. Head mounted display. *Displays*, 23(1):57–64, 2002.

[377] D. Shibli. Using cognitive load theory to improve the use of slideshow presentations and support a more efficient learning process. *Blended Learning in Practice (2019)*, 50, 2019.

[378] S. Shimojo and L. Shams. Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 11(4):505–509, 2001.

[379] 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.

[380] D. Spelmezan, M. Jacobs, A. Hilgers, and J. Borchers. Tactile motion instructions for physical activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2243–2252. ACM, 2009.

[381] C. Spence, M. E. Nicholls, and J. Driver. The cost of expecting events in the wrong sensory modality. *Perception & psychophysics*, 63(2):330–336, 2001.

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

[383] C. Spence, D. Senkowski, and B. Röder. Crossmodal processing, 2009.

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

[385] J. Sreng, A. Lecuyer, C. Megard, and C. Andriot. Using visual cues of contact to improve interactive manipulation of virtual objects in industrial assembly/maintenance simulations. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1013–1020, 2006.

[386] A. A. Stanley and K. J. Kuchenbecker. Evaluation of tactile feedback methods for wrist rotation guidance. *EEE Trans. Haptics*, 5(3):240–251, Jan. 2012.

[387] L. Stark, K. Ezumi, T. Nguyen, R. Paul, G. Tharp, and H. Yamashita. Visual search in virtual environments. In *Human vision, visual processing, and digital display III*, volume 1666, pages 577–589. International Society for Optics and Photonics, 1992.

[388] A. Steed, F. R. Ortega, A. S. Williams, E. Kruijff, W. Stuerzlinger, A. U. Batmaz, A. S. Won, E. S. Rosenberg, A. L. Simeone, and A. Hayes. Evaluating immersive experiences during covid-19 and beyond. *Interactions*, 27(4):62–67, July 2020.

[389] B. E. Stein. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, 123(1):124–135, 1998.

[390] 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.

[391] B. E. Stein and M. A. Meredith. *The merging of the senses.* The MIT press, 1993.

[392] B. E. Stein and M. T. Wallace. Comparisons of cross-modality integration in midbrain and cortex. *Progress in brain research*, 112:289–299, 1996.

[393] M. Straub, A. Riener, and A. Ferscha. Distance encoding in vibro-tactile guidance cues. In *2009 6th Annual International Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous*, pages 1–2. IEEE, 2009.

[394] E. Swan, A. Jones, E. Kolstad, M. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE transactions on visualization and computer graphics*, 13(3):429–442, 2007.

[395] 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.

[396] E. Tamaki, T. Miyaki, and J. Rekimoto. Possessedhand: Techniques for controlling human hands using electrical muscles stimuli. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 543–552, 2011.

[397] 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, pages 73–80. ACM, 2003.

[398] X. Tang, J. Wu, and Y. Shen. The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, 61:208–224, 2016.

[399] 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)*, pages 83–92. IEEE, 2016.

[400] The Interaction Design Foundation. Beyond ar vs. vr: What is the difference between ar vs. mr vs. vr vs. xr? https://www.interaction-design.org/literature/article/beyond-ar-vs-vr-what-is-the-difference-between-ar-vs-mr-vs-vr-vs-xr.

[401] The Interaction Design Foundation. What is virtuality continuum? https://www.interaction-design.org/literature/topics/virtuality-continuum.

[402] 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), pages 56–59. IEEE, 2005.

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

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

[405] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[406] C. Trepkowski, D. 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)*, pages 575–584. IEEE, 2019.

[407] C. Trepkowski, A. Marquardt, T. D. Eibich, Y. Shikanai, J. Maiero, K. Kiyokawa, E. Kruijff, J. Schöning, and P. König. Multisensory proximity and transition cues for improving target awareness in narrow field of view augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 28(2):1342–1362, 2021.

[408] J. Truxal. *The Age of Electronic Messages*. Lionel Robbins Lectures for 1989. MIT Press, 1990.

[409] J. Turner. Evaluation of an eye-slaved area-of-interest display for tactical combat simulation. In *The 6th interservice/industry training equipment conference and exhibition*, pages 75–86, 1984.

[410] 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*, pages 336–339. ACM, 2008.

[411] H. Uematsu, D. Ogawa, R. Okazaki, T. Hachisu, and H. Kajimoto. Halux: projection-based interactive skin for digital sports. In *SIGGRAPH Emerging Technologies*, 2016.

[412] N. van Atteveldt, M. Murray, G. Thut, and C. Schroeder. Multisensory integration: Flexible use of general operations. *Neuron*, 81(6):1240 – 1253, 2014.

[413] 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.

[414] E. Van der Burg, C. Olivers, A. Bronkhorst, and J. Theeuwes. Pip and pop: non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5):1053, 2008.

[415] E. Van der Burg, C. N. Olivers, A. W. Bronkhorst, and J. Theeuwes. Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience letters*, 450(1):60–64, 2009.

[416] R. Van der Linde, P. Lammertse, E. Frederiksen, and B. Ruiter. The hapticmaster, a new high-performance haptic interface. In *Proc. Eurohaptics*, pages 1–5, 2002.

[417] G. H. Van Doorn, V. Dubaj, D. B. Wuillemin, B. L. Richardson, and M. A. Symmons. Cognitive load can explain differences in active and passive touch. In P. Isokoski and J. Springare, editors, *Haptics: Perception, Devices, Mobility, and Communication*, pages 91–102. Springer Berlin Heidelberg, 2012.

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

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

[420] D. Van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1–20, 2010.

[421] E. Veas, E. Mendez, S. 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, pages 1471–1480, Vancouver, BC, Canada, 2011. ACM.

[422] R. T. Verrillo. Sensory substitution. In *Handbook of Signal Processing in Acoustics*, pages 1231–1244. Springer, 2008.

[423] 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.

[424] S. Vishniakou, B. Lewis, X. Niu, A. Kargar, K. Sun, M. Kalajian, N. Park, M. Yang, Y. Jing, P. Brochu, et al. Tactile feedback display with spatial and temporal resolutions. *Scientific reports*, 3:2521, 2013.

[425] H. Vitense, J. Jacko, and V. Emery. Multimodal feedback: Establishing a performance baseline for improved access by individuals with visual impairments. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies*, Assets '02, pages 49–56. ACM, 2002.

[426] R. Volcic, C. Fantoni, C. Caudek, J. A. Assad, and F. Domini. Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *Journal of Neuroscience*, 33(43):17081–17088, 2013.

[427] T. M. Voong and M. Oehler. Auditory spatial perception using bone conduction headphones along with fitted head related transfer functions. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1211–1212. IEEE, 2019.

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

[429] 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.

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

[431] B. N. Walker, R. M. Stanley, N. Iyer, B. D. Simpson, and D. S. Brungart. Evaluation of bone-conduction headsets for use in multitalker communication environments. *Proceedings of the human factors and ergonomics society annual meeting*, 49(17):1615–1619, 2005.

[432] 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*, pages 1123–1132, 2006.

[433] M. Ward, A. Barde, P. Russell, M. Billinghurst, and W. Helton. Visual cues to reorient attention from head mounted displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60:1574–1578, 09 2016.

[434] 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.

[435] 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*, volume 1116, pages 126–138. International Society for Optics and Photonics, 1989.

[436] 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.

[437] E. Wetzel, J. Boehnke, and A. Brown. *Response biases*, pages 349–363. Oxford University Press, 2016.

[438] B. W. White, F. A. Saunders, L. Scadden, C. C. Collins, et al. Seeing with the skin. *Perception & Psychophysics*, 7(1):23–27, 1970.

[439] C. D. Wickens, J. D. Lee, Y. Liu, and S. E. G. Becker. Visual search and detection. *Introduction to Human Factors Engineering, 2nd Edition*, pages 78–90, 2004.

[440] 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*, pages 98–101, Oct 1993.

[441] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. ii: Psychophysical validation. *The Journal of the Acoustical Society of America*, 85(2):868–878, 1989.

[442] 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.

[443] A. Wilberz, D. Leschtschow, C. Trepkowski, J. Maiero, E. Kruijff, and B. Riecke. Facehaptics: Robot arm based versatile facial haptics for immersive environments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14. Association for Computing Machinery, 2020.

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

[445] U. Windhorst. Sensory systems. *Encyclopedia of Neuroscience*, pages 3663–3673, 2009.

[446] 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*, pages 143–146, 2011.

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

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

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

[450] J. M. Wolfe. Visual search. *Current Biology*, 20(8):R346–R349, 2010.

[451] J. M. Wolfe. Visual search: How do we find what we are looking for. *Annual review of vision science*, 6(1):539–562, 2020.

[452] J. M. Wolfe and T. S. Horowitz. Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3):1–8, 2017.

[453] J. Woodward and J. Ruiz. Analytic review of using augmented reality for situational awareness. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[454] B. Wu, T. Ooi, and Z. He. Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428:73–7, 04 2004.

[455] J. Xu and S. Yue. Mimicking visual searching with integrated top down cues and low-level features. *Neurocomputing*, 133:1–17, 2014.

[456] S. Yamanaka, W. Stuerzlinger, and H. Miyashita. Steering through successive objects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 603:1–603:13. ACM, 2018.

[457] J. M. Yau, G. C. DeAngelis, and D. E. Angelaki. Dissecting neural circuits for multisensory integration and crossmodal processing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1677):20140203, 2015.

[458] V. Yem, R. Okazaki, and H. Kajimoto. Vibrotactile and pseudo force presentation using motor rotational acceleration. In *2016 IEEE Haptics Symposium (HAPTICS)*, pages 47–51. IEEE, 2016.

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

[460] M. Yuan, S. Ong, and A. Nee. Augmented reality for assembly guidance using a virtual interactive tool. *International journal of production research*, 46(7):1745–1767, 2008.

[461] 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*, pages 838–839, 2003.

[462] L. Zhang and W. Lin. *Selective visual attention: computational models and applications*. John Wiley & Sons, 2013.

[463] J. Zheng, Y.and Morrell. A vibrotactile feedback approach to posture guidance. In *Haptics Symposium, 2010 IEEE*, pages 351–358. IEEE, 2010.

[464] 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.

[465] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE, 2008.

[466] M. Zhou, D. Jones, S. Schwaitzberg, and C. Cao. Role of haptic feedback and cognitive load in surgical skill acquisition. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51(11):631–635, 2007.

A Supplementary Material for Chapter 6

User Preference Scores in Study 1

Table A.1 shows the number of times a feedback mode was preferred over another mode in for all individual comparisons and in total. The column "Pairing" lists which two modes competed against each other. That is, these two modes were presented to the user, who then chose the preferred mode (forced-choice procedure) that was considered more appropriate to increase awareness of an incoming out-of-view object. Each pairing was presented twice to 10 users. The columns "Mode 1" and "Mode 2" refer to the respective mode in the column "Pairing" and count the number of times in which the respective mode was preferred over the other. The maximum total score a mode could achieve was 100 for both reduced and increased noise.

Table A.1: User preference ratings in study 1.

	Reduced Noise		Increased Noise	
Pairing Mode 1 - Mode 2	Mode 1	Mode 2	Mode 1	Mode 2
AA-AT	6	18	2	22
AA-AV	21	3	17	7
AA-VA	10	14	15	9
AA-VT	7	17	5	19
AA-VV	21	3	23	1
AT-AV	20	4	21	3
AT-VA	18	6	19	5
AT-VT	10	14	20	4
AT-VV	24	0	22	2
AV-VA	7	17	7	17
AV-VT	5	19	4	20
AV-VV	20	4	22	2
VA-VT	5	19	5	19
VA-VV	23	1	21	3
VT-VV	23	1	23	1

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

Performance Measures for Study 2

Table A.2 shows the average reaction time (RT) in the awareness task and the hit rate (HR) of the focal attention task with standard deviations (SD) and interquartile ranges (IQR) for VA, VT, and AT modes by noise condition in Study 2. ***$p < .001$.

Table A.2: Performance measures in study 2.

	Reduced noise			Increased noise		
Mean (SD)	VA	VT	AT	VA	VT	AT
RT	0.73 (0.76)	0.37 (0.55)	0.51 (0.66)	1.54 (1.62)	0.90 (1.22)	0.87 (1.22)
HR	0.84 (0.08)	0.84 (0.11)	0.85 (0.08)	0.81 (0.15)	0.82 (0.17)	0.84 (0.10)
Median (IQR)						
RT	0.69 (0.11)	0.49 (0.12)***	0.47 (0.13)***	0.87 (0.76)***	0.53 (0.22)***, **	0.45 (0.45)***
HR	0.85 (0.10)	0.87 (0.17)	0.84 (0.12)	0.86 (0.16)	0.84 (0.13)	0.87 (0.13)

*$p < .05$, **$p < .01$,***=$p < .001$, black = VT, AT vs. VA, gray = same mode: reduced vs. increased noise

Post-hoc Questionnaire Ratings

Users indicated their level of agreement on a 7-point Likert scale with annotated end points 1 = “strongly disagree” and 7 = “strongly agree”.

Items on the Usefulness of Cues for Spatial and Temporal Perception

Table A.3 shows the items on cue usefulness for perceiving temporal events and the spatial location of the AR object. Items were included in the post-hoc questionnaire in study 1 (items 1 to 6) and study 2 (1, 7, 8).

Table A.3: Items on the usefulness of cues.

Items
1. The X proximity feedback helped me to perceive the direction from which the sphere was approaching the AR field of view
2. The X proximity feedback helped me to estimate the speed of the sphere that was approaching the AR field of view
3. The X transition feedback helped me to perceive the exact moment when the sphere entered the AR field of view
4. The X transition feedback helped me to perceive where the sphere entered the AR field of view
5. I did not have to concentrate much to interpret the X proximity feedback
6. I did not have to concentrate much to interpret the X transition feedback
7. The X proximity feedback helped me to be able to estimate early on, when exactly the sphere would reach the field of view
8. I could quickly react in response to the X transition feedback
Respective items were presented for each of the tested proximity cues (visual, audio) and transition cue (study 1: visual, audio, tactile; study 2: audio, tactile)

Table A.4 shows the comparison of mean ratings and standard deviations of visual, audio, and vibration **proximity cues** between conditions with reduced and with increased noise in study 1.

Table A.4: Ratings of proximity cues in study 1.

Proximity Feedback				
	Visual		Audio	
	Reduced Noise	Increased Noise	Reduced Noise	Increased Noise
Ease of perception	5.4 (1.4)**	2.8 (1.6)	6.2 (0.9)	5.9 (1.2)
Target direction estimation	5.9 (1.0)***	2.9 (1.7)	6.2 (0.8)	6.0 (0.8)
Target speed estimation	5.5 (1.5)***	2.9 (1.6)	5.2 (1.9)	4.8 (1.9)
Ease of interpretation	5.2 (1.5)**	3.0 (1.9)	6.2 (0.8)	5.9 (1.3)
No interference through noise	5.9 (0.7)**	1.7 (1.3)	6.5 (0.7)*	5.5 (1.5)
Overall cue suitability	5.7 (1.2)*	3.9 (1.6)	5.5 (1.6)	4.9 (1.7)

* p<.05, ** p <.01, *** p <.001

Table A.5 shows the comparison of mean ratings and standard deviations of visual, audio, and vibration **transition cues** between conditions with reduced and with increased noise in study 1.

Table A.5: Ratings of transition cues in study 1.

Transition Feedback						
	Visual		Audio		Vibration	
	Reduced Noise	Increased Noise	Reduced noise	Increased Noise	Reduced Noise	Increased Noise
Ease of perception	5.4 (1.4)**	2.8 (1.6)	6.5 (0.7)	6.4 (0.7)	6.8 (0.4)	6.8 (0.5)
Target direction estimation	6.1 (0.9)*	2.4 (1.8)	6.1 (0.9)	6.2 (0.8)	6.4 (0.7)	6.4 (0.5)
Perceive time of transition	5.5 (1.5)**	2.3 (1.8)	6.0 (0.8)	5.8 (0.9)	6.3 (0.7)	6.3 (0.8)
Ease of interpretation	5.1 (1.8)*	2.3 (1.7)	6.4 (0.8)	6.3 (0.8)	6.5 (0.7)	6.7 (0.7)
Overall cue suitability	5.2 (1.6)*	3.5 (1.9)	6.4 (0.7)	6.1 (0.7)	6.2 (0.8)	6.1 (1.2)

* p<.05, ** p <.01, *** p <.001.

Items on Comfort and Usability

Table A.6 shows the items on usability and comfort. User indicated their level of agreement on a 7-point Likert scale with annotated end points 1 = "strongly disagree" and 7 = "strongly agree". Items were included in the post-hoc questionnaire in study 1 and 2.

Table A.6: Items on usability and comfort.

Items on comfort and usability
1. My sitting position was comfortable
2. The interface was comfortable to wear
3. I could concentrate well on the task
4. I think I would prefer the same cue combination(s) if I performed the same study for the second time in the future
5. I enjoyed using the interface
6. Overall, the different feedback cues were easy to learn
7. Overall, the different feedback cues were easy to use
8. The task was not tiring to perform
9. I could imagine using one or more of the provided feedback combinations for a longer period of time
10. I expect to further improve in using the feedback once I get used to it

Table A.7 shows the mean ratings and standard deviations for items on comfort and usability in study 1 (Mode preference) and study 2 (Mode performance).

Table A.7: Usability and comfort ratings.

Statement	Study 1: Preference	Study 2: Performance
Sitting comfort	6.4 (0.7)	5.2 (1.1)
Wearing comfort	5.1(1.7)	3.6 (1.6)
Concentration	5.8 (0.8)	5.9 (1.3)
Confidence	5.8 (1.2)	-
Enjoyment	5.3 (1.2)	4.5 (1.5)
Ease of learning	6.3 (0.7)	4.0 (1.7)
Ease of interpretation	6.3 (0.7)	4.9 (1.4)
Not tiring	4.9 (1.7)	3.9 (2.0)
Long-term use	5.6 (1.4)	5.2 (1.3)
Improvement over time	5.6 (1.4)	5.0 (1.6)