Check for updates

Synergy and medial effects of multimodal cueing with auditory and electrostatic force stimuli on visual field guidance in 360° VR

Yasuhito Sawahata*, Masamitsu Harasawa and Kazuteru Komine

Science and Technology Research Laboratories, Japan Broadcasting Corporation, Tokyo, Japan

This study investigates the effects of multimodal cues on visual field guidance in 360° virtual reality (VR). Although this technology provides highly immersive visual experiences through spontaneous viewing, this capability can disrupt the quality of experience and cause users to miss important objects or scenes. Multimodal cueing using non-visual stimuli to guide the users' heading, or their visual field, has the potential to preserve the spontaneous viewing experience without interfering with the original content. In this study, we present a visual field guidance method that imparts auditory and haptic stimulations using an artificial electrostatic force that can induce a subtle "fluffy" sensation on the skin. We conducted a visual search experiment in VR, wherein the participants attempted to find visual target stimuli both with and without multimodal cues, to investigate the behavioral characteristics produced by the guidance method. The results showed that the cues aided the participants in locating the target stimuli. However, the performance with simultaneous auditory and electrostatic cues was situated between those obtained when each cue was presented individually (*medial effect*), and no improvement was observed even when multiple cue stimuli pointed to the same target. In addition, a simulation analysis showed that this intermediate performance can be explained by the integrated perception model; that is, it is caused by an imbalanced perceptual uncertainty in each sensory cue for orienting to the correct view direction. The simulation analysis also showed that an improved performance (*synergy effect*) can be observed depending on the balance of the uncertainty, suggesting that a relative amount of uncertainty for each cue determines the performance. These results suggest that electrostatic force can be used to guide 360° viewing in VR, and that the performance of visual field guidance can be improved by introducing multimodal cues, the uncertainty of which is modulated to be less than or comparable to that of other cues. Our findings on the conditions that modulate multimodal cueing effects contribute to maximizing the quality of spontaneous 360° viewing experiences with multimodal guidance.

1 Introduction

The presentation of multimodal sensory information in virtual reality (VR) can considerably enhance the sense of presence and immersion. In daily life, we perceive the surrounding physical world through multiple senses, such as visual, auditory, and haptic senses, and interact with it based on these perceptions (Gibson, 1979; Flach and Holden, 1998; Dalgarno and Lee, 2010). Therefore, introducing multimodal stimulations into VR can enhance realism and significantly improve the experience. In fact, numerous studies have reported the benefits of multimodal VR (Mikropoulos and Natsis, 2011; Murray et al., 2016; Wang et al., 2016; Martin et al., 2022; Melo et al., 2022).

Head-mounted displays (HMDs) offer a highly immersive visual experience by spontaneously allowing users to view a 360° visual world; however, this feature may disrupt the 360° viewing experience, causing users to miss important objects or scenes that are located outside their visual field, thereby resulting in the "out-of-view" problem (Gruenefeld, El Ali, et al., 2017b). In 360° VR, the visual field of the user is defined by the viewport of the HMD. As nothing is presented outside the viewport, users have no opportunity for perception without changing the head direction. To address this problem, the presentation of arrows (Lin Y.-C et al., 2017; Schmitz et al., 2020; Wallgrun et al., 2020), peripheral flickering (Schmitz et al., 2020; Wallgrun et al., 2020), and picture-in-picture previews and thumbnails (Lin Y. T. et al., 2017; Yamaguchi et al., 2021) have been employed and shown to guide the gaze and visual attention effectively. However, these approaches also exhibit the problem of inevitably interfering with the video content, potentially disrupting the spontaneous viewing experience and mitigating the benefit of 360° video viewing (Sheikh et al., 2016; Pavel et al., 2017; Tong et al., 2019). Addressing this problem will significantly improve the 360° video viewing experience, especially for VR content with fixed time events, such as live scenes, movies, and dramas.

Several studies have explored the potential of multimodal stimuli to guide user behavior in 360° VR while preserving the original content (Rothe et al., 2019; Malpica, Serrano, Allue, et al., 2020b). Diegetic cues based on non-visual sensory stimuli such as directional sound emanating from a VR scene provide natural and intuitive guidance that feels appropriate in VR settings (Nielsen et al., 2016; Sheikh et al., 2016; Rothe et al., 2017; Rothe and

Hußmann, 2018; Tong et al., 2019), exhibiting good compatibility with immersive 360° video viewing. At present, audio output is usually supported by any available HMDs and is the most common cue for visual field guidance. Because visual field guidance using non-visual stimuli is expected to provide high-quality VR experiences (Rothe et al., 2019), extensive research on various multimodal stimulation methods, including haptic stimulation, can aid the design of better VR experiences.

This study introduces electrostatic force stimuli to guide user behavior in selecting visual images that are displayed on the HMD (Figure 1). Previous studies have shown that applying an electrostatic force to the human body can induce a "fluffy" haptic sensation (Fukushima and Kajimoto, 2012a; 2012b; Suzuki et al., 2020; Karasawa and Kajimoto, 2021). Unlike some species of fish, amphibians, and mammals, humans do not possess electroreceptive abilities that allow them to perceive electric fields directly (Proske et al., 1998; Newton et al., 2019; Hüttner et al., 2023). However, as discussed in Karasawa and Kajimoto (2021), the haptic sensations that are produced through electrostatic stimulation are strongly related to the hair on the skin. Therefore, humans can indirectly perceive electrostatic stimulation through cutaneous mechanoreceptors (Horch et al., 1977; Johnson, 2001; Zimmerman et al., 2014), which are primarily stimulated by hair movements owing to electrostatic forces. Perceiving the physical world through cutaneous haptic sensations is a common experience in daily life, such as feeling the movement of air, and is expected to be a candidate method to guide user behavior naturally.

Many studies have proposed various methods of providing haptic sensations for visual field guidance, such as vibrations (Matsuda et al., 2020), normal forces on the face (Chang et al., 2018), and muscle stimulation (Tanaka et al., 2022), demonstrating that multimodal stimulation can improve the VR experience. Electrostatic force stimulation also provides haptic sensations, but can stimulate a relatively large area of the human body in a "fluffy" and subtle manner, which differs significantly from stimuli produced by other tactile stimulation methods, such as direct vibration stimulation through actuators. Karasawa and Kajimoto (2021) showed that electrostatic force stimulation can provide a feeling of *presence*. Previously, Slater (2009) and Slater et al. (2022) provided two views of immersive VR experiences, namely, place illusion (PI) and plausibility illusion (Psi), which refer to the

FIGURE 1
Visual field guidance in 360° VR using electrostatic force stimuli to mitigate the out-of-view problem. **(A)** Gentle visual field guidance. A user is viewing the scene depicted in the orange frame, whereas an important situation exists in the scene depicted in the red frame. Guiding the visual field to the proper direction will improve the user experience. **(B)** Haptic stimulus presentation using electrostatic forces. Electrostatic force helps the user to discover the important scene without affecting the original 360° VR content.

sensation of being in a real place and the illusion that the depicted scenario is actually occurring, respectively. In this sense, the effects of haptic stimulation on user experiences in VR belong to Psi. Such fluffy, subtle stimulation of the skin by electrostatic force has the potential to simulate the sensations of airflow, chills, and goosebumps, which are common daily-life experiences. The introduction of such modalities will enhance the plausibility of VR and lead to better VR experiences.

In this study, we presented electrostatic force stimuli using corona discharge, which is a phenomenon wherein ions are continuously emitted from a needle electrode at high voltages, allowing the provision of stimuli from a distance. Specifically, we placed the electrode above the user's head to stimulate a large area, from the head to the body (Figure 1B). Previous studies have employed plate- or pole-shaped electrodes to present such stimuli (Fukushima and Kajimoto, 2012a; 2012b; Karasawa and Kajimoto, 2021) and required the user to place their forearm close to the electrodes of the stimulation device, thereby limiting their body movement. The force becomes imperceptible even if the body parts are located 10 cm from the electrode (Karasawa and Kajimoto, 2021). As a typical VR user moves more than this distance, these conventional methods are not suitable for some VR applications that require physical movement. In addition, these devices are too bulky to be worn on the body. The proposed method can potentially overcome this limitation of distance and provide haptic sensations to VR users from a distance, thereby enabling the use of electrostatic force stimulation for visual field guidance in VR.

We evaluated the proposed visual field guidance method using multimodal cues in a psychophysical experiment. Previous studies have systematically evaluated visual field guidance using visual cues (Gruenefeld, Ennenga, et al., 2017a; Gruenefeld, El Ali, et al., 2017b; Danieau et al., 2017; Gruenefeld et al., 2018; 2019; Harada and Ohyama, 2022) in VR versions of visual search experiments (Treisman and Gelade, 1980; McElree and Carrasco, 1999). This study similarly investigated the effects of multimodal cues on visual searching.

Although numerous studies have shown that multiple modalities in VR can significantly improve the immersive experience (Ranasinghe et al., 2017; 2018; Cooper et al., 2018), it is unclear whether visual field guidance can also be improved by introducing multiple non-overt cues. We believe that multiple overt cues, such as visual arrows and halos, would help users to perform search tasks. However, this is not necessarily true for non-overt, subtle, and vague cues. Although guidance through subtle cues can minimize content intrusion (Bailey et al., 2009; Nielsen et al., 2016; Sheikh et al., 2016; Bala et al., 2019), it is not always guaranteed to be effective (Rothe et al., 2018). However, employing multiple subtle cues and integrating them into a coherent cue may provide effective overall guidance. In this study, in addition to electrostatic forces, we introduced weak auditory stimuli as subtle environmental cues to investigate the interaction effects of electrostatic and auditory cues on the guidance performance in VR as well as whether they improve, worsen, or have no effect on the guidance performance.

The nature of multimodal perception, which involves the integration of various sensory inputs to produce a coherent perception, has been understood using statistical models, such as maximum likelihood estimation and integration based on Bayes' theorem (Ernst and Banks, 2002; Ernst, 2006; 2007; Spence, 2011).

Although such computational modeling approaches are also expected to aid in comprehending the underlying mechanisms of multimodal cueing effects on visual field guidance, to the best of our knowledge, this aspect remains unexplored. Therefore, we adopted a similar approach using computational models and investigated the effects of various cueing conditions on visual field guidance. Thus, this study offers a detailed understanding of multimodal visual field guidance and knowledge for predicting user behavior under various cue conditions.

We first introduce electrostatic force and auditory stimuli as multimodal cues in a visual search task and then show that electrostatic force can potentially address the out-of-view problem. Because auditory stimuli have been commonly used in previous studies to guide user behavior (Walker and Lindsay, 2003; Rothe et al., 2017; Bala et al., 2018; Malpica, Serrano, Allue, et al., 2020a; Malpica, Serrano, Gutierrez, et al., 2020b; Chao et al., 2020; Masia et al., 2021), a baseline is provided for comparisons. In the visual search task, the participants were instructed to find a specific visual target as quickly as possible in 360° VR, both with and without sensory cues. We anticipated that the cueing would reduce the cumulative travel angles associated with updating the head direction during the search. Therefore, a comparison of the task performances in each condition revealed the effect of multimodal cueing on the visual field guidance.

In this study, we hypothesized that performance with multimodal cueing in the visual search task in VR would show one of the following three effects: 1) a performance improvement compared to that with electrostatic force or auditory cues (*synergy effect*); 2) the same performance as that with the better cue, not considering the performance worth the other cue (*masking effect*); and 3) performance between the individual performances with each cue (*medial effect*). We conducted a psychophysical experiment to investigate which of these effects were observed with multimodal cues. Subsequently, through the psychophysical experiment and an additional simulation analysis, we demonstrated that both the *synergy* and *medial effects* can be observed depending on the balance of perceptual uncertainties for each cue and the variance in the selection of the head direction. Finally, we investigated the conditions for effective multimodal visual field guidance.

2 Materials and methods

2.1 Visual search experiment with multimodal cues

This subsection describes the experiment that was conducted to investigate the effects of visual field guidance on visual search performance in 360° VR using haptic and auditory cues. In addition, the multimodal effects of simultaneous cueing using haptic and auditory stimuli were investigated. The search performance was measured based on the travel angles, which are the cumulative rotation angles of the head direction, as described in detail in Section 2.2.1.1. Finally, we determined which of the effects, namely, *synergy*, *medial*, or *masking*, were likely by comparing the travel angles obtained in each cue condition.

2.1.1 Participants

Fifteen participants (seven male, eight female; aged 21–33 years, mean: 24.4) were recruited for this experiment. All participants had normal or corrected-to-normal vision. Two participants were excluded because their psychological thresholds for the electrostatic force stimuli were too high and exceeded the intensity range that our apparatus could present. Informed consent was obtained from all participants, and the study design was approved by the ethics committee of the Science and Technology Research Laboratories, Japan Broadcasting Corporation.

2.1.2 Apparatus

A corona charging gun (GC90N, Green Techno, Japan) was used to present electrostatic force stimuli. This device comprises a needle-shaped electrode (ion-emitting gun) and a high-voltage power supply unit (rated voltage range: 0 to −90 [kV]). The electrostatic force intensity was modulated by adjusting the applied voltage. The gun was hung from the ceiling and placed approximately 50 cm above the participant's head, as shown in Figure 1B. In addition, the participant wore a wristband attached to the ground to avoid accidental shocks owing to unintentional charging.

A standalone HMD, Meta Quest 2 (Meta, United States), was used to present the 360° visual images and auditory stimuli, and the controller joystick (for the right hand) was used to collect the responses. The HMD communicated with the corona charging gun via an Arduino-based microcomputer (M5Stick-C PLUS, M5Stack Technology, China) to control the analog inputs for the gun. The delay between the auditory and electrostatic force stimuli was a maximum of 20 ms, which was sufficiently small to perform the task. The participants viewed the 360° images while sitting in a swivel chair to facilitate viewing. They wore wired earphones (SE215, SURE, United States), which were connected to the HMD and used to present auditory stimuli using functions provided in Unity (Unity Technologies, United States) throughout the experiment, even when no auditory stimuli were presented. The experimental room was soundproof. Participant safety was duly considered; the floor was covered with an electrically grounded conductive mat, which collected ions that were not meant for the participant, thereby preventing unintentional charging of other objects in the room.

2.1.3 Stimuli
2.1.3.1 Visual stimuli

The target and distractor stimuli were presented in a VR environment implemented in Unity (2021.3.2 f1). The target stimulus included a randomly selected white symbol among "├", "┤", "┬", and "┴," whereas the distractor stimuli included white "┼" symbols. These stimuli were displayed on a gray background and distributed within a range of -10°–10° relative to each intersection of the latitudes and longitudes of a sphere with a 5-m radius that was centered at the origin. The referential latitudes and longitudes were placed at each 36° position of the horizontal 360° view and 22.5° positions between the elevation angles of −45° and 45°. Thus, 1 target and 39 distractor stimuli were presented at 10 × 4 locations. The stimuli sizes were randomly selected from visual angles ranging from 2.86° ± 1.43°, both horizontally and vertically. The difficulty of the task was modulated by varying the stimulus size and placement and

FIGURE 2
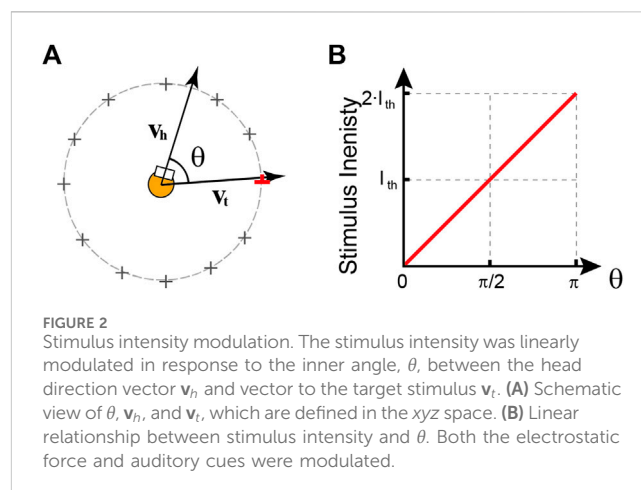Stimulus intensity modulation. The stimulus intensity was linearly modulated in response to the inner angle, θ, between the head direction vector $\mathbf{v}_h$ and vector to the target stimulus $\mathbf{v}_t$. **(A)** Schematic view of θ, $\mathbf{v}_h$, and $\mathbf{v}_t$, which are defined in the *xyz* space. **(B)** Linear relationship between stimulus intensity and θ. Both the electrostatic force and auditory cues were modulated.

the parameter values were selected based on our preliminary experiments.

2.1.3.2 Electrostatic force stimuli

In this study, the electrostatic force stimuli are referred to as haptic stimuli induced by the corona charging gun. The electrostatic force intensity was determined based on the gun voltage. We selected the physical intensity of the electrostatic force for each participant based on their psychological threshold; the intensity ranged from zero to twice the threshold. Thus, we ensured that the stimulus intensity was psychologically equivalent among all participants. The threshold I_{th}, which was largely dependent on each participant, was measured before the experiments using the method of staircase, and it typically ranged from −10 to −30 kV. We linearly modulated the stimulus intensity in response to the inner angle θ between the head-direction vector $\mathbf{v}_h$ and target stimulus vector $\mathbf{v}_t$, as shown in Figure 2A. When the target was in front, i.e., $\theta = 0$, no electrostatic force was presented, whereas when it was behind, i.e., $\theta = \pi$, the strongest electrostatic force of $2I_{th}$ was presented. Therefore, the electrostatic force was regarded as a cue stimulus because participants could potentially find the target stimulus by updating their head direction to avoid the subtle haptic sensations. That is, when the haptic sensations were sufficiently weak, the target stimulus was likely to be within the participant's visual field. This is the natural behavior of most people because a strong electrostatic stimulus is typically considered unpleasant.

2.1.3.3 Auditory stimuli

Monaural white noise was used as the auditory stimulus. We used the same modulation method for the auditory stimuli as that for the electrostatic force stimuli, as shown in Figure 2. Specifically, we linearly modulated the stimulus intensity in response to the inner angle θ between the head-direction vector $\mathbf{v}_h$ and target stimulus vector $\mathbf{v}_t$. When the target was in front, i.e., $\theta = 0$, no sound was presented, whereas when it was behind, i.e., $\theta = \pi$, the maximum amplitude (volume) of the stimulus of $2I_{th}$ was presented. As with the electrostatic force stimuli, the threshold I_{th} for auditory stimuli was measured for each participant before the experiments using the method of staircase.

2.1.4 Task and conditions

We designed a within-participant experiment to compare the effects of haptic and auditory guidance in a visual search task. The participants were instructed to find the target stimulus and indicate its direction using the joystick on the VR controller. For example, when they discovered a target stimulus "⊥," they tilted the joystick upward as quickly as possible. The trial was terminated once the joystick was manipulated. Feedback was provided between sessions, showing the success rate of the previous session, to encourage participants to complete the task. The task was conducted both with and without sensory cues, resulting in four conditions based on the combinations of cue stimuli: visual only (V), vision with auditory (A), vision with electrostatic force (E), and vision with auditory and electrostatic force (AE) cues.

2.1.5 Procedure

The experiment included 12 sessions comprising 12 visual search trials, for a total of 144 trials per participant. Therefore, each condition (V, A, E, and AE) was presented 36 times in one experiment. In three of the 12 sessions, only condition V was presented, whereas in the other sessions, conditions A, E, and AE were presented in a pseudo-random order. Before each session, we informed the participants whether the next session would be a V-only session or a session with the non-visual-cued conditions. This prevented participants from waiting for non-visual cues during condition V and inadvertently wasting search time.

Each trial comprised a rest period of variable-length (3–6 s) and a 10-s search period. In the rest period, 40 randomly generated distractors were presented, whereas in the following search period, one of the distractors was replaced with a target stimulus. The trials progressed as soon as the target stimulus was found or when the 10-s time limit was reached. Note that the participants underwent two practice sessions to understand the task and response methods prior to these sessions.

2.2 Analysis

2.2.1 Behavioral data analysis
2.2.1.1 Modeling

We recorded the participants' responses and extents of their head movements during the search period. The trials with a correct response were labeled as successful, whereas those with an incorrect or no response were labeled as failed. The travel angle was defined as the accumulated rotational changes in the head direction during the target search. If guidance by electrostatic forces and auditory cues is effective, the travel angles should be shorter than those with no cues. Therefore, we investigated the modulation efficiency of the target discovery according to cue type.

The travel angle allowed us to model the participants' behavior in the visual search experiment with non-overt multimodal cues appropriately. In the original visual search experiment (Treisman and Gelade, 1980; McElree and Carrasco, 1999), wherein participants had to find the target stimuli with specified visual features as quickly as possible, the performance was measured by the reaction time required for identification. These experimental paradigms have recently been extended to investigate user behavior in VR. Cue-based visual search experiments in VR involve the

analysis of reaction times and/or movement angles towards a target object (Gruenefeld, Ennenga, et al., 2017a; Gruenefeld, El Ali, et al., 2017b; Danieau et al., 2017; Gruenefeld et al., 2018; 2019; Schmitz et al., 2020; Harada and Ohyama, 2022). In addition, previous studies employed overt cues that directly indicated the target location, whereas we employed non-overt cues that weakly indicated them, without interfering with the visuals. This difference could have affected the behavior of participants, depending on their individual traits. For example, some participants may have adopted a scanning strategy wherein they sequentially scanned the surrounding visual world, ignoring the cues because they considered subtle cues to be unreliable. Participants with better physical ability could have completed the task faster using this strategy. In such cases, the reaction time would not accurately reflect the effects of cueing on the visual search performance and the effects would differ significantly from those we were investigating. Because behaviors including scanning that are not based on presented cues would result in larger travel angles, the effects of cues would likely be better reflected in the travel angle than in the reaction time. Therefore, we employed travel angles instead of reaction times to evaluate the performance.

We employed Bayesian modeling to evaluate the efficacy of each cue, as follows:

$$p(k|\lambda, \Phi) = \frac{(\lambda\Phi)^k}{k!}e^{-\lambda\Phi}, \tag{1}$$

where k is the number of discoveries (successful trials), λ is the expected target discovery rate, and Φ is the total travel angle. The probability of k given λ and Φ was calculated using the Poisson process (see 2.2.1.2). Note that $\Phi = \sum_{i=1}^{n}\phi_i$, where n is the number of trials and ϕ_i is the travel angle during the i-th trial. By applying the Bayes theorem to Eq. 1, the posterior distribution of λ can be expressed as $p(\lambda|k, \Phi) \propto p(k|\lambda, \Phi)p(\lambda)$. By assuming a noninformative prior on $p(\lambda)$, $p(\lambda|k, \Phi)$ is proportional to the right side of Eq. 1. Therefore, the expectation of $p(\lambda|k, \Phi)$ represents the target discovery rate, as follows:

$$\lambda = E[p(\lambda|k, \Phi)] = \frac{k}{\Phi} \tag{2}$$

Thus, λ was interpreted as the number of discoveries per travel angle.

2.2.1.2 Poisson process model derivation

The total travel angle Φ was divided into N bins of width $\Delta\phi = \Phi/N$. The probability of finding a target stimulus in a bin with the expected λ is $\Delta\phi\lambda$. Therefore, the probability of finding targets in k from N bins is represented by the following binomial distribution:

$$p(k|\lambda) = \frac{N!}{(N-k)!k!}(1 - \Delta\phi\lambda)^{N-k}(\Delta\phi\lambda)^k. \tag{3}$$

By minimizing the bin width using $N \to \infty$, we obtain $\lim_{N\to\infty}(1 - \Delta\phi\lambda)^{N-k} = \lim_{\epsilon\to0}(1 + \epsilon)^{-1/\epsilon} = e^{-\lambda\Phi}$ using the following relationships: $\epsilon = -\Delta\phi\lambda$ and $N - k \approx N = \Phi/\Delta\phi$. In addition, a relationship exists between $N!/(N-k)! \approx N^k = (\Phi/\Delta\phi)^k$ and $N \to \infty$. Finally, we obtain the following Poisson process:

$$p(k|\lambda) = \frac{1}{k!}\left(\frac{\Phi}{\Delta\phi}\right)^k e^{-\lambda\Phi}(\Delta\phi\lambda)^k = \frac{(\lambda\Phi)^k}{k!}e^{-\lambda\Phi}. \tag{4}$$

FIGURE 3
Perceptual model of visual search with multimodal cues. The possible head directions were estimated separately based on the synthesized auditory and electrostatic force sensations generated by $g_a(\cdot)$ and $g_e(\cdot)$. The final head direction in each iteration was determined by averaging the estimated directions.
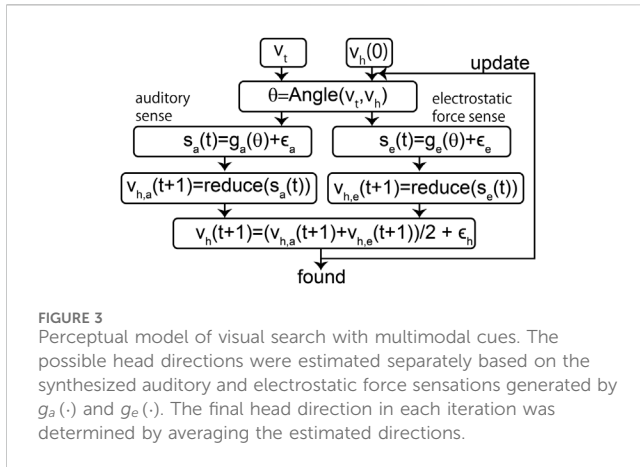
2.2.1.3 Statistics

We created a dataset by pooling all observations that were obtained from the participants. Thereafter, we obtained the posterior distributions of the target discovery rate, $p(\lambda|k, \Phi)$, for each condition. Subsequently, the significance of the visual field guidance was assessed by comparing the distribution shapes. For example, when λ was larger for condition E than that for condition V and their distributions overlapped slightly, we concluded that the electrostatic force-based guidance significantly affected the visual field guidance. The overlap was quantified by the area under the curve (AUC) metric, the value of which ranged from 0 to 1; a smaller overlap resulted in an AUC value closer to 1. We compared the posterior distribution of λ in condition AE with those in conditions A and E to identify the multimodal effect.

2.3 Simulation analysis

2.3.1 Overview

To better comprehend how participants processed the multimodal inputs in the experiment, we conducted a simulation analysis assuming a perceptual model wherein a participant determined the head direction by simply averaging two vectors directed towards the target induced through auditory and haptic sensations, as shown in Figure 3, constituting the most typical explanation of the multimodal effect (Ernst and Banks, 2002; Ernst, 2006; 2007). We manipulated the noise levels ϵ_a, ϵ_e, and ϵ_h, assumed for the auditory sensations, haptic sensations, and orienting head directions, respectively, as shown in Figure 3. Thereafter, we examined the relationship between the noise levels and target discovery rates for each stimulus condition.

We implemented a computational model to determine the target stimulus direction based on the synthesized sensations. The head direction vector at time t is represented as $\mathbf{v}_h(t)$ ($\|\mathbf{v}_h(t)\| = 1$), and the auditory and electrostatic force sensory inputs for the model are denoted as $s_a(t)$ and $s_e(t)$, respectively. The model estimated the next head direction $\mathbf{v}_h(t+1)$ such that the sensory inputs were reduced. By iterating these procedures, $s_a(\cdot)$ and/or $s_e(\cdot)$ were minimized and the target stimulus in the direction of $\mathbf{v}_t$ could be identified. The detailed procedure is presented in Section 2.3.2.

The simulation was initially conducted using randomly generated $\mathbf{v}_h(t)$ and $\mathbf{v}_t$ values. The search was iterated according to the synthesized sensations $s_a(t)$ and $s_e(t)$, using different noise levels ϵ_a and ϵ_e, as shown in Figure 3. To simulate multimodal processing, the model estimated $\mathbf{v}_h(t+1)$ by averaging the $\mathbf{v}_{h,a}(t+1)$ and $\mathbf{v}_{h,e}(t+1)$ estimations. The term $\boldsymbol{\epsilon}_h \in \mathbb{R}^3$ was introduced to represent orienting errors between the estimated and actual directions owing to the physical constraints and other factors during the real experiment. Note that in the unimodal conditions, $\mathbf{v}_h(t+1) = \mathbf{v}_{h,a}(t+1)$ or $\mathbf{v}_h(t+1) = \mathbf{v}_{h,e}(t+1)$. An inner angle of $< \pm 30°$ between $\mathbf{v}_h(t+1)$ and $\mathbf{v}_t$ indicated that the target stimulus was found and the iteration was terminated.

We ran the simulation using the parameter settings that were closest to those used in the real experiment; for example, the maximum amount and speed of head rotation and the number of trials were appropriately selected. The simulation was performed 468 times for each condition, corresponding to the setup in the real experiment (36 trials × 13 participants). The travel angle and target discovery rate were computed using the methods described in Section 2.2.1.1. To examine the effects of ϵ_a, ϵ_e, and ϵ_h on λ for conditions A, E, and AE, we generated them using the following parameters: $\epsilon_a \sim \mathcal{N}(0, \sigma_a^2)$ with $0.05^2 < \sigma_a^2 < 0.50^2$, $\epsilon_e \sim \mathcal{N}(0, \sigma_e^2)$ with $0.05^2 < \sigma_e^2 < 0.50^2$, and $\boldsymbol{\epsilon}_h \sim \mathcal{N}(0, \sigma_h^2 \mathrm{I})$ with $\sigma_h^2 = 0.01^2$ and 0.1^2.

2.3.2 Procedure

In this section, we describe the details of the simulation, as summarized in Section 2.3.1 and Figure 3.

The simulation model iteratively updated the head direction vector $\mathbf{v}_h(\cdot)$ based on synthetic sensory inputs. The next head direction $\mathbf{v}_h(t+1)$ was determined using two steps: first, the possible head directions that minimized the target vector ($\mathbf{v}_t$) error were estimated independently for each modality, and thereafter, $\mathbf{v}_h(t+1)$ was obtained by averaging the estimated directions. In reality, because $\mathbf{v}_t$ was unknown, it was substituted with its estimate, which was obtained using an auditory or electrostatic force sensation, i.e., $\hat{\mathbf{v}}_{t,a}$ or $\hat{\mathbf{v}}_{t,e}$, respectively. Thus, the model determined the next head direction using a gradient descent search, as follows:

$$\mathbf{v}_{h,a}(t+1) = \mathbf{v}_h(t) - \alpha \nabla \left(\|\hat{\mathbf{v}}_{t,a}(t) - \mathbf{v}_h\|^2 \right)\Big|_{\mathbf{v}_h = \mathbf{v}_h(t)}, \quad (5)$$

$$\mathbf{v}_{h,e}(t+1) = \mathbf{v}_h(t) - \alpha \nabla \left(\|\hat{\mathbf{v}}_{t,e}(t) - \mathbf{v}_h\|^2 \right)\Big|_{\mathbf{v}_h = \mathbf{v}_h(t)}, \quad (6)$$

where $\alpha (> 0)$ is a step-size parameter. The value of α corresponds to the head rotation speed during the experiment. Thus, Eqs 5, 6 can be rewritten as:

$$\mathbf{v}_{h,a}(t+1) = \mathbf{v}_h(t) - 2\alpha \left(\hat{\mathbf{v}}_{t,a}(t) - \mathbf{v}_h(t) \right), \quad (7)$$

$$\mathbf{v}_{h,e}(t+1) = \mathbf{v}_h(t) - 2\alpha \left(\hat{\mathbf{v}}_{t,e}(t) - \mathbf{v}_h(t) \right). \quad (8)$$

Finally, the next head direction vector was obtained as follows:

$$\mathbf{v}_h(t+1) = \frac{1}{2} \left(\mathbf{v}_{h,a}(t+1) + \mathbf{v}_{h,e}(t+1) \right) + \boldsymbol{\epsilon}_h, \quad (9)$$

where $\boldsymbol{\epsilon}_h$ follows $\mathcal{N}(0, \sigma_h^2 \mathrm{I})$ and represents the fluctuations associated with the head motion. Note that Eq. 9 is normalized before the next iteration. In unimodal simulations, Eq. 9 can be substituted with $\mathbf{v}_h(t+1) = \mathbf{v}_{h,a}(t+1) + \boldsymbol{\epsilon}_h$ or $\mathbf{v}_h(t+1) = \mathbf{v}_{h,e}(t+1) + \boldsymbol{\epsilon}_h$.

$\mathbf{v}_t$ was estimated using past auditory and somatosensory observations $s_a(t), ..., s_a(t - N)$ and $s_e(t), ..., s_e(t - N)$, respectively, where N is the number of observations used for the estimation. Because the stimulus intensities are given by the inner angle between the head and target directions, the simulated sensory inputs $s_a(t)$ and $s_e(t)$ can be expressed as

$$s_a(t) = \cos^{-1}(\mathbf{v}_t \cdot \mathbf{v}_h(t)) + \epsilon_a, \tag{10}$$

$$s_e(t) = \cos^{-1}(\mathbf{v}_t \cdot \mathbf{v}_h(t)) + \epsilon_e, \tag{11}$$

where ϵ_a and ϵ_e are the noise terms that follow normal distributions with $\mathcal{N}(0, \sigma_a^2)$ and $\mathcal{N}(0, \sigma_e^2)$, respectively, and σ_a^2 and σ_e^2 indicate the amount of noise generated.

We define a head-direction matrix $V_h = [\,\mathbf{v}_h(t) \,\cdots\, \mathbf{v}_h(t - T)\,]^{\mathrm{t}}$ and observation vectors $\mathbf{s}_a = [\,s_a(t) \,\cdots\, s_a(t - T)\,]^{\mathrm{t}}$ and $\mathbf{s}_e = [\,s_e(t) \,\cdots\, s_e(t - T)\,]^{\mathrm{t}}$. If we assume that $\mathbf{s}_a$ and $\mathbf{s}_e$ are negatively correlated with $V_h\mathbf{v}_t$, we can estimate $\mathbf{v}_t$ such that $\mathbf{s}_a^{\mathrm{t}}V_h\mathbf{v}_t$ and $\mathbf{s}_e^{\mathrm{t}}V_h\mathbf{v}_t$ are minimized under the constraint of $\|\mathbf{v}_t\| = 1$, assuming that V_h, $\mathbf{s}_a$, and $\mathbf{s}_e$ are centered in advance. Letting $\mathbf{v}_{t,a}$ and $\mathbf{v}_{t,e}$ be the target vectors obtained through auditory and haptic signals, the estimation is then tractable using the method of the Lagrange multiplier method, as follows:

$$L_a = \mathbf{s}_a^{\mathrm{t}}V_h\mathbf{v}_{t,a} + \lambda_a(\mathbf{v}_{t,a}^{\mathrm{t}}\mathbf{v}_{t,a} - 1), \tag{12}$$

$$L_e = \mathbf{s}_e^{\mathrm{t}}V_h\mathbf{v}_{t,e} + \lambda_e(\mathbf{v}_{t,e}^{\mathrm{t}}\mathbf{v}_{t,e} - 1), \tag{13}$$

where L_a and L_e are the Lagrangian functions for each modality, and λ_a and λ_e are the Lagrange multipliers. By considering $\hat{\mathbf{v}}_{t,a}$ and $\hat{\mathbf{v}}_{t,e}$ as the estimators of $\mathbf{v}_{t,a}$ and $\mathbf{v}_{t,e}$, respectively, in Eqs 12, 13, we obtain their values by considering $\partial L_a/\partial\mathbf{v}_{t,a} = \mathbf{0}$ and $\partial L_e/\partial\mathbf{v}_{t,e} = \mathbf{0}$ with $\|\mathbf{v}_{t,a}\| = 1$ and $\|\mathbf{v}_{t,e}\| = 1$, respectively:

$$\hat{\mathbf{v}}_{t,a} = -\frac{V_h^{\mathrm{t}}\mathbf{s}_a}{\sqrt{\mathbf{s}_a^{\mathrm{t}}V_hV_h^{\mathrm{t}}\mathbf{s}_a}}, \tag{14}$$

$$\hat{\mathbf{v}}_{t,e} = -\frac{V_h^{\mathrm{t}}\mathbf{s}_e}{\sqrt{\mathbf{s}_e^{\mathrm{t}}V_hV_h^{\mathrm{t}}\mathbf{s}_e}}. \tag{15}$$

Therefore, by substituting $\hat{\mathbf{v}}_{t,a}(t)$ and $\hat{\mathbf{v}}_{t,e}(t)$ in Eqs 7, 8 with Eqs 14, 15, the next head direction vectors could be estimated.

Initially, $\mathbf{v}_t$ and $\mathbf{v}_h(0)$ were randomly selected from the 360° omnidirectional candidates. In addition, $\mathbf{v}_h(t)$ and $1 < t < N$ were also generated around $\mathbf{v}_h(0)$. Based on the sampled $\mathbf{v}_t$ and $\mathbf{v}_h(t)$, synthetic sensations were generated using Eqs 9–11.

The target search was conducted using a maximum of 1000 steps. The simulation parameter values of $N = 10$ and $\alpha = 0.01$ were selected to ensure that the target discovery rates were similar to those observed in the real experiments.

3 Results

3.1 Behavioral results

We pooled all data obtained from the 13 participants. The number of successful trials k and accumulated travel distances Φ for each condition were $k = 352$ and $\Phi = 3.12 \times 10^3$ for V; $k = 422$ and $\Phi = 2.44 \times 10^3$ for A; $k = 391$ and $\Phi = 2.80 \times 10^3$ for E; and $k = 417$ and $\Phi = 2.66 \times 10^3$ for AE. Note that the numbers of trials

with response errors, which appeared to be owing to manipulation errors, were: 6, 7, 10, and 4 for the V, A, E, and AE, respectively. Figure 4 shows histograms of the successful and failed trials plotted against the travel angles, wherein the blue and red plots denote successful and failed trials, respectively. The target stimuli were identified in all conditions even if the travel angles were short or close to zero because they could appear in the participant's visual field at the beginning of the trial, as the target locations were determined randomly. The failed trials featured longer travel angles, suggesting that the participants looked around but were unable to complete the task within the time limit.
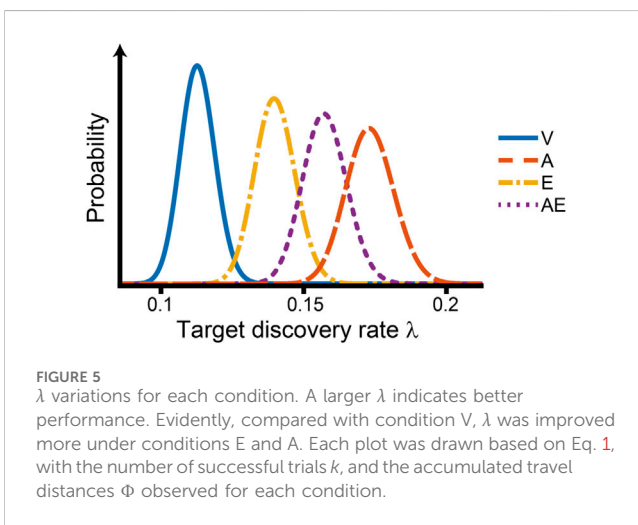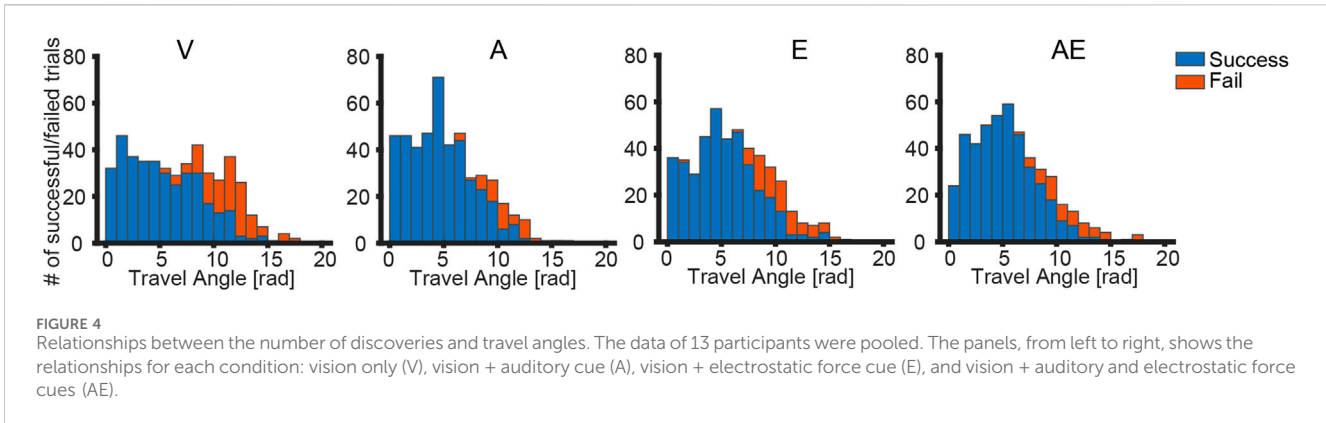
Figure 5 shows the posterior distributions of λ for each condition. The expected discovery rates λ_V, λ_A, λ_E, and λ_{AE} for each condition were 0.113, 0.173, 0.140, and 0.157, respectively. As predicted, guidance with the auditory cues significantly improved the target discovery rate compared to condition V, and no overlap was observed between the distributions. In addition, the target discovery rate improved significantly in condition E compared with condition V, although not as much as that in condition A. The AUC between the distributions of conditions V and E was 0.998, suggesting that λ for condition E was significantly higher than that for condition V. In addition, there was no overlap between the distributions of condition V and the other conditions, A and AE, indicating that the AUCs were 1. This indicates that visual field guidance using electrostatic force is effective even in a VR environment wherein users view a 360° world using both head and body movements.

We observed that the performance in condition AE was situated between those in conditions A and E, thereby rejecting the possibilities of *synergy* and *masking effects* because the search using both cues did not enhance the performance and the participants could not ignore the other cue. This result supports the *medial effect*, which was one of the anticipated candidates.

3.2 Simulation results

Figure 6 shows λ for each cue condition, plotted against the uncertainty ratio for the AE cues for varying electrostatic force uncertainty under constant auditory cue uncertainty. σ_a^2, σ_e^2, and σ_h^2 represent the perceptual uncertainty, i.e., the variances for ϵ_a, ϵ_e, and ϵ_h, respectively, as shown in Figure 3. Summaries of the parameters and observations shown in Figures 6A, D are presented in Tables 1, 2, respectively.

As shown in Figures 6A, D, the expected λ values for unimodal cueing in condition E and multimodal cueing in condition AE asymptotically decreased as σ_e^2/σ_a^2 increased in both cases with $\sigma_h^2 = 0.01^2$ and 0.1^2. The *medial effect*, which was elicited in our behavioral data, was particularly observed for larger ratios of σ_e^2/σ_a^2 and σ_h^2, as shown in Figures 6D, F. Notably, when the σ_e^2/σ_a^2 ratio was close to 1 under $\sigma_h^2 = 0.01^2$ (Figures 6A, B), we observed the *synergy effect*, wherein λ with multimodal cueing was better than that with unimodal cueing. The magnitude of the σ_e^2/σ_a^2 ratio indicates the bias level of the uncertainty of the electrostatic force sensation over the auditory sensation. Therefore, these results suggest that both *medial* and *synergy effects* were observable under the assumption of the typical multimodal integration model (Ernst and Banks, 2002; Ernst, 2006; 2007) (Figure 3), depending on the uncertainty bias for each sensation and head motion.

FIGURE 4
Relationships between the number of discoveries and travel angles. The data of 13 participants were pooled. The panels, from left to right, shows the relationships for each condition: vision only (V), vision + auditory cue (A), vision + electrostatic force cue (E), and vision + auditory and electrostatic force cues (AE).

FIGURE 5
λ variations for each condition. A larger λ indicates better performance. Evidently, compared with condition V, λ was improved more under conditions E and A. Each plot was drawn based on Eq. 1, with the number of successful trials k, and the accumulated travel distances Φ observed for each condition.

4 Discussion

We demonstrated the multimodal effects of AE cues on visual field guidance in 360° VR, and found that both *medial* and *synergy effects* were observable depending on the uncertainty of the cue stimuli through the psychophysical experiment and the simulation analysis. Specifically, guidance performance with multimodal cueing is modulated by balancing the perceptual uncertainty elicited by each cue stimulus. We also demonstrated that the applicability of the electrostatic force-based stimulation method in VR applications; electrostatic stimulation through the corona charging gun allowed users to make large body movements. These results suggest that multimodal cueing with electrostatic force has sufficient potential to guide user behavior in 360° VR gently, offering a highly immersive visual experience through spontaneous viewing.

We showed that electrostatic force can be used as a haptic cue to guide the visual field. However, the search performance did not reach that with the auditory cue, even though we selected cue intensities that varied equally in small ranges around the supra- and sub-thresholds, with no significant difference in the perceptual domain. In the informal post-experiment interviews, some participants reported that the sensation induced by the electrostatic force was attenuated, especially while moving. In addition, most participants reported that the auditory

cue made it easier to identify the target location. This suggests that the haptic sensation was affected by body motion that inevitably accompanied the updating of the head direction. The increased uncertainty for the haptic sensation was estimated to be approximately five times greater than that for the auditory sensation, as suggested by the simulation results (Figure 6F). Thus, the perception of changes in the stimulus intensity associated with visual field updates acts as a cue for estimating the target direction, which means that increasing the electrostatic field intensity such that it is strong enough to resist the effects of body motion could mitigate this uncertainty. As suggested by the simulation results presented in Section 3.2, reducing the perceptual uncertainty improves the search performance. This finding has been overlooked in previous studies that mainly focused on visual field guidance using overt cue stimuli (Gruenefeld, Ennenga, et al., 2017a; Gruenefeld, El Ali, et al., 2017b; Danieau et al., 2017; Gruenefeld et al., 2018; 2019; Harada and Ohyama, 2022). This results in the requirement for the property of cue stimuli to improve performance in multimodal visual field guidance.

The *medial effect* might have been counterintuitive because participants received more information regarding the target stimulus from multimodal cues than unimodal cues. Because the cues conveyed the same information, the *synergy effect* was more likely if participants used the received information properly. The simulation analysis showed that both effects could be observed under specific noise settings. This can also be explained theoretically: let A and E be random variables for auditory and electrostatic force sensations, respectively. Then, $V(A)$ and $V(E)$ represent the variances for each sensation. According to the integrated perception model (Ernst and Banks, 2002; Ernst, 2006; 2007), the total sensation variance can be expressed as

$$V\left(\frac{A+E}{2}\right) = \frac{1}{4}\left(V(A) + V(E) + 2Cov(A,E)\right), \qquad (16)$$

where $Cov(A,E)$ denotes the covariance between A and E. If A and E are independent, i.e., $Cov(A,E) = 0$, and $V(A)$ and $V(E)$ are equal, according to Eq. 16, $V((A+E)/2)$ is less than both $V(A)$ and $V(E)$, indicating a more efficient search performance than unimodal cueing (*synergy effect*) because smaller variances improve the performance. For example, if $V(E) = 5V(A)$, $V((A+E)/2)$ should be $3/2 \cdot V(A)$, suggesting intermediate performance if $V(A) < V((A+E)/2) < V(E)$ is used (*medial effect*). However, if $V(A)$ and $V(E)$ are not independent and $Cov(A,E)$ has a certain value, $V((A+E)/2)$
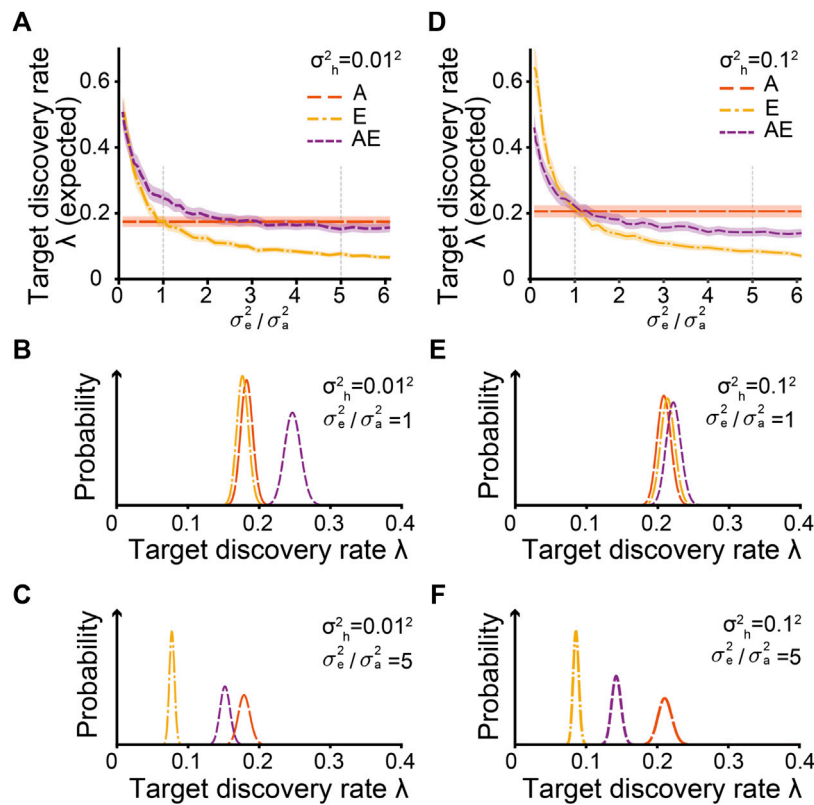
FIGURE 6
Comparisons of λ in the simulation analysis. **(A–C)** and **(D–F)** show the effects of $\sigma_h^2 = 0.01^2$ and 0.1^2 on λ, respectively. The expected λ values under each condition are plotted against the σ_e^2/σ_a^2 ratio using a representative value of $\sigma_a^2 = 0.17^2$. As no electrostatic-force stimuli were presented in condition A, λ could not be technically plotted against σ_e^2/σ_a^2. However, for reference, as λ for condition A was independent of σ_e^2, we plotted the expected λ values for condition A as straight lines through σ_e^2/σ_a^2 using a constant σ_a^2 value. The shaded areas behind the plots denote 95% credible intervals of the posterior distribution of λ. **(B,E)**, and **(C,F)** show the posterior distributions of λ for $\sigma_e^2/\sigma_a^2 = 1$ and 5, respectively. The original values presented in **(A–C)** and **(D–F)** are shown in Tables 1, 2, respectively.

increases and the *synergy effect* fades. In reality, ϵ_h in Figure 3 controlled the dependence between A and E, as the variance of ϵ_h is determined based on the observation of *synergy* or *medial effects*. These results support the validity of the integrated perception model shown in Figure 3 as the underlying mechanism of visual search tasks with multimodal cues.

Addressing the out-of-view problem has been a major challenge in 360° VR video viewing (Lin Y. T. et al., 2017; Schmitz et al., 2020; Wallgrun et al., 2020; Yamaguchi et al., 2021). Gentle and diegetic guidance that does not interfere with the visual content has received substantial attention from VR content providers (Nielsen et al., 2016; Sheikh et al., 2016; Rothe et al., 2017; Rothe and Hußmann, 2018; Bala et al., 2019; Tong et al., 2019). This study showed that subtle cues using artificial electrostatic force can guide the visual field, thereby demonstrating the application potential for 360° VR. Whereas previous studies using static electricity have severely limited the movements of the user (Fukushima and Kajimoto, 2012b; 2012a; Karasawa and Kajimoto, 2021), the use of the corona discharge phenomenon mitigated this limitation. The simulation analysis using the computational model helped to provide an understanding of the mechanisms of multimodal cueing. Similar to the observations in this study, previous studies using non-overt cues with perceptual uncertainty have reported both positive and negative effects of

TABLE 1 Summary of parameters and observations in simulation analysis ($\sigma_h^2 = 0.01^2$).

σ_e^2	σ_e^2/σ_a^2	E			AE		
		k	Φ	λ	k	Φ	λ
0.05^2	0.09	468	952.9	0.491	468	921.3	0.508
0.08^2	0.22	468	1105.4	0.423	468	1133.7	0.413
0.11^2	0.42	468	1583.2	0.296	468	1362.7	0.343
0.14^2	0.68	468	2176.8	0.215	468	1741.6	0.269
0.17^2	1.00	468	2651.1	0.177	468	1897.4	0.247
0.20^2	1.38	468	3106.3	0.151	468	2093.9	0.224
0.23^2	1.83	467	3761.7	0.124	468	2339.1	0.200
0.26^2	2.34	466	4249.6	0.110	468	2577.8	0.182
0.29^2	2.91	464	4691.1	0.099	468	2616.1	0.179
0.32^2	3.54	461	5150.7	0.090	468	2837.3	0.165
0.35^2	4.24	460	5707.8	0.081	467	2818.1	0.166
0.38^2	5.00	457	5920.3	0.077	468	3082.3	0.152
0.41^2	5.82	446	6698.8	0.067	467	3045.0	0.154

TABLE 2 Summary of parameters and observations in simulation analysis ($\sigma_h^2 = 0.1^2$).

σ_e^2	σ_e^2/σ_a^2	E			AE		
		k	Φ	λ	k	Φ	λ
0.05^2	0.09	468	727.3	0.643	468	1016.1	0.461
0.08^2	0.22	468	882.3	0.530	468	1285.8	0.364
0.11^2	0.42	468	1338.2	0.350	468	1595.7	0.293
0.14^2	0.68	468	1753.9	0.267	468	1899.7	0.246
0.17^2	1.00	468	2191.7	0.214	468	2111.3	0.222
0.20^2	1.38	468	2906.4	0.161	468	2386.7	0.196
0.23^2	1.83	468	3337.2	0.140	468	2567.0	0.182
0.26^2	2.34	468	3567.1	0.131	468	2811.7	0.166
0.29^2	2.91	468	4272.9	0.110	468	2990.2	0.157
0.32^2	3.54	467	4723.8	0.099	468	2925.0	0.160
0.35^2	4.24	467	5069.9	0.092	468	3162.1	0.148
0.38^2	5.00	466	5443.8	0.086	468	3288.5	0.142
0.41^2	5.82	462	5884.6	0.079	468	3418.1	0.137

multimodal cueing in 360° VR (Sheikh et al., 2016; Rothe and Hußmann, 2018; Bala et al., 2019; Malpica, Serrano, Gutierrez, et al., 2020a). We believe that our results also provide a rational explanation for these previous findings.

However, this study had some limitations. Some participants exhibited insufficient sensitivity to the electrostatic force stimuli. Although their hair moved when they were exposed to static electricity, they reported low sensations, which may be caused by skin moisture or other factors; however, this phenomenon has not yet been investigated. Furthermore, as humans are incapable of electroreception, it is reasonable to believe that the mechanoreceptors in the skin are involved in providing the sensations (Horch et al., 1977; Johnson, 2001; Zimmerman et al., 2014); however, this must be investigated further. In addition, the wristband used to tether the participants to the ground may have restricted free body movement; this can be addressed by introducing an ionizer that remotely neutralizes the charge level (Ohsawa, 2005), thereby allowing participants to move freely. Finally, the results presented in this study were obtained under reductive conditions. While the results provide insight into stimulus design, further experiments are required to demonstrate the effectiveness in real-world VR applications such as video viewing and gaming, which will be the focus of our future study.

In future work, we will implement electrostatic stimulation in a VR application. We believe that haptic stimulation by electrostatic force could be used not only to guide the visual field, but also to enhance the user's subjective impression. Although this has not been discussed here, we have experimentally implemented a VR game wherein a user shoots zombies charged with static electricity approaching from all sides. The electrostatic force-based stimulus can result in unpleasant sensations. Other haptic stimuli, such as vibrations, could also be used to cue the zombies. However, we believe that these stimuli are too obvious and artificial, and may detract from the subjective quality of experience to a certain extent. The use of static electricity can result in an unsettling experience for users when charged zombies approach them from behind. Thus, by comparing the effects of electrostatic force and other haptic stimuli on subjective impressions, we will be able to demonstrate the availability of electrostatic force-based stimulation to provide a highly immersive experience.

5 Conclusion

We investigated the multimodal effects of auditory and electrostatic force-based haptic cues on visual field guidance in 360° VR, demonstrating the potential for a visual field guidance method that does not interfere with the visual content. We found that modulating the degree of perceptual uncertainty for each cue improves the overall guidance performance under simultaneous multimodal cueing. Moreover, we presented a simple haptic stimulation method using only a single channel of a corona charging gun. In the future, we will increase the number of channels to present more complex stimulations in a larger area by dynamically controlling the electric fields, allowing for remote haptic stimulation under a six-degrees-of-freedom viewing condition. Finally, our results showed that multimodal stimuli have the potential to increase the richness in VR environments.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, upon reasonable request.

Ethics statement

The studies involving humans were approved by Japan Broadcasting Corporation. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Conflict of interest

Authors YS, MH, and KK were employed by Japan Broadcasting Corporation.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bailey, R., McNamara, A., Sudarsanam, N., and Grimm, C. (2009). Subtle gaze direction. *ACM Trans. Graph.* 28 (4), 1–14. doi:10.1145/1559755.1559757

Bala, P., Masu, R., Nisi, V., and Nunes, N. (2018). "Cue control: interactive sound spatialization for 360° videos," in *Interactive storytelling. ICIDS 2018. Lecture notes in computer science, vol 11318* Editors R. Rouse, H. Koenitz, and M. Haahr (Cham: Springer), 333–337. doi:10.1007/978-3-030-04028-4_36

Bala, P., Masu, R., Nisi, V., and Nunes, N. (2019). "When the elephant trumps," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, May 4-9, 2019, 1–13.

Chang, H.-Y., Tseng, W.-J., Tsai, C.-E., Chen, H.-Y., Peiris, R. L., and Chan, L. (2018). "FacePush: introducing normal force on face with head-mounted displays," in Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, 927–935.

Chao, F. Y., Ozcinar, C., Wang, C., Zerman, E., Zhang, L., Hamidouche, W., et al. (2020). "Audio-visual perception of omnidirectional video for virtual reality applications," in 2020 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2020, 2–7.

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., and Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLOS ONE* 13 (2), e0191846. doi:10.1371/journal.pone.0191846

Dalgarno, B., and Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *Br. J. Educ. Technol.* 41 (1), 10–32. doi:10.1111/j.1467-8535.2009.01038.x

Danieau, F., Guillo, A., and Doré, R. (2017). "Attention guidance for immersive video content in head-mounted displays," in 2017 IEEE Virtual Reality (VR), 205–206.

Ernst, M. O. (2006). "A bayesian view on multimodal cue integration," in *Human body perception from the inside out: advances in visual cognition* (Oxford University Press), 105–131.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7 (5), 7. doi:10.1167/7.5.7

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415 (6870), 429–433. doi:10.1038/415429a

Flach, J. M., and Holden, J. G. (1998). The reality of experience: gibson's way. *Presence* 7 (1), 90–95. doi:10.1162/105474698565550

Fukushima, S., and Kajimoto, H. (2012a). "Chilly chair: facilitating an emotional feeling with artificial piloerection," in *ACM SIGGRAPH 2012 emerging Technologies (SIGGRAPH '12), 1, article*, 5–1. doi:10.1145/2343456.2343461

Fukushima, S., and Kajimoto, H. (2012b). "Facilitating a surprised feeling by artificial control of piloerection on the forearm," in Proceedings of the 3rd Augmented Human International Conference (AH '12), Article 8, 1–4.

Gibson, J. J. (1979) *The ecological approach to visual perception 3.* Houghton Mifflin.

Gruenefeld, U., El Ali, A., Boll, S., and Heuten, W. (2018). "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 (MobileHCI '18), 1–11.

Gruenefeld, U., El Ali, A., Heuten, W., and Boll, S. (2017a). "Visualizing out-of-view objects in head-mounted augmented reality," in Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17), Article 87, 1–7.

Gruenefeld, U., Ennenga, D., Ali, A.El, Heuten, W., and Boll, S. (2017b). "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), 109–118.

Gruenefeld, U., Koethe, I., Lange, D., Weirb, S., and Heuten, W. (2019). "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), 742–746.

Harada, Y., and Ohyama, J. (2022). Quantitative evaluation of visual guidance effects for 360-degree directions. *Virtual Real.* 26 (2), 759–770. doi:10.1007/s10055-021-00574-7

Horch, K. W., Tuckett, R. P., and Burgess, P. R. (1977). A key to the classification of cutaneous mechanoreceptors. *J. Investigative Dermatology* 69 (1), 75–82. doi:10.1111/1523-1747.ep12497887

Hüttner, T., von Fersen, L., Miersch, L., and Dehnhardt, G. (2023). Passive electroreception in bottlenose dolphins (*Tursiops truncatus*): implication for micro- and large-scale orientation. *J. Exp. Biol.* 226 (22), jeb245845. doi:10.1242/jeb.245845

Johnson, K. (2001). The roles and functions of cutaneous mechanoreceptors. *Curr. Opin. Neurobiol.* 11 (4), 455–461. doi:10.1016/S0959-4388(00)00234-8

Karasawa, M., and Kajimoto, H. (2021). "Presentation of a feeling of presence using an electrostatic field: presence-like sensation presentation using an electrostatic field," in Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21), Article 285, 1–4.

Lin, Y.-C., Chang, Y.-J., Hu, H.-N., Cheng, H.-T., Huang, C.-W., and Sun, M. (2017). "Tell me where to look," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2535–2545.

Lin, Y. T., Liao, Y. C., Teng, S. Y., Chung, Y. J., Chan, L., and Chen, B. Y. (2017). "Outside-in: visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17), 255–265.

Malpica, S., Serrano, A., Allue, M., Bedia, M. G., and Masia, B. (2020a). Crossmodal perception in virtual reality. *Multimedia Tools Appl.* 79 (5–6), 3311–3331. doi:10.1007/s11042-019-7331-z

Malpica, S., Serrano, A., Gutierrez, D., and Masia, B. (2020b). Auditory stimuli degrade visual performance in virtual reality. *Sci. Rep.* 10 (1), 12363–12369. doi:10.1038/s41598-020-69135-3

Martin, D., Malpica, S., Gutierrez, D., Masia, B., and Serrano, A. (2022). Multimodality in VR: a survey. *ACM Comput. Surv.* 54 (10s), 1–36. doi:10.1145/3508361

Masia, B., Camon, J., Gutierrez, D., and Serrano, A. (2021). Influence of directional sound cues on users' exploration across 360° movie cuts. *IEEE Comput. Graph. Appl.* 41 (4), 64–75. doi:10.1109/MCG.2021.3064688

Matsuda, A., Nozawa, K., Takata, K., Izumihara, A., and Rekimoto, J. (2020). "HapticPointer," in Proceedings of the Augmented Humans International Conference (AHs '20), 1–10.

McElree, B., and Carrasco, M. (1999). The temporal dynamics of visual search: evidence for parallel processing in feature and conjunction searches. *J. Exp. Psychol. Hum. Percept. Perform.* 25 (6), 1517–1539. doi:10.1037/0096-1523.25.6.1517

Melo, M., Goncalves, G., Monteiro, P., Coelho, H., Vasconcelos-Raposo, J., and Bessa, M. (2022). Do multisensory stimuli benefit the virtual reality experience? A systematic review. *IEEE Trans. Vis. Comput. Graph.* 28 (2), 1428–1442. doi:10.1109/TVCG.2020.3010088

Mikropoulos, T. A., and Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). *Comput. Educ.* 56 (3), 769–780. doi:10.1016/j.compedu.2010.10.020

Murray, N., Lee, B., Qiao, Y., and Muntean, G. M. (2016). Olfaction-enhanced multimedia: a survey of application domains, displays, and research challenges. *ACM Comput. Surv.* 48 (4), 1–34. doi:10.1145/2816454

Newton, K. C., Gill, A. B., and Kajiura, S. M. (2019). Electroreception in marine fishes: chondrichthyans. *J. Fish Biol.* 95 (1), 135–154. doi:10.1111/jfb.14068

Nielsen, L. T., Møller, M. B., Hartmeyer, S. D., Ljung, T. C. M., Nilsson, N. C., Nordahl, R., et al. (2016). "Missing the point," in Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, 229–232.

Ohsawa, A. (2005). Modeling of charge neutralization by ionizer. *J. Electrost.* 63 (6–10), 767–773. doi:10.1016/j.elstat.2005.03.043

Pavel, A., Hartmann, B., and Agrawala, M. (2017). "Shot orientation controls for interactive cinematography with 360° video," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17), 289–297.

Proske, U., Gregory, J. E., and Iggo, A. (1998). Sensory receptors in monotremes. *Philosophical Trans. R. Soc. B Biol. Sci.* 353 (1372), 1187–1198. doi:10.1098/rstb.1998.0275

Ranasinghe, N., Jain, P., Karwita, S., Tolley, D., and Do, E. Y.-L. (2017). "Ambiotherm," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 1731–1742.

Ranasinghe, N., Jain, P., Thi Ngoc Tram, N., Koh, K. C. R., Tolley, D., Karwita, S., et al. (2018). "Season traveller," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–13.

Rothe, S., Althammer, F., and Khamis, M. (2018). "GazeRecall: using gaze direction to increase recall of details in cinematic virtual reality," in Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, 115–119.

Rothe, S., Buschek, D., and Hußmann, H. (2019). Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technol. Interact.* 3 (1), 19. doi:10.3390/mti3010019

Rothe, S., and Hußmann, H. (2018). "Guiding the viewer in cinematic virtual reality by diegetic cues," in *Augmented reality, virtual reality, and computer graphics. AVR 2018. Lecture notes in computer science* Editors L. De Paolis, and P. Bourdot (Cham: Springer), 101–117. doi:10.1007/978-3-319-95270-3_7

Rothe, S., Hußmann, H., and Allary, M. (2017). "Diegetic cues for guiding the viewer in cinematic virtual reality," in Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, 1–2.

Schmitz, A., Macquarrie, A., Julier, S., Binetti, N., and Steed, A. (2020). "Directing versus attracting attention: exploring the effectiveness of central and peripheral cues in panoramic videos," in Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2020, 63–72.

Sheikh, A., Brown, A., Watson, Z., and Evans, M. (2016). Directing attention in 360-degree video. *IBC 2016 Conf.*, 1–9. doi:10.1049/ibc.2016.0029

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Trans. R. Soc. B Biol. Sci.* 364 (1535), 3549–3557. doi:10.1098/rstb.2009.0138

Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., and Oliva, R. (2022). A separate reality: an update on place illusion and plausibility in virtual reality. *Front. Virtual Real.* 3. doi:10.3389/frvir.2022.914392

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73 (4), 971–995. doi:10.3758/s13414-010-0073-7

Suzuki, K., Abe, K., and Sato, H. (2020). Proposal of perception method of existence of objects in 3D space using quasi-electrostatic field. *Int. Conf. Human-Computer Interact.*, 561–571. doi:10.1007/978-3-030-49760-6_40

Tanaka, Y., Nishida, J., and Lopes, P. (2022). Electrical head actuation: enabling interactive systems to directly manipulate head orientation. *Proc. 2022 CHI Conf. Hum. Factors Comput. Syst.* 1, 1–15. doi:10.1145/3491102.3501910

Tong, L., Jung, S., and Lindeman, R. W. (2019). "Action units: directing user attention in 360-degree video based VR," in 25th ACM Symposium on Virtual Reality Software and Technology, 1–2.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12 (1), 97–136. doi:10.1016/0010-0285(80)90005-5

Walker, B. N., and Lindsay, J. (2003). "Effect of beacon sounds on navigation performance in a virtual reality environment," in Proceedings of the 9th International Conference on Auditory Display (ICAD2003), July, 204–207.

Wallgrun, J. O., Bagher, M. M., Sajjadi, P., and Klippel, A. (2020). "A comparison of visual attention guiding approaches for 360° image-based VR tours," in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 83–91.

Wang, D., Li, T., Zhang, Y., and Hou, J. (2016). Survey on multisensory feedback virtual reality dental training systems. *Eur. J. Dent. Educ.* 20 (4), 248–260. doi:10.1111/eje.12173

Yamaguchi, S., Ogawa, N., and Narumi, T. (2021). "Now I'm not afraid: reducing fear of missing out in 360° videos on a head-mounted display using a panoramic thumbnail," in 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 176–183.

Zimmerman, A., Bai, L., and Ginty, D. D. (2014). The gentle touch receptors of mammalian skin. *Science* 346 (6212), 950–954. doi:10.1126/science.1254229