Annals of the New York Academy of Sciences

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

ORIGINAL ARTICLE

Value-Driven Crossmodal Spatial Attention: The Role of Perceptual Load

Qiongyu Wang[1] | Lihan Chen[1,2,3,4]

[1]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China | [2]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China | [3]National Engineering Laboratory for Big Data Analysis and Applications, Peking University, Beijing, China | [4]Beijing Key Laboratory of Brain-Computer Interface and Mental Health Modulation, Peking University, Beijing, China

Correspondence: Lihan Chen (CLH@pku.edu.cn)

Received: 26 February 2026 | **Revised:** 6 June 2026 | **Accepted:** 9 June 2026

Keywords: attention | auditory localization | multisensory processing | perceptual load | value | ventriloquism effect

ABSTRACT

Reward-associated stimuli capture attention in a value-driven manner, interfering with multisensory integration. This study investigated the mechanisms of value-driven crossmodal attention and the moderating role of perceptual load using the spatial ventriloquism effect. Across two experiments, participants localized target sounds—defined by frequency (low load) or frequency and duration (high load)—while ignoring reward-linked visual distractors. Experiment 1 employed a binaural localization task with peripheral distractors, whereas Experiment 2 used a continuous localization task with distractors positioned near auditory targets. High-value distractors successfully captured crossmodal attention, producing slower reaction times and stronger spatial bias toward visual locations. Crucially, perceptual load effects were modulated by spatial configuration: increasing load reduced visual interference when distractors were peripheral (Experiment 1) but intensified interference when distractors were colocated with targets (Experiment 2). Notably, load did not affect localization accuracy. These findings demonstrate that value-driven crossmodal processing is jointly shaped by spatial relationships between modalities and cognitive demands, offering insights into how specific perceptual load conditions constrain selective attention during multisensory integration. Given the sample size limitations of this study, these results should be interpreted cautiously pending future replication.

1 | Introduction

In our daily lives, we are constantly bombarded by a vast amount of sensory information. Because the human brain has a finite capacity for information processing, it is neither possible nor necessary to process every stimulus. Selective attention, therefore, serves as a critical mechanism for prioritizing important information and allocating limited cognitive resources to enhance the efficiency of our interactions with the environment. Understanding the factors that determine attentional priority and resource allocation remains a central objective in cognitive psychology [1].

Traditional theories of selective attention have long distinguished between two primary mechanisms: goal-directed and stimulus-driven attention [2, 3]. Goal-directed (top-down) attention is a voluntary process where individuals actively select information based on current intentions, such as a driver looking for a specific traffic sign. In contrast, stimulus-driven (bottom-up) attention is an involuntary capture by physically salient or novel stimuli, such as a sudden flash of light. While goal-directed selection allows for task-relevant processing, it is cognitively demanding and slow; stimulus-driven selection is automatic and rapid, helping us detect unexpected threats but also making us susceptible to distraction [4, 5].

Despite the widespread acceptance of this dichotomy, emerging evidence suggests it is insufficient. Researchers have identified a third category: selection history [6]. This framework incorporates attentional biases that arise from previous experiences, such as intertrial priming, where searching for a specific feature becomes more efficient over repeated trials [7, 8], and statistical learning, where individuals implicitly learn the probability of target or distractor locations [9, 10].

Within the realm of selection history, the role of reward has gained significant prominence. Anderson proposed that attentional priority is fundamentally determined by the value of a stimulus [11]. Through reward learning, stimuli that are neither task-relevant nor physically salient can acquire the power to capture attention if they have been previously associated with high rewards. This phenomenon, termed *value-driven attentional capture* (VDAC), represents a mechanism independent of traditional top-down or bottom-up processes [4].

While VDAC was initially studied within the visual modality, it is now recognized as a crossmodal phenomenon. Our environment is multisensory, and reward associations in one modality can influence attention in another. For instance, high-value auditory stimuli can significantly interfere with performance on a visual search task [12]. Furthermore, value-driven capture can alter multisensory integration. In phenomena such as the spatial ventriloquism effect—where visual location captures the perceived location of a sound—reward associations can modulate the degree of bias, either by strengthening the integration or by drawing attention away from the target modality [13, 14].

A critical factor that governs how we allocate resources among competing stimuli is perceptual load. According to Lavie's perceptual load theory, the level of selection (early vs. late) depends on the demands of the primary task. Under high perceptual load, the task consumes all available resources, leaving none for task-irrelevant distractors, thereby facilitating early selection. Under low perceptual load, spare resources "spill over," resulting in the automatic processing of distractors, leading to late selection and increased interference [15]. However, the efficacy of perceptual load is modulated by various factors. Studies show that load-dependent filtering is most effective for peripheral distractors but often fails when distractors are centrally located [16, 17]. Additionally, stimuli with high biological or social significance, such as faces, often bypass the load-induced filter [18]. Crucially, the relationship between perceptual load and VDAC remains complex. Some evidence suggests that high-value stimuli are more resistant to load effects [19], and in heterogeneous contexts, increasing load might even enhance the capture of reward-associated distractors [20].

Finally, the question of whether attentional resources are shared across sensory modalities or are modality-specific remains a subject of intense debate. While some crossmodal studies report phenomena like inattentional deafness under high visual load [21], others suggest that processing in one modality (e.g., audition) is largely unaffected by the load in another [22, 23].

Addressing these gaps in the literature, this study aims to investigate how value-driven crossmodal capture is modulated by perceptual load within the context of audiovisual spatial localization. The first objective of this research is to determine whether value-driven crossmodal attention capture and subsequent multisensory integration can occur independently of operant conditioning. While traditional training-test paradigms [12, 24, 25] demonstrate that reward-linked stimuli in one modality can interfere with another, these effects may stem from reinforced response habits rather than the intrinsic value of the stimulus [26]. Here, reward values are associated with task-irrelevant visual distractors that never serve as targets. In this context, the presence of value-driven capture would provide strong evidence for a robust, automatic selection mechanism that enhances multisensory integration (e.g., the spatial ventriloquism effect) based purely on reward history [14, 27].

The second objective focuses on how perceptual load and the spatial configuration of stimuli interact to modulate this capture. According to perceptual load theory, high task demands should exhaust attentional resources and suppress distractor processing [28]. However, this filtering effect may be compromised by the high attentional priority of reward-associated stimuli [19] or by the spatial proximity of the distractor. By comparing binaural (peripheral; Experiment 1) and continuous (proximal; Experiment 2) localization tasks, we evaluate how "attentional zoom" dictates the impact of perceptual load on crossmodal processing. We test whether spatial proximity leads to object-level integration, potentially reversing the expected benefits of high auditory load on visual interference [29]. These results will distinguish between shared and modality-specific attentional resources while illustrating how the brain balances task-directed demands with value-based priorities in complex environments.

2 | Experiment 1: Binaural Spatial Localization Task

2.1 | Methods

In Experiment 1, we used a binaural spatial ventriloquism task integrated with a response-competition paradigm, a standard approach in selective attention research. Auditory targets were presented at ±5° azimuth, while visual distractors appeared at ±10° azimuth, both positioned to the left and right of the central fixation (0°).

The spatial relationship between the audiovisual stimuli was manipulated to be either congruent (same side) or incongruent (opposite sides), with auditory-only trials serving as a neutral baseline. Participants were tasked with ignoring the visual distractors to identify and localize the auditory targets. By associating visual colors with varying reward magnitudes and manipulating the auditory discrimination requirements, this experiment investigated how task-irrelevant value associations and task-relevant perceptual load interact to influence auditory localization speed and accuracy.

2.1.1 | Participants

Thirty-one university students (15 males, age range 18–28 years, $M = 22.8$) participated in the study. A power analysis using G*Power 3.1 indicated the sample was sufficiently powered, achieving a power of 0.94 to detect a medium effect size (f = 0.25). All participants were right-handed, reported normal or corrected-to-normal vision and hearing, and had no history of color blindness or deficiency. Two participants were excluded from the final analysis because their error rates in at least one experimental condition exceeded three standard deviations from the group mean. Consequently, the final sample consisted of 29 participants. The study protocol was approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences of Peking University.

2.1.2 | Stimuli and Apparatus

The experiment was conducted in a sound-attenuated darkroom. The apparatus consisted of a semi-circular rack with a 1.6-meter radius, housing an array of LED lights and speakers (see Figure S1 for the Experiment 1 setup). To prevent visual cues regarding the hardware, the entire rack was concealed behind a black cloth. Visual stimuli consisted of 50 or 200 ms flickers produced by red or green LEDs positioned at ±10° from the center. A yellow LED at 0° served as the fixation point, while two vertical columns of eight LEDs (one red, one green) provided trial-by-trial score feedback. Auditory stimuli were pure tones (750 or 3000 Hz) with durations of 50 or 200 ms, including 5 ms rise/fall times. Tones were delivered via speakers positioned at ±5°.

2.1.3 | Design and Procedure

The experiment adopted a 2 × 2 × 2 within-subjects design, manipulating auditory perceptual load (low vs. high), audiovisual congruency (congruent vs. incongruent), and visual distractor value (low vs. high). Each trial began with a central yellow fixation point (500–800 ms), followed by the presentation of the stimuli. For half of the participants, red distractors were associated with high value and green with low value; this mapping was reversed for the remaining half. Upon hearing a target sound, participants localized it (go) by judging whether the sound originated from the left or right side and pressing the corresponding left or right key. If the sound was a nontarget, they withheld localization (no-go) and pressed the spacebar. Participants were instructed to respond as quickly and accurately as possible. A feedback screen was presented after each trial (Figure 1).

Participants were instructed to ignore the visual distractors and focus solely on the auditory task, though they were informed that the distractor's color determined the trial's reward—an eightfold difference in potential earnings—which would be converted into their final compensation. Points were converted to monetary compensation at a rate of 1 CNY per 40 points. On average, participants earned approximately 2500 points (range: 2000–3000 points), which resulted in a final payment ranging from 50 to 80 CNY (M = 62.5 CNY).

Perceptual load was manipulated via a go/no-go task. In the low-load condition, participants discriminated only the tone frequency; for example, 750 Hz tones required a localization response (go, 75% of trials), while 3000 Hz tones required a spacebar press (no-go, 25%). In the high-load condition, participants were required to identify specific combinations of frequency and duration (e.g., 750 Hz/50 ms and 3000 Hz/200 ms as go targets), significantly increasing processing demands. Following the value-driven capture paradigm [26], correct responses within 1000 ms earned points matching the distractor's value (1 or 8 points), indicated by the number and color of lit LEDs in the feedback display. Errors or slow responses yielded zero points. The experiment consisted of 768 trials divided into four blocks of 192 trials, with perceptual load counterbalanced in an ABBA/BAAB sequence.

2.1.4 | Data Analysis

Data were preprocessed and categorized using MATLAB R2018a (Mathworks, Inc). Initially, sound discrimination error rates in the go/no-go task were analyzed across all trials to verify the efficacy of the load manipulation. For trials requiring a localization response (Go trials), outliers—defined as reaction times (RTs) exceeding three standard deviations from an individual's mean—were excluded (accounting for 1.08% of Go trials). RTs and error rates for target localization were then calculated for each condition. The audiovisual congruency effect, used as a measure of visual interference, was derived by subtracting the performance in congruent trials from that in incongruent trials. Statistical analyses were conducted in jamovi 1.2.17 (The jamovi project, 2020) using repeated-measures analysis of variance (ANOVA).

2.1.5 | Results

The analysis of sound discrimination (go/no-go task) error rates confirmed a robust manipulation of perceptual load: the error rate under low auditory load (3.31%) was significantly lower than under high load (9.53%), $F(1, 28) = 79.45$, $p < 0.001$, $\eta_p^2 = 0.739$, 95% CI for the mean difference [−7.64, −4.79] (Figure 2). For the target localization RTs, a significant interaction between auditory load and congruency revealed that the visual interference was more pronounced under low load than under high load ($F(2, 56) = 4.33$, $p = 0.018$, $\eta_p^2 = 0.134$). The congruency effect (Incongruent—Congruent) was 59.16 ms under low load [95% CI: 42.13, 76.19] and 34.55 ms under high load [95% CI: 20.04, 49.07]; the interaction contrast was 24.60 ms [95% CI: 5.19, 44.02] (Figure 3).

Most critically, the audiovisual congruency effect analysis demonstrated the impact of reward value. In RTs, high-value visual distractors produced a significantly larger congruency effect (62 ms) than low-value distractors (32 ms), ($F(1, 28) = 8.95$, $p = 0.006$, $\eta_p^2 = 0.242$), 95% CI [9.47, 50.53] ms, indicating value-driven attention capture. While increasing perceptual load reduced the RT congruency effect ($F(1, 28) = 6.74$, $p = 0.015$, $\eta_p^2 = 0.194$), 95% CI [5.19, 44.02] ms, it did not modulate the impact of value ($p = 0.173$) (Figures 4 and 5). Similarly, in error rates—a measure of the spatial ventriloquism effect—high-value distractors caused greater interference (18.6%) than low-value ones (15.9%; $F(1, 28) = 5.21$, $p = 0.030$, $\eta_p^2 = 0.157$), 95% CI [0.28, 5.12]%), but this accuracy-based interference was unaffected by perceptual load ($p = 0.098$) (Figure 6).

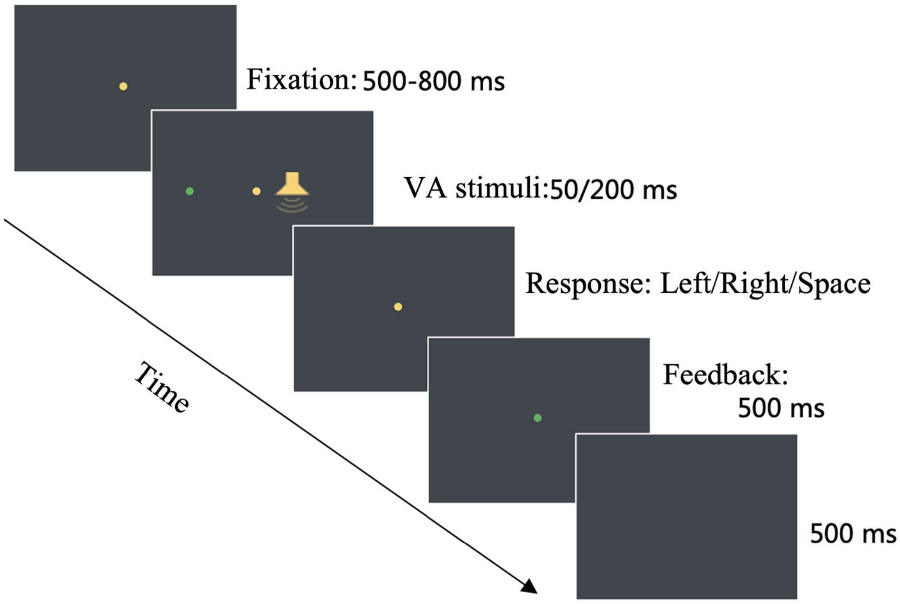

FIGURE 1 | Stimuli and procedure for Experiment 1 (see text for details).

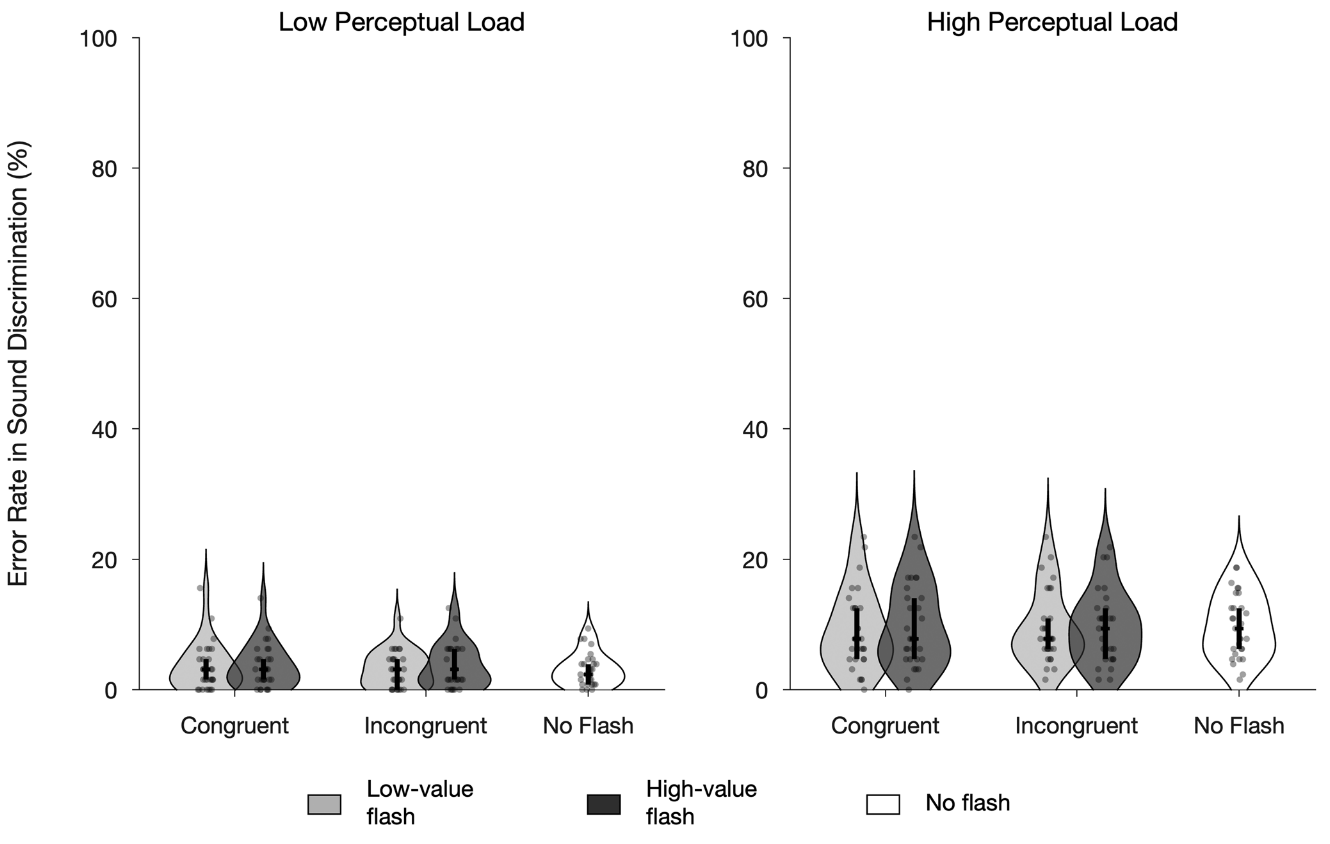

FIGURE 2 | Error rates in sound discrimination (go/no-go task) as a function of audiovisual congruency (congruent, incongruent, no flash) and visual distractor value (low-value vs. high-value flash), separately for low perceptual load (left panel) and high perceptual load (right panel) for Experiment 1. Violin plots show the kernel-density distribution of individual participant means. Thick vertical bars indicate the interquartile range (IQR), and horizontal bars mark the median. Individual data points are overlaid as jittered dots. No-flash trials serve as the baseline condition without visual distractors.

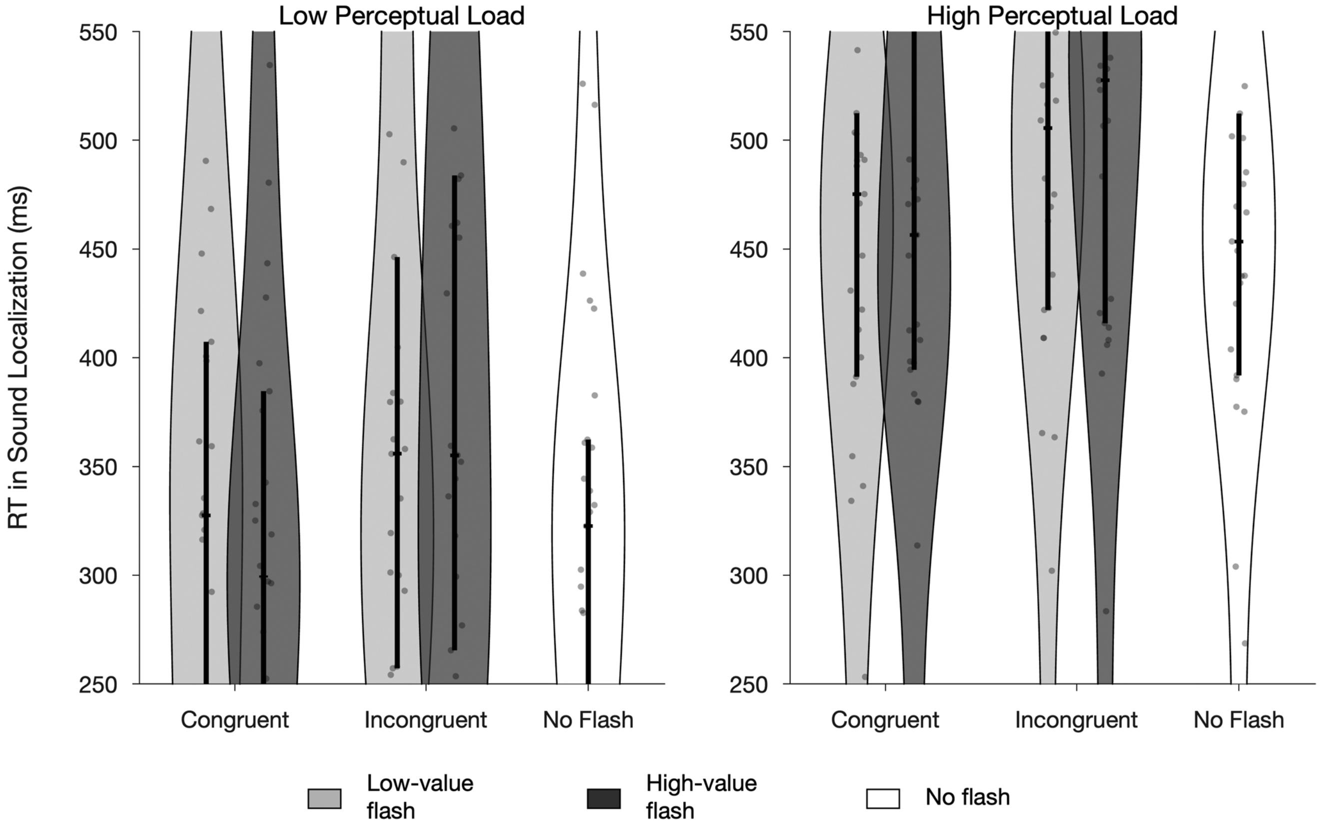

FIGURE 3 | Reaction times (ms) in sound localization as a function of audiovisual congruency (congruent, incongruent, no flash) and visual distractor value (low-value vs. high-value flash), separately for low perceptual load (left panel) and high perceptual load (right panel) for Experiment 1. Violin plots show the kernel-density distribution of individual participant mean RTs. Thick vertical bars indicate the IQR, and horizontal bars mark the median. Individual data points are overlaid as jittered dots.

To further track the behavioral differences arising from reward learning throughout the experimental process, the two experimental blocks for each of the high and low auditory perceptual load conditions were analyzed separately, and a 2 (experimental block) × 2 (auditory perceptual load) × 2 (visual distractor value) three-way ANOVA was conducted. The results showed that for RTs (Figure S2), there was no main effect of experimental block ($F(1, 28) = 0.51$, $p = 0.482$, $\eta_p^2 = 0.018$), 95% CI [−23.05, 11.13] or interaction effect ($p > 0.2$). For error rates (Figure S3), the main effect of experimental block was significant, $F(1, 28) = 12.37$, $p = 0.002$, $\eta_p^2 = 0.306$, with the congruency effect decreasing by 10.99 percentage points from block 1 to block 2 (95% CI [−17.39, −4.59]), reflecting a decrease in the congruency effect on error rates as the experiment progressed; however, none of the interaction effects were significant ($p > 0.3$).

2.2 | Discussion

Through the manipulation of auditory discrimination requirements in the go/no-go task, we successfully operationalized auditory perceptual load. Sound discrimination error rates were significantly higher in the high-load condition (discriminating frequency and duration) compared to the low-load condition (discriminating frequency only). Notably, the presence of visual distractors did not affect the accuracy of sound discrimination itself, suggesting that participants strictly adhered to task instructions by focusing on the auditory targets while attempting to ignore visual interference. Despite this focus, visual distractors strongly influenced auditory localization performance, underscoring the inherent dominance of the visual modality in spatial (localization) tasks.

By associating specific colors with varying reward magnitudes, we observed VDAC in the absence of prior response training or reinforcement. High-value visual stimuli induced significantly greater interference in auditory localization than low-value stimuli across both RTs and error rates. The error rates, specifically indexing the audiovisual congruency effect, demonstrate that high-value visual distractors elicited a stronger spatial ventriloquism effect. This indicates that reward-associated stimuli in a task-irrelevant modality not only impair the response efficiency of the target modality but also modulate the integration of multisensory information, ultimately altering perceptual outcomes. This confirms that VDAC operates across sensory modalities and modulates multisensory integration.

While increasing auditory perceptual load did not influence the magnitude of this crossmodal integration effect (accuracy), it did reduce the interference of value-driven capture on RTs, a finding consistent with perceptual load theory. Further analysis across the experimental timeline revealed that the congruency effect on RTs remained stable throughout both halves of the session, indicating that the moderating role of perceptual load on VDAC

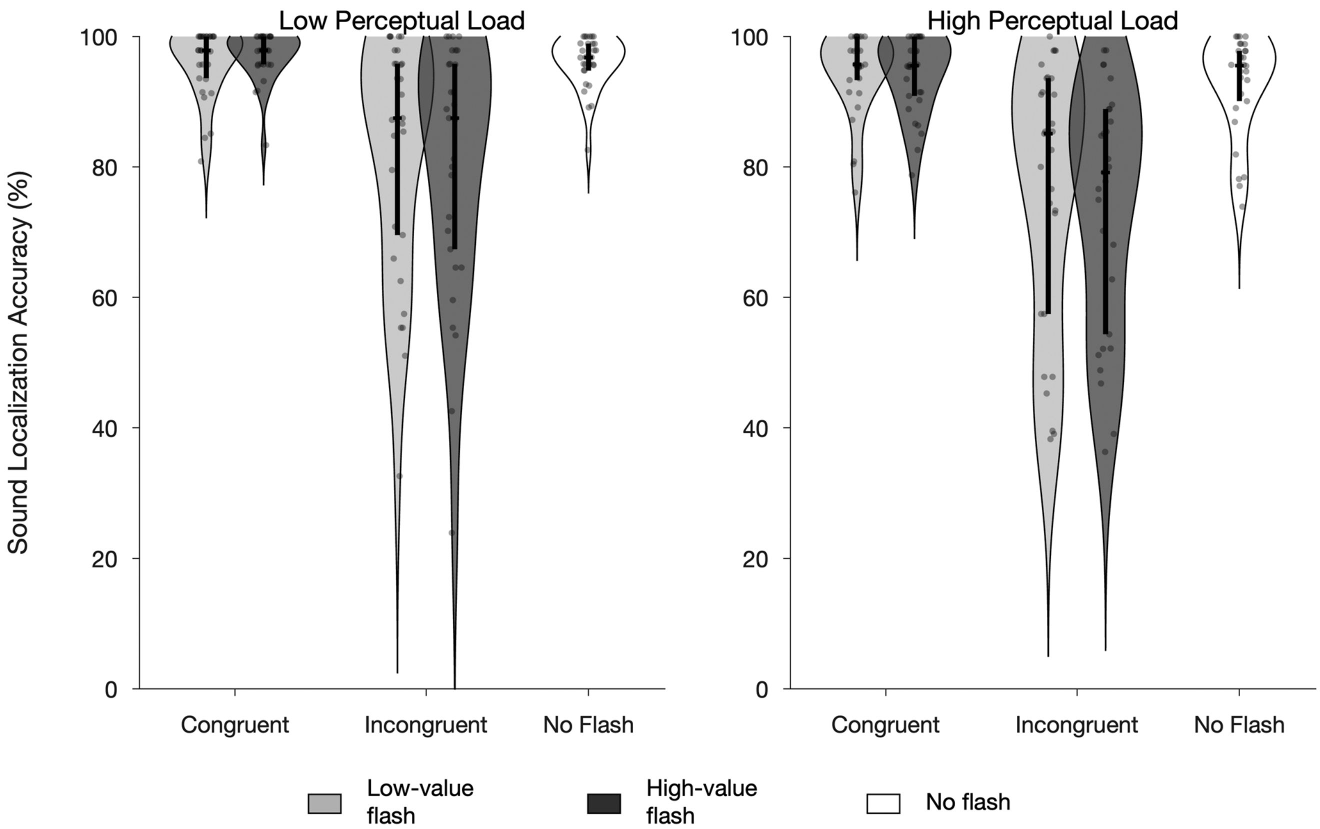

FIGURE 4 | Sound localization accuracy (%) as a function of audiovisual congruency (congruent, incongruent, no flash) and visual distractor value (low-value vs. high-value flash), separately for low perceptual load (left panel) and high perceptual load (right panel) for Experiment 1. Violin plots show the kernel-density distribution of individual participant mean accuracy. Thick vertical bars indicate the IQR, and horizontal bars mark the median. Individual data points are overlaid as jittered dots. Higher values indicate better localization performance.

is relatively consistent. Interestingly, the congruency effect did not vanish entirely under high-load conditions but remained at a substantial level, reflecting the competitive advantage of high-value stimuli in capturing attentional resources.

The dissociation between accuracy and RT effects under load manipulation suggests that crossmodal spatial integration and VDAC are governed by distinct neurocognitive principles. The ventriloquism effect reflects a mandatory, automatic remapping of auditory space onto visual coordinates [30–32] that appears to proceed with minimal attentional capacity. In contrast, the RT congruency effect reflects the competition between task-relevant processing and value-driven distractor capture for limited central resources [15, 24]. Under high perceptual load, the visual system exhausts its perceptual capacity on the primary task, leaving fewer residual resources for the value-driven priority signal to influence response selection, thereby reducing RT interference without abolishing the underlying multisensory spatial bias.

Our findings differ slightly from those of Gupta et al., who found no significant difference in interference from reward-linked faces across load levels [19]. This discrepancy may stem from the spatial positioning of the distractors. In their study, reward-associated faces were presented centrally within the search array, making them difficult to ignore regardless of load. In our Experiment 1, however, auditory targets were central, while visual distractors were peripheral. This configuration likely allowed participants to focus their attentional zoom more narrowly on the central field under high load, thereby effectively mitigating peripheral interference [17].

To further explore this, in Experiment 2, we adjusted the spatial relationship of the stimuli by placing visual distractors within the attentional range of the auditory task. Furthermore, while the flanker-like style task used in Experiment 1 is a standard laboratory paradigm, it lacks ecological validity for real-world audiovisual scenarios where sound sources and visual distractors are not restricted to central and peripheral fixations. Therefore, Experiment 2 employed a continuous auditory localization task [13] to investigate these mechanisms within a more refined and ecologically valid spatial ventriloquism framework.

3 | Experiment 2: Continuous Auditory Localization Task

3.1 | Methods

Experiment 2 was designed to build upon the findings of the first experiment by refining the spatial parameters to more closely resemble real-world auditory localization scenarios. While Experiment 1 utilized a discrete dichotic task, Experiment 2 presented auditory stimuli across a broader field (±36°) and implemented a fine-grained continuous localization task (see Figure S4 for

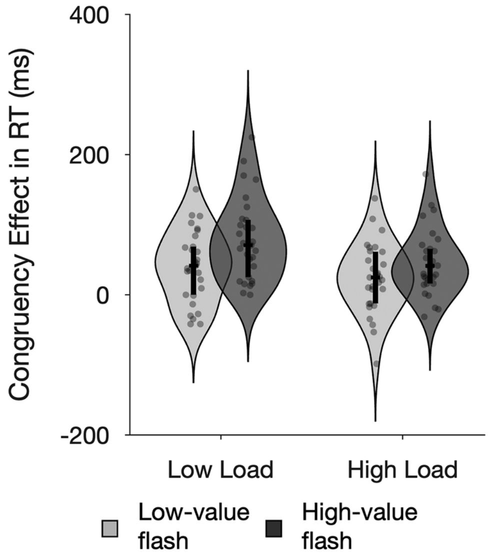

FIGURE 5 | Congruency effect in sound localization reaction time (RT), calculated as the difference between incongruent and congruent trials (Incongruent—Congruent), plotted separately for low perceptual load (left cluster) and high perceptual load (right cluster) for Experiment 1. Within each load level, effects are shown for low-value flash (light gray) and high-value flash (dark gray) conditions. Positive values indicate slower responses on incongruent relative to congruent trials. Violin plots show the kernel-density distribution of individual participant congruency effects.

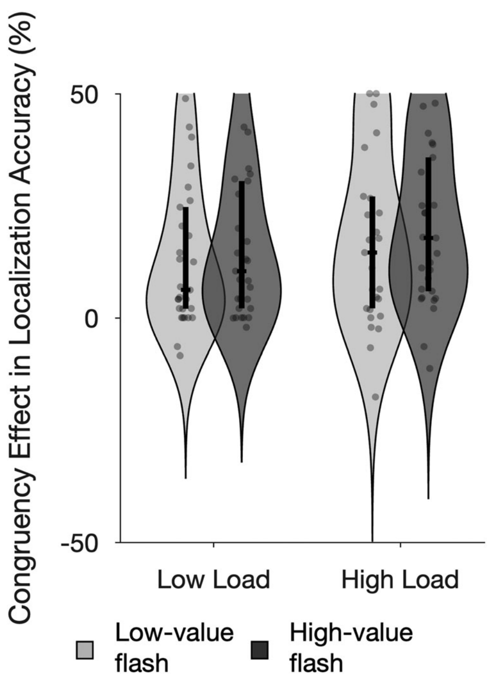

FIGURE 6 | Congruency effect in sound localization accuracy, calculated as the difference between congruent and incongruent trials (Congruent—Incongruent), plotted separately for low perceptual load (left cluster) and high perceptual load (right cluster) for Experiment 1. Within each load level, effects are shown for low-value flash (light gray) and high-value flash (dark gray) conditions. Positive values indicate better performance (higher accuracy) on congruent relative to incongruent trials. Violin plots show the kernel-density distribution of individual participant congruency effects.

setup). By placing visual distractors in close proximity (±9°) to the auditory targets, this experiment sought to measure the visual bias (spatial ventriloquism effect) with higher precision. This setup allowed us to investigate whether the moderating effect of perceptual load changes when the distractor is positioned within the immediate attentional range of the target task.

3.1.1 | Participants and Apparatus

Data collection for Experiment 2 was curtailed due to unforeseen institutional lockdowns and social distancing mandates resulting from the COVID-19 pandemic. Sixteen university students (5 males, age range 18–25 years, M = 21.9) participated in this experiment. Consequently, we relied on post-hoc power analyses to evaluate the sensitivity of our statistical tests. We acknowledge that post-hoc power calculations cannot substitute for rigorous a priori sample planning, and this reduced sample size inherently restricts our statistical power, particularly for detecting complex interaction effects. Despite this limitation, a post-hoc power analysis indicated that this sample size still provided a power of 0.91 to detect large effect sizes (f = 0.40) within our repeated-measures design, ensuring sufficient sensitivity for the primary experimental manipulations. All were right-handed with normal or corrected-to-normal vision and hearing, and none reported color vision deficiencies. The experimental apparatus remained consistent with the semi-circular rack used in Experiment 1, but the speaker configuration was expanded to seven locations (0°, ±12°, ±24°, and ±36°). For each speaker, red and green LEDs were positioned 9° to the left and right to serve as visual distractors. Participants provided responses using a specialized device consisting of five keys and a rotatable pointer equipped with a laser, which projected a precise point onto the concealing black cloth to facilitate accurate spatial localization. The study protocol was approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences of Peking University.

3.1.2 | Design and Procedure

We implemented a 2 × 2 within-subjects design, manipulating auditory perceptual load (low vs. high) and visual distractor value (low vs. high). In each trial, auditory and visual stimuli were presented simultaneously. For go trials (target identification), participants were required to use their right hand to rotate the laser pointer to the perceived origin of the sound and press a confirmation key with their left hand. For no-go trials (non-target identification), participants pressed another key without performing a localization. Unlike Experiment 1, the emphasis was placed entirely on spatial accuracy rather than response speed. After each response, participants were required to return the pointer to a neutral "home zone" within the central −5° to 5° range.

In the absence of a unimodal auditory baseline, we adapted our protocol based on the paradigms of Wang et al. and Wang and Theeuwes. While Wang et al. demonstrate that the ability of a stimulus to capture attention is gated by the interaction between perceptual salience and value [33], Wang and Theeuwes highlight how spatial statistical regularities, even when implicit, induce location-based suppression that influences attentional capture and target selection efficiency [10]. Our study utilizes these principles to investigate how value-driven crossmodal attention is modulated by perceptual load and spatial configuration. By examining the differential capture effect under varying distractor locations (i.e., disparities of audiovisual locations), we emphasize the impact of reward associations on multisensory integration while minimizing participant fatigue and eliminate the potential for overlearning effects.

The reward structure was adjusted to a fourfold difference between high- and low-value colors to maintain motivation. Points were awarded based on localization precision: errors within 4° earned maximum points (2 or 8), while errors between 4° and 8° earned partial points (1 or 4). Localization errors exceeding 8° or incorrect discrimination resulted in zero points. Perceptual load was managed in a block design across four sessions of 224 trials each, counterbalanced in an ABBA/BAAB sequence. Each block balanced the seven auditory locations and the spatial positioning of the visual distractors (left vs. right). A total of 24 practice trials were provided before each load condition to ensure task proficiency over the approximately 90-min experimental session.

3.1.3 | Data Analysis

Data preprocessing and categorization were conducted using MATLAB R2018a (Mathworks, Inc.), while statistical analyses were conducted using repeated-measures ANOVA in jamovi 1.2.17. Initially, sound discrimination error rates in the go/no-go task were examined to verify task engagement and load efficacy. For trials requiring target localization, outliers—defined as trials where either RT or absolute localization error exceeded three standard deviations from an individual's mean—were excluded (representing 3.63% of Go trials). The primary dependent variable of interest was the visual bias in auditory localization, a measure of the spatial ventriloquism effect. This was calculated by determining the difference between mean localization angles when the visual distractor was to the right of the sound versus the left, averaged across all seven auditory positions. This metric quantifies the degree to which visual stimuli capture auditory perception, ranging from zero (no influence) to twice the spatial disparity between the two stimuli (complete visual dependence).

3.1.4 | Results

The analysis of sound discrimination error rates and localization RTs confirmed that the manipulation of perceptual load was successful. Participants exhibited significantly higher error rates under high auditory load (M = 4.98%) than under low load (M = 1.77%), with a mean difference of 3.21 percentage points [95% CI: 2.06, 4.35], $p < 0.001$ (Figure 7). Similarly, localization RTs were slower under high load (M = 1434.02 ms) compared to low load (M = 1229.42 ms), with a mean difference of 204.60 ms [95% CI: 94.38, 314.84], $p < 0.001$ (Figure 8).

A significant interaction between auditory load and distractor value was observed for RTs, $F(1, 15) = 6.98$, $p = 0.018$, $\eta_p^2 = 0.318$, 95% CI for the interaction contrast [7.05, 65.92] ms. Follow-up paired-samples t-tests revealed that under high load, high-value visual distractors resulted in significantly longer RTs (M = 1450.37 ms) than low-value distractors (M = 1417.67 ms), $t(15) = 2.18$, $p = 0.046$, Cohen's $d = 0.54$, 95% CI [0.69, 64.70] ms. By contrast, under low load, high-value distractors (M = 1227.53 ms) did not differ significantly from low-value distractors (M = 1231.32 ms), $t(15) = -0.27$, $p = 0.792$, $d = -0.07$, 95% CI [−33.91, 26.33] ms. While RTs decreased significantly from block 1 to block 2 (mean decrease = 262.89 ms), $t(15) = 4.17$, $p = 0.001$, d = 1.04, 95% CI [128.47, 397.32] ms, the interaction effects with reward value remained nonsignificant. Furthermore, a linear contrast confirmed that localization RTs increased monotonically as the sound source moved further from central fixation, $t(15) = 2.17$, $p = 0.047$, $d = 0.54$, 95% CI [3.46, 399.47] ms.

The analysis of the visual bias (ventriloquism effect) revealed a robust value-driven effect: high-value visual distractors induced a significantly larger visual bias (7.93°) compared to low-value distractors (7.31°), $F(1, 15) = 13.09$, $p = 0.003$, $\eta_p^2 = 0.466$, 95% CI [0.25, 0.98]. Notably, this crossmodal capture of localization accuracy was not significantly modulated by auditory perceptual load (M = 7.67 vs. 7.57) or its interaction with reward value (95% CI [−0.32, 0.90], $p > 0.33$) (Figure 9). Although the magnitude of the visual bias decreased over the course of the experiment ($F(1, 15) = 13.66$, $p = 0.002$, $\eta_p^2 = 0.477$, 95% CI [0.91, 3.39]), the relative influence of distractor value remained constant (p>0.12) (Figure 10).

Finally, spatial position modulated the magnitude of the visual bias on auditory localizations.

For visual bias on sound localization, a 7 (sound location) × 2 (auditory perceptual load) × 2 (visual distractor value) repeated measures ANOVA showed that the main effect of sound location was marginally significant, $F(6, 90) = 2.12$, $p = 0.058$, $\eta_p^2 = 0.124$. Post-hoc comparisons with Bonferroni correction revealed that the visual bias at the central 0° position (8.67°) was significantly greater than at the leftmost 36° position (6.60°), with a mean difference of 2.07° (SE = 0.75, 95% CI [0.46, 3.68]), $p = 0.015$; although it was also numerically higher than other positions, the differences were not significant (ps> 0.8). The interaction between sound location and auditory perceptual load ($F(6, 90)$ = 1.63, $p = 0.148$, $\eta_p^2 = 0.098$), the interaction between sound location and visual distractor value ($F(6, 90) = 1.42$, $p = 0.214$, $\eta_p^2 = 0.087$), and the three-way interaction ($F(6, 90) = 0.10$, $p = 0.996$, η_p^2 = 0.007) were all nonsignificant (Figure 10).

For RTs, a 7 (sound location) × 2 (auditory perceptual load) × 2 (visual distractor value) repeated measures ANOVA showed that the main effect of sound location was significant, $F(6, 90)$ = 32.95, $p < 0.001$, $\eta_p^2 = 0.687$, with overall performance showing slower responses as the distance from the center increased. The interaction effects between sound location and auditory perceptual load ($F(6, 90) = 1.19$, $p = 0.318$, $\eta_p^2 = 0.074$), between sound location and visual distractor value ($F(6, 90) = 1.11$, p = 0.363, $\eta_p^2 = 0.069$), and the three-way interaction ($F(6, 90) = 1.04$, $p = 0.404$, η_p^2= 0.065) were all nonsignificant (Figure 11). These findings indicate that while reward value consistently enhances

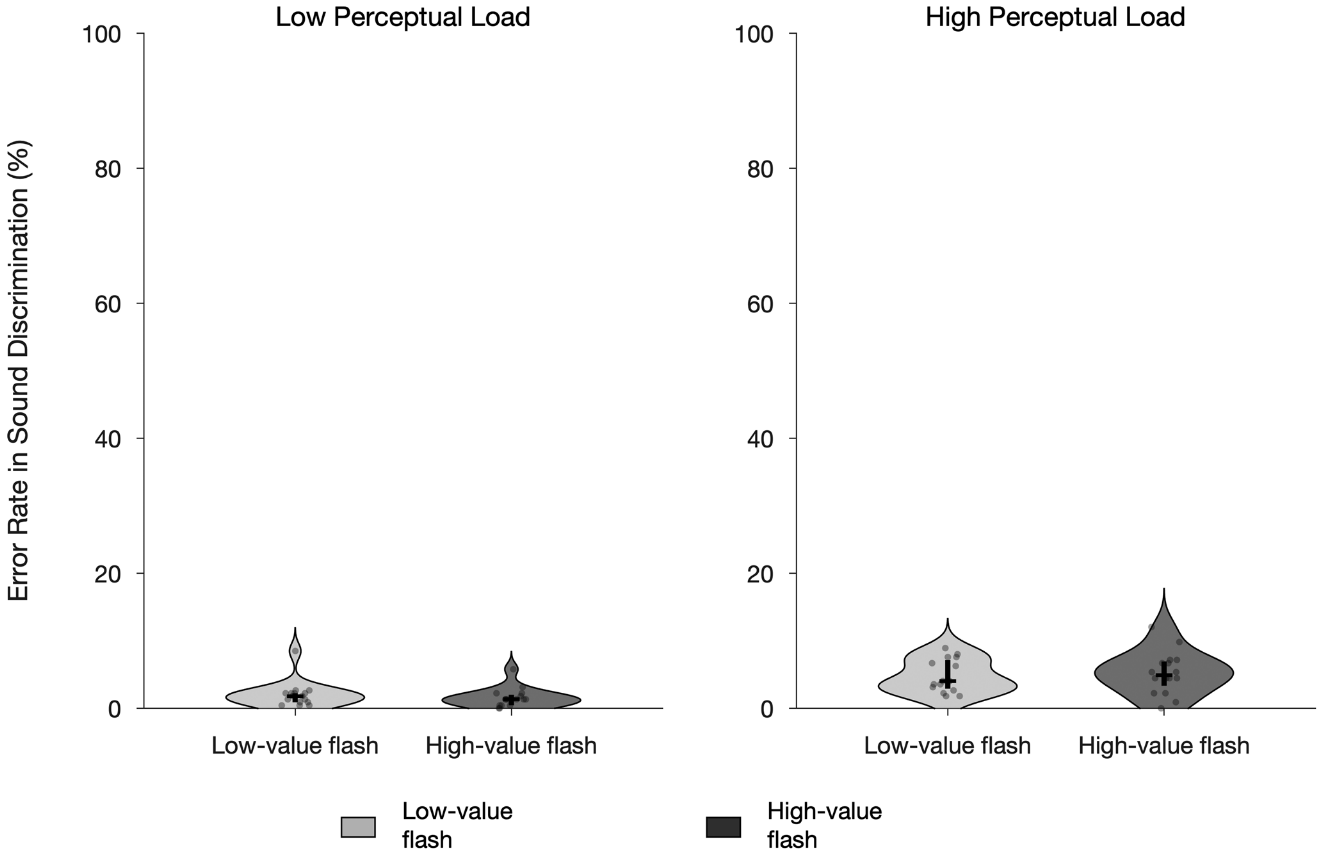

FIGURE 7 | Error rates in sound discrimination (go/no-go task) as a function of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value flash) in Experiment 2. Violin plots show the kernel-density distribution of individual participant means. Thick vertical bars indicate the interquartile range (IQR), and horizontal bars mark the median. Individual data points are overlaid as jittered dots.

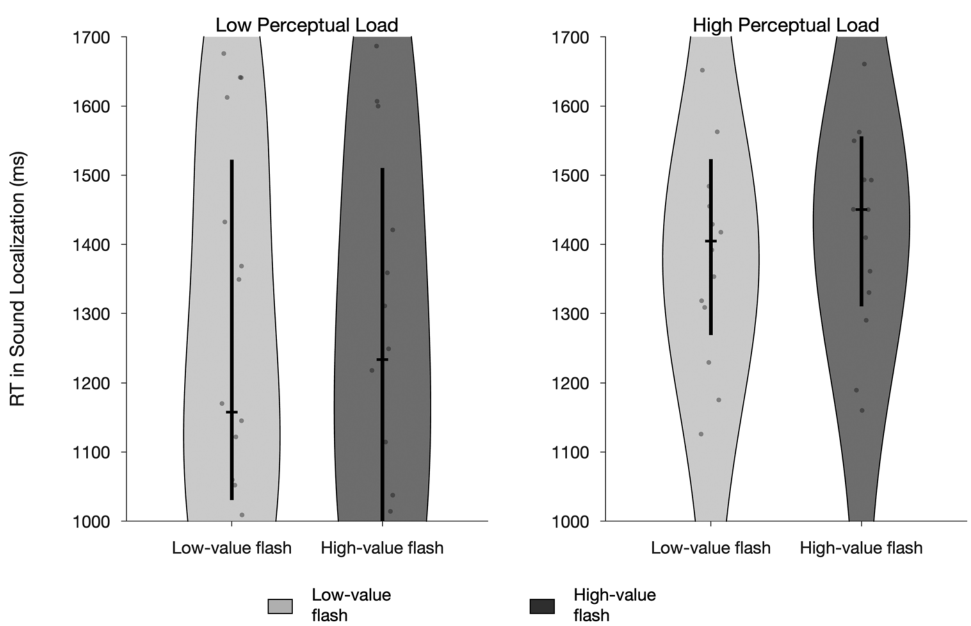

FIGURE 8 | Reaction times (RTs) in sound localization as a function of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value flash) in Experiment 2. Violin plots show the kernel-density distribution of individual participant mean RTs. Thick vertical bars indicate the IQR, and horizontal bars mark the median. Individual data points are overlaid as jittered dots.

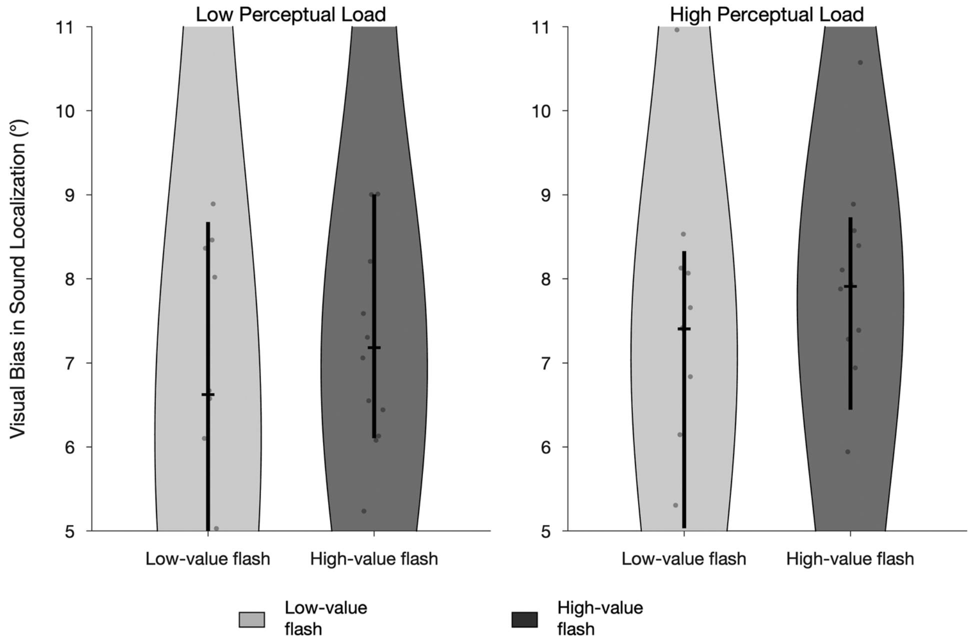

FIGURE 9 | Visual bias in sound localization (ventriloquism effect, in degrees) as a function of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value flash) in Experiment 2. Positive values indicate a bias toward the right visual flash. Violin plots show the kernel-density distribution of individual participant mean biases. Thick vertical bars indicate the IQR, and horizontal bars mark the median. Individual data points are overlaid as jittered dots.

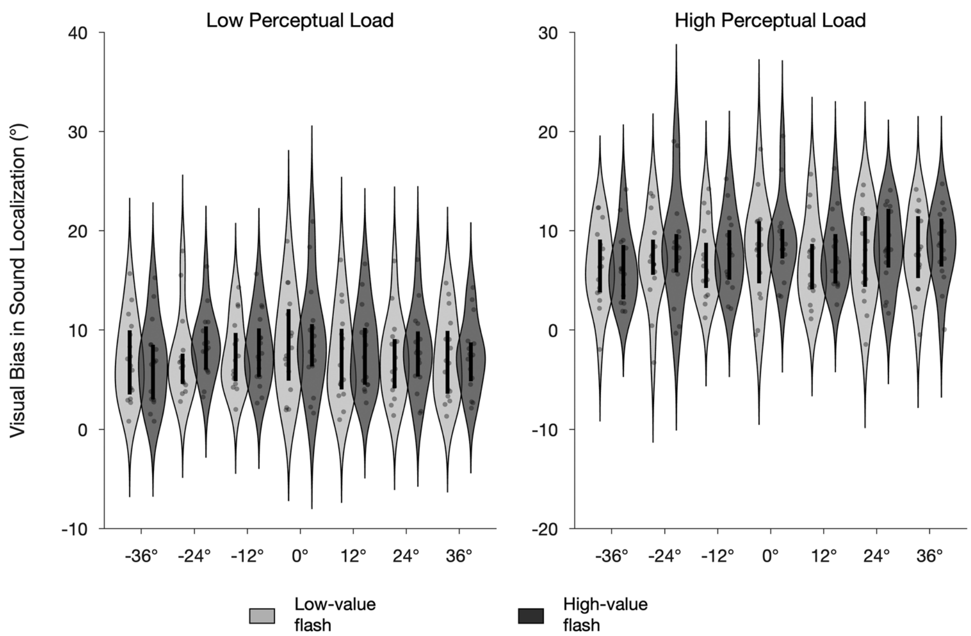

FIGURE 10 | Visual bias in sound localization (°) across seven sound locations (−36° to +36°), separately for low perceptual load (left panel) and high perceptual load (right panel) for Experiment 2. Within each panel, light gray violins represent low-value flash trials and dark gray violins represent high-value flash trials. Violin plots show the kernel-density distribution of individual participant mean biases. Thick vertical bars indicate the IQR, and horizontal bars mark the median. Individual data points are overlaid as jittered dots.

the capture of auditory space by visual distractors, increasing the cognitive load of the auditory task primarily impacts the speed of processing rather than the ultimate perceptual integration of the multisensory signals.

3.2 | Discussion

Consistent with the findings of Experiment 1, the results of Experiment 2 demonstrate that while the value of a visual distractor does not affect sound discrimination, it significantly impacts

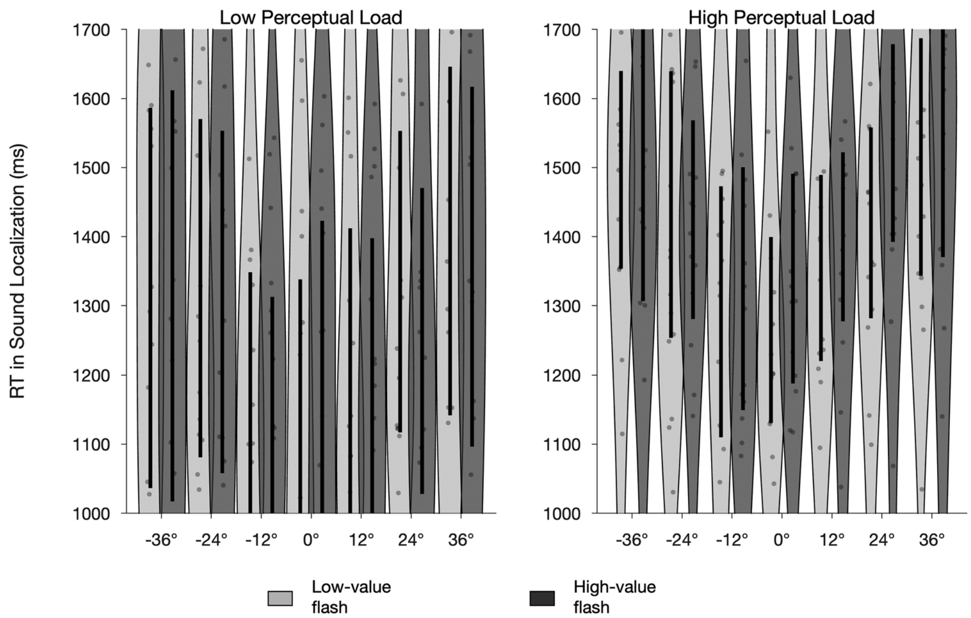

FIGURE 11 | Reaction times in sound localization (ms) across seven sound locations (−36° to +36°) in Experiment 2, separately for low perceptual load (left panel) and high perceptual load (right panel). Within each panel, light gray violins represent low-value flash trials and dark gray violins represent high-value flash trials. Violin plots show the kernel-density distribution of individual participant mean RTs. Thick vertical bars indicate the IQR, and horizontal bars mark the median. Individual data points are overlaid as jittered dots.

auditory localization. High-value visual distractors elicited a larger visual bias than low-value distractors, indicating a more pronounced spatial ventriloquism effect. This provides further evidence that VDAC can occur in crossmodal settings without prior response training or reinforcement of the task-irrelevant stimulus. Furthermore, this value-driven crossmodal integration remains robust and is not modulated by perceptual load.

A key divergence between Experiment 2 and Experiment 1 lies in the RT data pattern: the difference between high- and low-value visual stimuli was significant only under high perceptual load, disappearing under low-load conditions. This pattern aligns with previous research where value-based distractors were presented within the focus of attention [19, 20]. These results confirm that the spatial position of the distractor fundamentally alters the efficacy of perceptual load. When a distractor cannot be excluded through attentional zooming because it falls within the task's focus, increasing perceptual load fails to mitigate interference [17]. In such cases, high-reward stimuli possess a competitive advantage in capturing attention, specifically when cognitive demands are high.

Analysis of the sound position effect revealed that localization RTs increased as the source moved further from the center. This is expected, given that the laser pointer was reset to the central position before each trial. Regarding the spatial ventriloquism effect, the magnitude of visual bias was greatest at the 0° (central) position, differing significantly from the far-left position. This may be attributed to the center serving as the primary locus of visual attention, making central visual stimuli more likely to interfere with auditory localization. Alternatively, the current feedback mechanism may have induced a spatial bias; because reward feedback was always given centrally, this information might have increased attentional prioritization of the central area. Indeed, Chelazzi et al. demonstrated that reward-based learning can shift spatial priority maps, with high-reward locations receiving preferential attention [34].

The finding that the ventriloquism effect was minimized at the far-left position (while showing no significant difference between the center and far-right) suggests a degree of spatial asymmetry in attention. Similar asymmetries have been observed in studies of the spatial ventriloquism after-effects under perceptual load, where left-sided stimuli often fail to elicit significant shifts compared to other locations [35]. This rightward bias is well-documented in dual-task paradigms: adding an auditory task to a visual memory task can shift visual attention to the right [36], and introducing a secondary visual task can cause participants to perceive right-sided stimuli as appearing earlier in temporal order judgments [37].

Supplementary block analyses (Figures S5 and S6) confirm the stability of this dissociation over time. Sound localization RTs exhibited practice-related speeding, yet the interaction between reward value and perceptual load remained stable, and the load effect on localization RTs emerged reliably in block 2. Crucially, the ventriloquism bias remained statistically stable across blocks, with perceptual load never significantly modulating visual bias. Although a value effect on bias appeared under high load in block 1, it did not persist or interact with the block. Thus, the core pattern—load-insensitive spatial integration alongside load-sensitive value-driven RT interference—remained intact throughout the session, supporting an architectural rather than fatigue-based account.

3.3 | General Discussion

The current research utilized the spatial ventriloquism effect to investigate the mechanisms of value-driven crossmodal attentional processing and the potential moderating role of perceptual load. By manipulating the spatial relationship between auditory targets and visual distractors across two distinct localization tasks, we demonstrated that reward associations can trigger crossmodal capture and enhance multisensory integration even in the absence of prior response training. However, the influence of perceptual load on this capture is contingent upon the spatial configuration of the stimuli.

3.3.1 | Value-Driven Capture Without Response Reinforcement

Our findings show that high-value visual distractors interfere more significantly with auditory localization than low-value ones, even when these distractors never served as targets. Unlike the traditional training-test paradigm, which may be confounded by operant conditioning—where the habit of responding to the stimulus is reinforced [26]—our single-stage design associated rewards with task-irrelevant distractors. Because attending to these distractors led to slower responses and potential reward loss, the observed capture reflects a robust, automatic priority mechanism rather than learned motor reinforcement [24]. Furthermore, this value-driven capture enhanced multisensory integration, resulting in a stronger spatial ventriloquism effect for high-value stimuli [14, 27].

3.3.2 | The Mediating Role of Spatial Configuration on Perceptual Load

The interaction between perceptual load and value-driven capture was found to be contingent on stimulus location. In Experiment 1, which utilized a central target and peripheral distractors, increasing perceptual load reduced interference, consistent with the attentional zoom theory, where high load narrows the focus to the center [17]. Conversely, Experiment 2 revealed that when stimuli are proximal, and targets are less spatially certain, high load actually increased the interference gap between high- and low-value distractors. When distractors fall within the task's immediate attentional range, they cannot be filtered out; instead, the increased cognitive focus associated with high load may inadvertently heighten the capture by reward-linked stimuli [20].

3.3.3 | The Role of Automaticity in Crossmodal Integration

Notably, while perceptual load modulated RTs, it did not affect the accuracy-based spatial ventriloquism effect in either experiment. Under high perceptual load, the visual system could exhaust its perceptual capacity on the primary task, leaving fewer residual resources for the value-driven priority signal to influence response selection, thereby reducing RT interference but without abolishing the underlying multisensory spatial bias. This suggests that crossmodal integration may be an automatic process that persists even when attentional resources for the distractor modality are reduced [38]. While our findings demonstrate the persistence of the visual bias under the specific load conditions tested, they are insufficient to claim that multisensory integration is entirely automatic, nor is it wholly independent of general attentional resources, given that the intricate relationship between perceptual load and attention is beyond the scope of this study. Our results imply that discrepancies in previous crossmodal load studies [21–23] may depend on whether the stimuli across modalities are integrated into a single object or remain independent. This has practical implications for managing distractions: while increasing task difficulty (perceptual load) can help ignore peripheral distractions, it may paradoxically increase distraction if the valued object—such as a smartphone—remains within the immediate focus of attention.

3.3.4 | Limitations, Future Research, and Conclusion

It is important to interpret the findings of Experiment 2 with caution due to the limited sample size ($n = 16$) caused by pandemic-related testing curtailments. Because this early termination forced a reliance on post-hoc power analyses, our confidence in the robustness of the observed interaction claims is necessarily weakened. To avoid overstating these findings, we consider the interaction between perceptual load and distractor value in this experiment to be strictly exploratory and preliminary. Future research employing adequately powered, a priori sample planning is required to confidently replicate these interaction effects and validate the precise boundaries of this visual bias.

While this study provides evidence for value-driven crossmodal capture, several limitations remain. First, these conclusions may not generalize across all sensory pairings, such as olfactory or tactile modalities, which possess unique physiological and social value signatures [39, 40]. Future research should explore whether the immunity of integration to perceptual load holds true across more diverse multisensory contexts. Additionally, while we focused on positive monetary rewards, the role of negative valence—such as loss or pain—remains unexplored in this framework [19, 33].

From a methodological perspective, the use of a binary load manipulation (low vs. high) may not capture the threshold at which multisensory integration is finally affected [18]. Future studies should consider implementing multiple levels of load to determine if a breaking or transitional point exists for crossmodal integration. Finally, while behavioral data clearly distinguish between capture (speed) and integration (accuracy), more direct and stronger evidence could be gathered through electrophysiological or neuroimaging techniques. Analyzing ERP components such as the N2pc for attentional capture [41] or the N260 for ventriloquism-related auditory cortex activation would provide more precise data on the temporal and structural neural mechanisms at play [42].

Future directions should substantiate the temporal and network-level signatures of the present behavioral dissociation. Electrophysiological studies could determine if high-value distractors and load interact to modulate early sensory event-related potentials (ERPs) and late spatial tracking components [42–44]. Furthermore, functional magnetic resonance imaging (fMRI) and connectivity analyses might clarify whether load-dependent spatial processing [45] and value-driven cross-modal capture rely on dissociable superior temporal sulcus (STS) /intraparietal sulcus (IPS) and orbitofrontal networks [46]. Time-frequency analyses could also reveal whether fronto-medial theta oscillations track the conflict between task goals and reward-driven capture [47, 48].

Additionally, constraints on the current study's generalizability motivate several behavioral extensions. Because our design relied on monetary rewards, audiovisual dyads, and healthy young adults, future directions should explore whether these load-by-value interactions generalize to other reinforcers and across other multisensory domains [49]. Lifespan and clinical studies—particularly in schizophrenia, where multisensory temporal binding windows are altered [50]—are necessary to map the boundaries of reward-driven crossmodal capture. Finally, parametric manipulations of intersensory disparity could help determine whether value-driven capture engages the same causal-inference competition mechanisms as neutral spatial conflict [47].

In conclusion, this research reveals that in audiovisual spatial attention, visual distractors associated with reward can trigger value-driven crossmodal capture and enhance multisensory integration without prior response training. The effect of perceptual load on RT interference is dictated by the spatial positioning of the stimuli: high load reduces peripheral interference but amplifies interference from distractors within the attentional focus. Critically, perceptual load does not modulate the spatial ventriloquism effect, suggesting that value-driven crossmodal integration (realized in the spatial ventriloquism effect) operates on an automatic mechanism that is relatively independent of top-down attentional resource allocation. While the observed pattern aligns with our theoretical predictions, we acknowledge that the modest sample size (particularly for Experiment 2) limits statistical power, and future studies with larger cohorts are needed to confirm the robustness of these effects.

Author Contributions

Qiongyu Wang: Conceptualization, methodology, investigation, formal analysis, writing – original draft, visualization. Lihan Chen: Conceptualization, methodology, funding acquisition, project administration, formal analysis, writing – original draft, visualization, writing – review and editing, validation, supervision.

Funding

The study was funded by STI2030-Major Project (2021ZD0202600) and the Natural Science Foundation of China (T2192932, 31861133012) grant to L.C.

Ethics Statement

The study protocol was approved by the Committee for Protecting Human and Animal Subjects at the School of Psychological and Cognitive Sciences of Peking University.

Consent Statement

Informed consent was obtained from participants prior to the commencement of the experiments, and they agreed to their data for publication without disclosing their private information.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

Peer Review

For transparency, the peer review documents associated with this article are available at https://doi.org/10.1111/nyas.70346.

References

1. M. I. Posner and M. K. Rothbart, "Research on Attention Networks as a Model for the Integration of Psychological Science," *Annual Review of Psychology* 58 (2007): 1–23, https://doi.org/10.1146/annurev.psych.58.110405.085516.

2. M. Corbetta and G. L. Shulman, "Control of Goal-Directed and Stimulus-Driven Attention in the Brain," *Nature Reviews Neuroscience* 3, no. 3 (2002): 201–215, https://doi.org/10.1038/nrn755.

3. J. Theeuwes, "Top–Down and Bottom–Up Control of Visual Selection," *Acta Psychologica* 135, no. 2 (2010): 77–99, https://doi.org/10.1016/j.actpsy.2010.02.006.

4. B. A. Anderson, P. A. Laurent, and S. Yantis, "Value-Driven Attentional Capture," *Proceedings of the National Academy of Sciences* 108, no. 25 (2011): 10367–10371, https://doi.org/10.1073/pnas.1104047108.

5. M. Failing and J. Theeuwes, "Selection History: How Reward Modulates Selectivity of Visual Attention," *Psychonomic Bulletin & Review* 25, no. 2 (2018): 514–538, https://doi.org/10.3758/s13423-017-1380-y.

6. E. Awh, A. V. Belopolsky, and J. Theeuwes, "Top-Down Versus Bottom-Up Attentional Control: A Failed Theoretical Dichotomy," *Trends in Cognitive Sciences* 16, no. 8 (2012): 437–443, https://doi.org/10.1016/j.tics.2012.06.010.

7. V. Maljkovic and K. Nakayama, "Priming of Pop-Out: I. Role of Features," *Memory & Cognition* 22, no. 6 (1994): 657–672, https://doi.org/10.3758/BF03209251.

8. J. Theeuwes and E. Van der Burg, "On the Limits of Top-Down Control of Visual Selection," *Attention, Perception, & Psychophysics* 73, no. 7 (2011): 2092–2103, https://doi.org/10.3758/s13414-011-0176-9.

9. M. M. Chun and Y. Jiang, "Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention," *Cognitive Psychology* 36, no. 1 (1998): 28–71, https://doi.org/10.1006/cogp.1998.0681.

10. B. Wang and J. Theeuwes, "Statistical Regularities Modulate Attentional Capture," *Journal of Experimental Psychology: Human Perception and Performance* 44, no. 1 (2018): 13–17, https://doi.org/10.1037/xhp0000472.

11. B. A. Anderson, "A Value-Driven Mechanism of Attentional Selection," *Journal of Vision* 13, no. 3 (2013): 7, https://doi.org/10.1167/13.3.7.

12. B. A. Anderson, "Value-Driven Attentional Capture in the Auditory Domain," *Attention, Perception, & Psychophysics* 78, no. 1 (2016): 242–250, https://doi.org/10.3758/s13414-015-1001-7.

13. P. Bruns, M. Maiworm, and B. Röder, "Reward Expectation Influences Audiovisual Spatial Integration," *Attention, Perception, & Psychophysics* 76, no. 6 (2014): 1815–1827, https://doi.org/10.3758/s13414-014-0699-y.

14. X. Luo, G. Kang, Y. Guo, X. Yu, and X. Zhou, "A Value-Driven McGurk Effect: Value-Associated Faces Enhance the Influence of Visual Information on Audiovisual Speech Perception and Its Eye Movement Pattern," *Attention, Perception, & Psychophysics* 82, no. 4 (2020): 1928–1941, https://doi.org/10.3758/s13414-019-01918-x.

15. N. Lavie, "Distracted and Confused?: Selective Attention Under Load," *Trends in Cognitive Sciences* 9, no. 2 (2005): 75–82, https://doi.org/10.1016/j.tics.2004.12.004.

16. D. M. Beck and N. Lavie, "Look Here but Ignore What You See: Effects of Distractors at Fixation," *Journal of Experimental Psychology: Human Perception and Performance* 31, no. 3 (2005): 592–607, https://doi.org/10.1037/0096-1523.31.3.592.

17. Z. Chen and K. R. Cave, "Zooming in on the Cause of the Perceptual Load Effect in the Go/No-Go Paradigm," *Journal of Experimental Psychology: Human Perception and Performance* 42, no. 8 (2016): 1072–1087, https://doi.org/10.1037/xhp0000168.

18. N. Lavie, T. Ro, and C. Russell, "The Role of Perceptual Load in Processing Distractor Faces," *Psychological Science* 14, no. 5 (2003): 510–515, https://doi.org/10.1111/1467-9280.03453.

19. R. Gupta, Y. J. Hur, and N. Lavie, "Distracted by Pleasure: Effects of Positive Versus Negative Valence on Emotional Capture Under Load," *Emotion* 16, no. 3 (2016): 328–337, https://doi.org/10.1037/emo0000112.

20. T. Feldmann-Wüstefeld, R. Brandhofer, and A. Schubö, "Rewarded Visual Items Capture Attention Only in Heterogeneous Contexts," *Psychophysiology* 53, no. 7 (2016): 1063–1073, https://doi.org/10.1111/psyp.12641.

21. J. S. Macdonald and N. Lavie, "Visual Perceptual Load Induces Inattentional Deafness," *Attention, Perception, & Psychophysics* 73, no. 6 (2011): 1780–1789, https://doi.org/10.3758/s13414-011-0144-4.

22. N. A. Parks, M. R. Hilimire, and P. M. Corballis, "Steady-State Signatures of Visual Perceptual Load, Multimodal Distractor Filtering, and Neural Competition," *Journal of Cognitive Neuroscience* 23, no. 5 (2011): 1113–1124, https://doi.org/10.1162/jocn.2010.21460.

23. G. Rees, C. Frith, and N. Lavie, "Processing of Irrelevant Visual Motion During Performance of an Auditory Attention Task," *Neuropsychologia* 39, no. 9 (2001): 937–949, https://doi.org/10.1016/S0028-3932(01)00016-1.

24. B. A. Anderson, "The Attention Habit: How Reward Learning Shapes Attentional Selection," *Annals of the New York Academy of Sciences* 1369, no. 1 (2016): 24–39, https://doi.org/10.1111/nyas.12957.

25. B. A. Anderson, "Social Reward Shapes Attentional Biases," *Cognitive Neuroscience* 7, no. 1–4 (2016): 30–36, https://doi.org/10.1080/17588928.2015.1047823.

26. M. E. Le Pelley, D. Pearson, O. Griffiths, and T. Beesley, "When Goals Conflict With Values: Counterproductive Attentional and Oculomotor Capture by Reward-Related Stimuli," *Journal of Experimental Psychology General* 144, no. 1 (2015): 158–171.

27. L. R. D. Sanz, P. Vuilleumier, and A. Bourgeois, "Cross-Modal Integration During Value-Driven Attentional Capture," *Neuropsychologia* 120 (2018): 105–112, https://doi.org/10.1016/j.neuropsychologia.2018.10.014.

28. N. Lavie, "Perceptual Load as a Necessary Condition for Selective Attention," *Journal of Experimental Psychology: Human Perception and Performance* 21, no. 3 (1995): 451–468, https://doi.org/10.1037/0096-1523.21.3.451.

29. Z. Chen, "Attentional Focus, Processing Load, and Stroop Interference," *Perception & Psychophysics* 65, no. 6 (2003): 888–900, https://doi.org/10.3758/BF03194822.

30. P. Bertelson and G. Aschersleben, "Automatic Visual Bias of Perceived Auditory Location," *Psychonomic Bulletin & Review* 5, no. 3 (1998): 482–489, https://doi.org/10.3758/BF03208826.

31. B. E. Stein and T. R. Stanford, "Multisensory Integration: Current Issues From the Perspective of the Single Neuron," *Nature Reviews Neuroscience* 9, no. 4 (2008): 255–266, https://doi.org/10.1038/nrn2331.

32. T. Rohe and U. Noppeney, "Sensory Reliability Shapes Perceptual Inference via Two Mechanisms," *Journal of Vision* 15, no. 5 (2015): 22, https://doi.org/10.1167/15.5.22.

33. L. Wang, H. Yu, and X. Zhou, "Interaction Between Value and Perceptual Salience in Value-Driven Attentional Capture," *Journal of Vision* 13, no. 3 (2013): 5, https://doi.org/10.1167/13.3.5.

34. L. Chelazzi, J. E to inova, R. Calletti, et al., "Altering Spatial Priority Maps via Reward-Based Learning," *Journal of Neuroscience* 34, no. 25 (2014): 8594–8604, https://doi.org/10.1523/JNEUROSCI.0277-14.2014.

35. R. Eramudugolla, M. R. Kamke, S. Soto-Faraco, and J. B. Mattingley, "Perceptual Load Influences Auditory Space Perception in the Ventriloquist Aftereffect," *Cognition* 118, no. 1 (2011): 62–74, https://doi.org/10.1016/j.cognition.2010.09.009.

36. P. V. Peers, R. Cusack, and J. Duncan, "Modulation of Spatial Bias in the Dual Task Paradigm: Evidence From Patients With Unilateral Parietal Lesions and Controls," *Neuropsychologia* 44, no. 8 (2006): 1325–1335, https://doi.org/10.1016/j.neuropsychologia.2006.01.033.

37. A. Pérez, P. V. Peers, M. Valdés-Sosa, L. Galán, L. García, and E. Martínez-Montes, "Hemispheric Modulations of Alpha-Band Power Reflect the Rightward Shift in Attention Induced by Enhanced Attentional Load," *Neuropsychologia* 47, no. 1 (2009): 41–49, https://doi.org/10.1016/j.neuropsychologia.2008.08.017.

38. P. Bertelson, F. Pavani, E. Ladavas, J. Vroomen, and B. de Gelder, "Ventriloquism in Patients With Unilateral Visual Neglect," *Neuropsychologia* 38, no. 12 (2000): 1634–1642, https://doi.org/10.1016/S0028-3932(00)00067-1.

39. J. A. Gottfried and D. H. Zald, "On the Scent of Human Olfactory Orbitofrontal Cortex: Meta-Analysis and Comparison to Non-Human Primates," *Brain Research Reviews* 50, no. 2 (2005): 287–304, https://doi.org/10.1016/j.brainresrev.2005.08.004.

40. W. Li, I. Moallem, K. A. Paller, and J. A. Gottfried, "Subliminal Smells Can Guide Social Preferences," *Psychological Science* 18, no. 12 (2007): 1044–1049, https://doi.org/10.1111/j.1467-9280.2007.02023.x.

41. S. Qi, Q. Zeng, C. Ding, and H. Li, "Neural Correlates of Reward-Driven Attentional Capture in Visual Search," *Brain Research* 1532 (2013): 32–43, https://doi.org/10.1016/j.brainres.2013.07.044.

42. B. Bonath, T. Noesselt, A. Martinez, et al., "Neural Basis of the Ventriloquist Illusion," *Current Biology* 17, no. 19 (2007): 1697–1703, https://doi.org/10.1016/j.cub.2007.08.050.

43. W. A. Bainbridge, Z. Pounder, A. F. Eardley, and C. I. Baker, "Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory," *Cortex* 135 (2021): 159–172, https://doi.org/10.1016/j.cortex.2020.11.014.

44. X. Tang, T. Zhang, J. Sun, and S. Lu, "Echoes of the mind's eye: Reciprocal crossmodal interaction between auditory and visual processing," *iScience* 29, no. 3 (2026): 114990, https://doi.org/10.1016/j.isci.2026.114990.

45. A. Callan, D. Callan, and H. Ando, "An fMRI study of the ventriloquism effect," *Cerebral Cortex* 25, no. 11 (2015): 4248–4258, https://doi.org/10.1093/cercor/bhu306.

46. J. E. Antono, S. Dang, R. Auksztulewicz, and A. Pooresmaeili, "Distinct patterns of connectivity between brain regions underlie the intra-modal and cross-modal value-driven modulations of the visual cortex," *The Journal of Neuroscience* 43, no. 44 (2023): 7361–7375, https://doi.org/10.1523/JNEUROSCI.0355-23.2023.

47. A. Marly, A. Yazdjian, and S. Soto-Faraco, "The role of conflict processing in multisensory perception: Behavioural and electroencephalography evidence," *Philosophical Transactions of the Royal Society B* 378, no. 1886 (2023): 20220346, https://doi.org/10.1098/rstb.2022.0346.

48. M. Kaiser, D. Senkowski, and J. Keil, "Mediofrontal theta-band oscillations reflect top-down influence in the ventriloquist illusion," *Human Brain Mapping* 42, no. 2 (2021): 452–466, https://doi.org/10.1002/hbm.25236.

49. M. Overgaard, A. Kirkeby-Hinrup, T. L. Kvamme, and Q. J. Wang, "Multisensory interplay and theories of consciousness Psychology of Consciousness: Theory, Research, and Practice," *Psychology of Conscious-*

ness: Theory, Research, and Practice (2024), Advance online publication, https://doi.org/10.1037/cns0000407.

50. A. Borgolte, C. Sinke, L. Michalke, et al., "Neural correlates of audiovisual integration in schizophrenia – an ERP study," *Frontiers in Psychiatry* 15 (2024): 1372456, https://doi.org/10.3389/fpsyt.2024.1492266.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supplementary Figure S1: The apparatus for Experiment 1.
Supplementary Figure S2: Congruency effect in sound localization reaction time (RT; Incongruent – Congruent) across experimental blocks (Block 1 vs. Block 2), plotted separately for each combination of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value) for Experiment 1. **Supplementary Figure S3**: Congruency effect in sound localization accuracy (Congruent – Incongruent) across experimental blocks (Block 1 vs. Block 2), plotted separately for each combination of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value) for Experiment 1. **Supplementary Figure S4**: The apparatus for Experiment 2. **Supplementary Figure S5**: Sound localization RT (ms) across experimental blocks (Block 1 vs. Block 2) in Experiment 2, plotted separately for each combination of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value). **Supplementary Figure S6**: Visual bias in sound localization (°) across experimental blocks (Block 1 vs. Block 2) in Experiment 2, plotted separately for each combination of perceptual load (low vs. high) and visual distractor value (low-value vs. high-value).