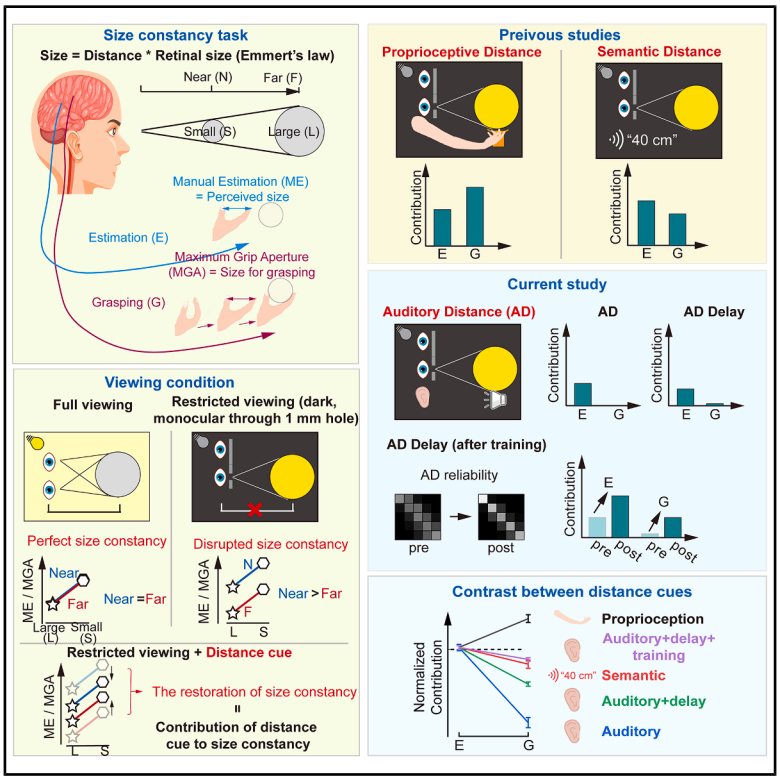
iScience

Article

Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing

Graphical abstract



Highlights

- In the restricted viewing, auditory cues improve perceptual size judgments
- In the restricted viewing, auditory cues do not affect grasping size constancy
- Grasping size constancy improved moderately after auditory distance training
- Grasping integrates auditory distance less easily than proprioceptive distance

Authors

Chao Zheng, Gexiu Wang, Xiaoming Zhou, Jie Gao, Irene Sperandio, Melvyn A. Goodale, Juan Chen

Correspondence

juanchen@m.scnu.edu.cn

In brief

Neuroscience; Sensory neuroscience; Cognitive neuroscience

Zheng et al., 2025, iScience ■■, 113341
■■, 2025
https://doi.org/10.1016/j.isci.2025.113341

CellPress

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

iScience

CellPress
OPEN ACCESS

Article

Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing

Chao Zheng,[1,2,3,7] Gexiu Wang,[1,7] Xiaoming Zhou,[4] Jie Gao,[1] Irene Sperandio,[5] Melvyn A. Goodale,[6] and Juan Chen[1,2,3,8,*]

[1]Key Laboratory of Brain, Cognition and Education Sciences (South China Normal University), Ministry of Education, Guangzhou, Guangdong Province 510631, China
[2]Center for the Study of Applied Psychology, Guangdong Key Laboratory of Mental Health and Cognitive Science, and the School of Psychology, South China Normal University, Guangzhou, Guangdong Province 510631, China
[3]Philosophy and Social Science Laboratory of Reading and Development in Children and Adolescents (South China Normal University), Ministry of Education, Guangzhou, China
[4]Key Laboratory of Brain Functional Genomics of Ministry of Education, Shanghai Key Laboratory of Brain Functional Genomics, School of Life Sciences, East China Normal University, Shanghai 200062, China
[5]Department of Psychology and Cognitive Science, University of Trento, 38068 Rovereto, TN, Italy
[6]Western Institute for Neuroscience and the Department of Psychology, The University of Western Ontario, London, ON N6A 5C2, Canada
[7]These authors contributed equally
[8]Lead contact
*Correspondence: juanchen@m.scnu.edu.cn
https://doi.org/10.1016/j.isci.2025.113341

SUMMARY

When vision is restricted, proprioceptive distance cues fully restore size constancy for scaling grip aperture when grasping objects, despite only limited improvement in perceptual judgments of object size. This suggests that specific task demands and associated neural mechanisms determine the relative weighting of cues during multisensory integration. Is this specific to proprioceptive cues? Here, we examined the contribution of auditory information to perception and action systems under restricted viewing conditions. Surprisingly, in contrast to proprioception, providing auditory distance information had no impact whatsoever on size constancy in grasping but did improve perceptual judgments of size. After participants received extensive training in discriminating distance from auditory cues, there was a modest improvement in grip scaling. Taken together, we suggest that the neural mechanisms mediating grasping cannot incorporate distance information from audition as easily as they can from proprioception when computing real-world object size, but this ability can be improved with training.

INTRODUCTION

Although vision is our dominant sense for perceiving and acting on objects beyond our body, there are cases where vision is absent or limited and we must rely on other sensory information. For example, if you want to pick up a bottle of pop that has fallen on the floor in a dark movie theater, you might use your hands to feel around the area where you heard it fall. In other words, sound and touch can be useful cues for locating objects in darkness.[1,2] There are other situations where your vision of a goal object might be partly obscured, reducing the reliability of visual distance cues. For example, you might be using your left hand to hold the bottle while watching the movie and using your right hand to unscrew the lid. Here, proprioceptive information about the posture of your left hand provides information about the location of the bottle.[3–5] To generate coherent distance and location information in these and other scenarios, we must integrate cues from different sensory systems. It has been suggested that the weighting of each modality in multisensory integration depends on the reliability of the spatial information provided by that sensory modality.[6–8]

Distance cues from visual, auditory, and proprioceptive modalities not only provide spatial information, but they can also influence the computation of object size for both perception and action. One classic example of this kind of integration is size constancy, whereby people integrate the distance information from multiple modalities to compensate for the changes in retinal size[9,10] so that objects in the world appear to be the same size regardless of viewing distance (size constancy for perception[11–13]). Similarly, when we reach out to grasp an object in near space, our grip aperture is tuned to the real size of the goal object regardless of its distance (size constancy for grasping[14,15]) when there is rich distance information. A long history of research, however, suggests that the visual perception of objects and the visual control of actions directed at those objects depend on different computations and are mediated by different

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

neural pathways (i.e., the Two Visual Streams theory).[16,17] Thus, it is entirely possible that cues from different modalities that contribute to size constancy in perception are weighted differently from those contributing to size constancy in grasping.[18]

Several years ago, we tested this possibility by having participants use their left hand to hold a pedestal on which a glowing sphere was mounted while they reached out and grasped the sphere with their right hand. The pedestal was placed randomly at different distances and the size of the glowing ball was varied randomly from trial to trial. Participants viewed the glowing sphere in complete darkness through a 1-mm hole using only one eye, which almost completely eliminated any visual cues to distance.[15] We found, however, that proprioceptive information from their left hand about the location of the pedestal in near space completely restored size constancy for grasping but only moderately improved size constancy for perceptual judgments of size. This finding not only provides compelling evidence for the two visual streams theory,[19] but also challenged the previous theory that the weighting of each modality in multisensory integration was determined by the reliability of the sensory information itself.[6–8] Instead, this finding supports our conjecture that the weighting of sensory signals in multisensory integration depends not only on the reliability of the information itself,[6,8] but also on the task, and thus the "consumer" of that information (i.e., the action system or the perceptual system).

It remains unclear, however, whether the influence of the task on the weighting of different sensory cues in this context is specific to proprioception, or whether the contribution of other sensory distance cues, or even semantic information about distance, could also vary as a function of the task (i.e., the consumer system). In a recent study,[20] we looked at the contribution of semantic information to size constancy (by having participants listen via headphones to verbal information about the distance of the target object) in a paradigm similar to the one we had used to investigate proprioception.[15] In this case, we found that knowing the distance of the target object provided only a modest improvement in size constancy for both perceptual judgments and grasping. Semantic information, it seems, is not a particularly effective way to improve size constancy in either perception or action when visual distance cues are severely limited. Indeed, the contribution of semantic information about distance is more likely to reflect the operation of explicit cognitive processes rather than multisensory integration mechanisms.

When visual cues to distance are limited or absent, it is possible that the sound emitted by a goal object could be used to compute the location of that object in near space, which could then be integrated with information about retinal image size to determine the real-world size of the object. Even so, given that auditory and visual integration is required constantly for things like speech perception but not nearly so often for our physical interactions with the world, we hypothesized that auditory distance information might contribute more to perceptual estimation of size than it would to the control of grasping. To test this, we adopted the same paradigm used in the previous two experiments,[15,20] but provided an auditory distance cue instead. In experiment 1, a burst of white noise was played from a speaker attached to a pedestal at the same time as a glowing sphere became visible through a pinhole (i.e., simultaneous presentation). We found that when auditory distance information was provided, there was a modest improvement in size constancy in perceptual judgments. In contrast, the provision of simultaneous auditory distance information had no impact on size constancy in grip aperture when people reached out to grasp the sphere. This result again suggests that the weighting of sensory signals in multisensory integration depends on the task, but this result is the opposite of what we previously observed with proprioception, which fully restored size constancy for grasping but not for perception.[15]

One possibility for the above finding is that although participants were not able to incorporate auditory cues simultaneously presented with the visual object for grip scaling, they might be able to do so if they were given more time to process the auditory distance cue. Therefore, in experiment 2, the visual stimulus was presented only after the auditory cue was presented. Participants once again showed a modest improvement in perceptual judgments, but there was no evidence of size constancy in their grasping, which confirmed that the grasping system cannot incorporate distance information when computing real-world object size.

Another possibility is that the absence of size constancy in grasping could have simply reflected the fact that participants had only limited experience with the auditory cue during grasping even though it was common to integrate auditory distance cue with the retinal image size during perception. Therefore, in experiment 3, we provided participants with several days of training using the auditory cue to see if we could improve their ability to discriminate between the different distances signaled by the auditory cue. Following training, participants now showed a significant improvement in size constancy in both the perceptual estimation and grasping tasks, similar to the modest contribution of semantic distance cues to both perception and grasping.[20]

At the end of the study, we compared the pattern of results we had obtained with auditory cues to those we had previously observed with proprioceptive[15] and semantic[20] cues with the aim of providing a synthesis of the contribution of various cues to the perception and action systems.

RESULTS

Setup and design of the size constancy test

To measure size constancy, spheres of two sizes (small: 2.5 cm, large: 5 cm in diameter) were placed at two distances (near: 20 cm, far: 40 cm). The near-small and far-large stimuli generated the same image size on the retina ($\sim 7^\circ$ of visual angle) (Figure 1A). To increase the variability of size and distance information and to enhance participant engagement, we included three additional spheres (1.25 cm, 3.75 cm, and 6.25 cm in diameter) and a third distance (30 cm; Figure 1B). However, only four conditions were considered as critical for the analysis of size constancy (i.e., spheres with a diameter of 2.5 cm or 5 cm placed at 20 cm or 40 cm, see Figure 1A). Theoretically, when perfect size constancy is maintained for both perception and action, the target will be perceived as the same size and grasped with the same maximum grip aperture (MGA), regardless of viewing distance[15,20] (Figure 1C, Perfect). When visual information

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

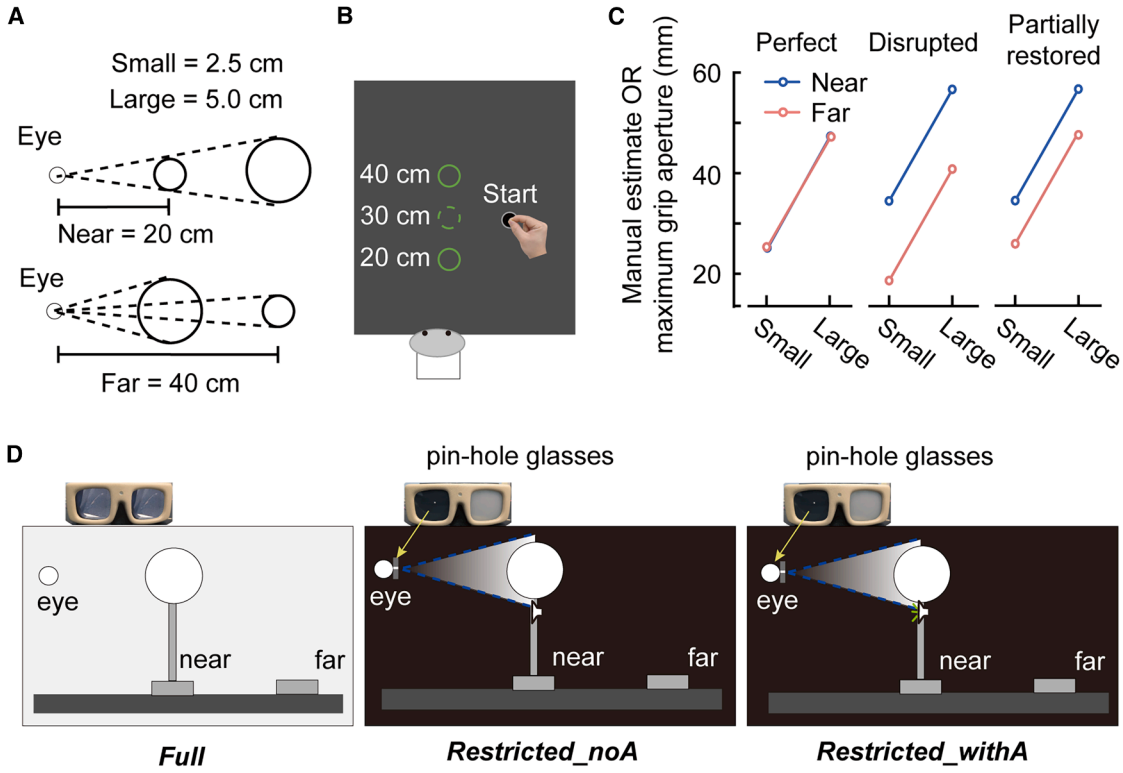


Figure 1. Setup and design of the size constancy test

(A) To measure size constancy, two sizes and two distances were used. The near-small and far-large glowing spheres had the same retinal size.

(B) The layout on the table. The stimulus positions were aligned with the right eye of the participants, so that in addition to the intensity level, they could also use interaural intensity and time differences as cues to judge the distance of the stimulus. In addition to the near (20 cm) and far (40 cm) distances, a third distance condition (30 cm) was added to increase the variability of the task. The distances from the start position of the hand to the near (20 cm) and far (40 cm) positions were identical.

(C) Size constancy in perception refers to the fact that objects are perceived to be the same size, despite changes in visual angle due to varying viewing distances. Similarly, size constancy in grasping refers to the fact that participants grasp object with the same grip aperture regardless of viewing distances. If participants exhibit perfect size constancy, the far and near lines should overlap. If there is insufficient distance information, participants are likely to rely mainly on retinal images to perceive and grasp objects. As a result, they may perceive the same object as larger when it is closer and open their fingers wider to grasp it (i.e., disrupted size constancy, indicated by blue lines on top of red lines). If observers can use distance information to some extent, they may show partial but not complete restoration of size constancy (the gap between blue and red lines gets smaller).

(D) Participants could perform the size constancy test under three distance-cue conditions: (1) full viewing (*Full*), (2) restricted viewing without auditory distance cues (*Restricted-noA*), and (3) restricted viewing with auditory distance cues (*Restricted-withA*). The target spheres were placed on a black pedestal, which was moved to different distances from the observer. In the *Full* condition, participants viewed the target binocularly with the room lights on. In the *Restricted-noA* condition, participants viewed the glowing target sphere monocularly through a 1 mm hole entirely in the dark. In the *Restricted-withA* condition, a burst of white noise was played through the speaker mounted in a hole in the rod of the pedestal. The front of the speaker faced the participant and was centered below the object.

about distance is limited, size constancy is disrupted, leading participants to rely more on retinal-image size for both perception and grasping. As a result, when an object is presented close to the observer, it will be perceived as larger and grasped with a larger grip aperture than when the same object is presented further away (Figure 1C, Disrupted). This was confirmed in our previous studies showing that, when distance information was largely compromised in the restricted-viewing condition, size constancy for both perception and action was disrupted.[15,20] Here, we tested whether providing auditory distance cues can contribute to size constancy mechanisms and if the contribution of the auditory signals can change as a function of the consumer system, namely perception vs. action systems. To put it simply, we tested whether or not the gap in manual estimates (MEs) and/or grip apertures between near and far distances would be reduced with the addition of auditory distance cues (Figure 1C, partially restored).

Participants performed the size constancy test under three distance-cue conditions: (1) full viewing (*Full*), (2) restricted viewing without auditory distance cues (*Restricted-noA*), and (3) restricted viewing with auditory distance cues (*Restricted-withA*) (Figure 1D). The target spheres were placed on a black pedestal, which was moved to different distances from the observer. In the *Full* condition, participants viewed the target binocularly with the room lights on. In the *Restricted-noA* condition, participants viewed the glowing target sphere monocularly through a 1 mm hole entirely in the dark. In the *Restricted-withA* condition, a burst of white noise was played through a speaker mounted in a hole in the rod of the

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

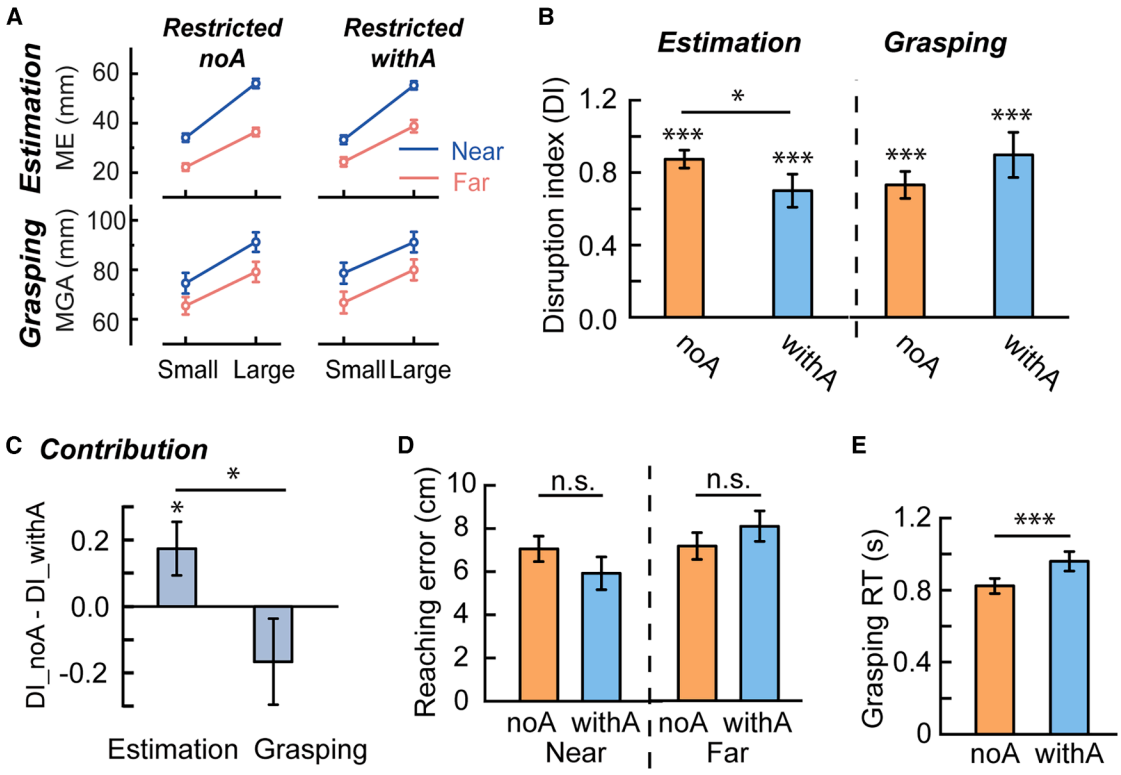


Figure 2. Results of size constancy for experiment 1 when sound and visual objects were presented at the same time (simultaneous presentation)

(A) Top: manual size estimates (ME) of the perceived size of small and large spheres at the near or far distances under two cue conditions: restricted viewing without auditory cue (*Restricted_noA*), and restricted viewing with auditory cue (*Restricted_withA*). Bottom: maximum grip aperture (MGA) during grasping in all conditions.

(B) The normalized size constancy disruption indices (DIs) of the two viewing conditions for both tasks.

(C) Contribution of the auditory cue to size constancy ($DI_{noA} - DI_{withA}$) for manual size estimation and grasping.

(D) The reaching error was defined as the difference between location at the end of the reach and the location of the target. The reaching distance was defined as the distance from the midpoint between the index finger and thumb to the participant on the table surface when the velocity dropped below 10% of the peak velocity.[21]

(E) The reaction time on grasping trials when auditory cue was provided (i.e., withA) or not provided (i.e., noA). Asterisks (*) and (***) indicate that the value is significantly different from zero or the values in two conditions are significantly different at $p < 0.05$ and $p < 0.001$, respectively. Error bars show standard error of mean.

The "n.s." indicates no significant difference between two conditions.

pedestal. The front of the speaker faced the participant and was centered below the object (Figure S1). In all three experiments, participants performed the size constancy tests under the *Restricted-noA*, and *Restricted-withA* cue conditions. In experiment 1, the full-viewing condition was not included because previous studies had demonstrated perfect size constancy under this condition. This was further confirmed in experiments 2 where the full-viewing condition ("*Full*") was included (Figure 3A). In experiment 3, not all participants completed the full-viewing condition because this condition is not directly related to the research question. Therefore, the results were not reported.

Results of size constancy for experiment 1 when sound and visual objects were presented at the same time (simultaneous presentation)

MEs and MGAs

Figure 2A shows MEs and MGAs in the restricted-viewing without auditory cues (*Restricted-noA*) and restricted viewing with auditory information (*Restricted-withA*), respectively. The *Full* condition was not included because previous studies[15,20] consistently demonstrated perfect size constancy in the full-viewing condition. This was further confirmed in experiment 2 in the current study.

In the restricted-viewing condition, when participants performed the tasks monocularly through a 1-mm pinhole in complete darkness without auditory information (i.e., *Restricted-noA*), size constancy was disrupted for both manual estimation and grasping. This was evidenced by a significant main effect of distance (ME: $F_{(1, 21)} = 312.511$, $p < 0.001$, $\eta_p^2 = 0.937$; MGA: $F_{(1, 21)} = 94.030$, $p < 0.001$, $\eta_p^2 = 0.817$) when a 2 sizes $\times$2 distances repeated-measures ANOVA was performed separately for ME and MGA.

The key question is whether the addition of the auditory cue would restore size constancy (i.e., participants' MEs and/or grip apertures reflected the object's real size). As it turned out, in the *Restricted-withA* condition, the main effect of distance

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

was again significant for both ME ($F_{(1, 21)} = 59.006$, $p < 0.001$, $\eta_p^2 = 0.738$) and MGA ($F_{(1, 21)} = 51.236$, $p < 0.001$, $\eta_p^2 = 0.709$), suggesting that size constancy was still disrupted even when auditory cues were provided in the restricted-viewing condition.

To assess any improvement in size constancy after providing auditory cues, we calculated a size-constancy disruption index (DI; Figure 2B), which reflects the difference in ME or MGA between the near and far distances. To compare the results between ME and MGA, the DI was corrected by the slope for MEs or MGAs as a function of object size. This correction is needed because MGA and ME could have different slopes, such that a 1-mm difference for MGA is different from a 1-mm difference for ME (see Methods for details). For ME, the DI decreased significantly when auditory cues were added ($t_{(21)} = 2.160$, $p = 0.043$, Cohen's $d = 0.460$, two-tailed). In contrast, for MGA, there was no significant difference between the *Restricted-withA* and *Restricted-noA* conditions ($t_{(21)} = -1.283$, $p = 0.214$, Cohen's $d = -0.273$, two-tailed). These results suggest that auditory cues contributed to the size constancy mechanisms during perception but did not contribute to size constancy in grasping when visual cues were limited.

Contribution of auditory cues to size constancy in perception and action

To quantify the contribution of auditory cues to size constancy in perception and action, we considered the difference in DI between the *Restricted-withA* and *Restricted-noA* conditions (Figure 2C). We found that the contribution of auditory cues (i.e., DI_{noA}-DI_{withA}) to the estimation task was significantly larger than that for grasping ($t_{(21)} = 2.448$, $p = 0.023$, Cohen's $d = 0.522$, two-tailed). Overall, these results suggest that the auditory cue had more significant contribution to size constancy in estimation than to size constancy in grasping when vision was compromised, which is in contrast to the results of proprioception which completely restored size constancy in grasping but had a moderate contribution to estimation.[15]

It might seem surprising that the provision of auditory cues contributed to a partial restoration of size constancy in perception but did nothing for size constancy in grasping. Given that grasping involves two distinct components: a transport component (i.e., reaching) and a prehension component (i.e., grasping) which are mediated by different neural circuits.[22,23] One possibility is that the auditory distance cue was not incorporated into the calculation of both where to direct the grasp (i.e., the transport component of prehension) and how much to open the finger and thumb in flight (the grasp component of prehension). The other possibility is that the auditory distance cue could support the transport component of prehension but not the grasp component.

Contribution of the auditory cue to reaching distances

To statistically test the contribution of the auditory cue to reaching distances, we calculated the reaching error, which was defined as the reaching distance minus target distance during grasping for the near (20 cm) and far (40 cm) distances separately (Figure 2D). The reaching distance was defined as the distance from the midpoint between the index finger and thumb to the participant on the table surface when the velocity dropped below 10% of the peak velocity.[21,24] With auditory cues, the error of reaching distance was not significantly reduced by auditory cues (Near: NoA vs. withA, $t_{(21)} = 1.354$, $p = 0.190$, Cohen's $d = 0.289$, two-tailed; Far: NoA vs. withA, $t_{(21)} = -1.292$, $p = 0.210$, Cohen's $d = -0.275$, two-tailed). These results suggest that the transport component of prehension did not benefit from the provision of auditory distance cues. However, these findings could not be simply attributed to the poor auditory information because size estimation benefited from the same auditory distance information. Instead, they suggest again that the action system including both the reaching and grasping components could not make use of the auditory information.

Contribution of the auditory cue to reaction time for grasping

We also analyzed whether or not the addition of auditory cues also affects the reaction time for grasping (i.e., the interval between stimulus onset and the movement onset[24]) using repeated-measures ANOVA with Cue condition (withA or noA), size, and distance as within-subject factors. The main effect of Cue was significant ($F_{(1, 20)} = 14.95$, $p < 0.001$, $\eta_p^2 = 0.428$) (Figure 2E), but none of the interaction between Cue and other factors were significant (all $p > 0.204$). The reaction time for grasping in the Restricted_noA (823 ± 194 ms) was shorter than that in the Restricted_withA (960 ± 250 ms). This suggests that participants needed time to incorporate auditory distance cues for grasping.

To test this, we manipulated the presentation time of the visual target in a second experiment. Instead of presenting the visual stimuli at the same time as the auditory cues (i.e., simultaneous presentation), in experiment 2, the burst of white noise was presented right before the opening of the goggles (i.e., delayed presentation).

Results of size constancy for experiment 2 when sound was played right before the presentation of the visual object (delayed presentation)

In experiment 2, the sound and visual target were presented sequentially rather than simultaneously. The same design and analyses were employed as in experiment 1, except that the *Full* condition was included to confirm that size constancy occurred with the current stimulus and testing environment. It should be noted that the *Full* condition was not the focus of our analyses. Our purpose was to investigate the extent to which the addition of auditory cues could restore the disrupted size constancy in the restricted viewing condition (i.e., *Restricted-noA* vs. *Restricted-withA*).

Contribution of auditory cues to size constancy in perception and action

The results of experiment 2 are presented in Figure 3. Similar to what we found in our earlier studies,[15,20] the main effect of distance was not significant in the *Full* viewing condition for both tasks (ME: $F_{(1, 25)} = 1.494$, $p = 0.223$, $\eta_p^2 = 0.056$; MGA: $F_{(1, 25)} = 3.643$, $p = 0.068$, $\eta_p^2 = 0.127$) (Figure 3A). Consistent with experiment 1, the main effect of distance was significant for both ME and MGA in both restricted viewing conditions regardless of the availability of auditory cues (*Restricted_noA*: ME, $F_{(1, 25)} = 302.970$, $p < 0.001$, $\eta_p^2 = 0.924$; MGA, $F_{(1, 25)} = 62.659$, $p < 0.001$, $\eta_p^2 = 0.715$. *Restricted_withA*: ME, $F_{(1, 25)} = 131.050$, $p < 0.001$, $\eta_p^2 = 0.840$; MGA, $F_{(1, 25)} = 86.101$,

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

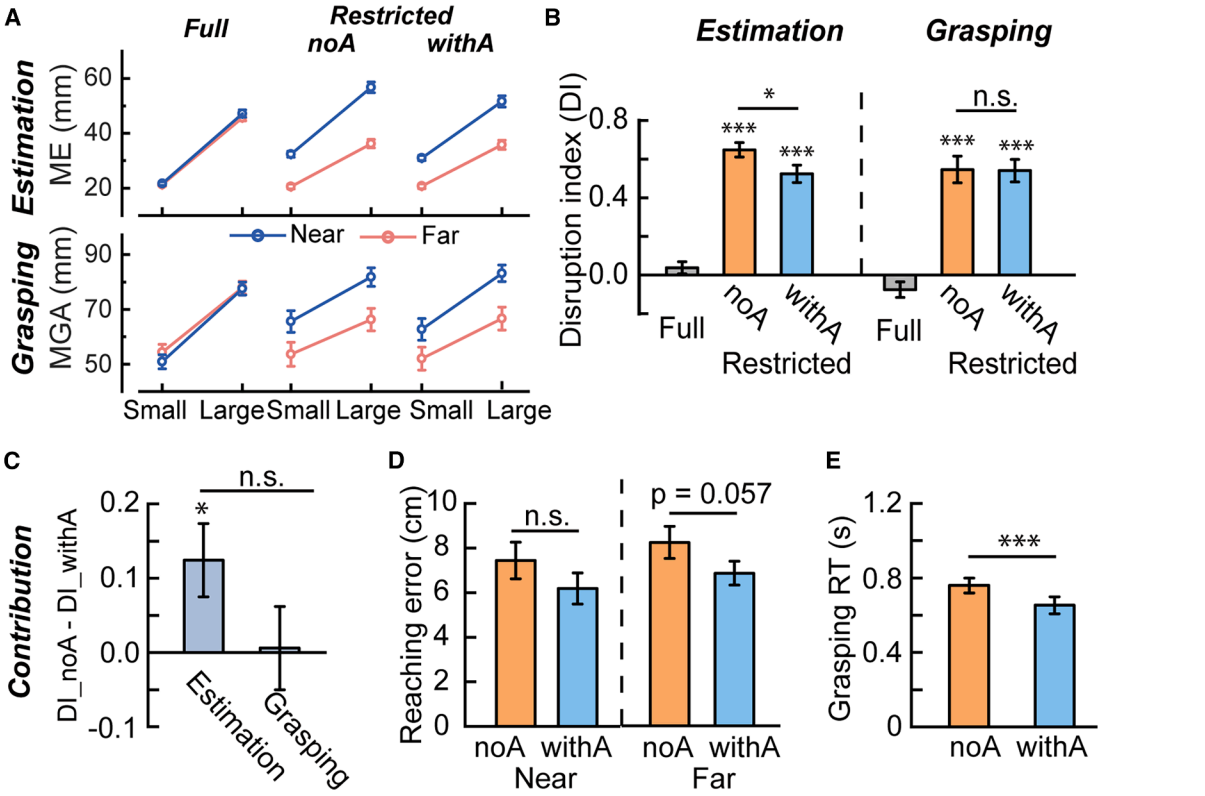


Figure 3. Results of size constancy for experiment 2 when sound was played right before the presentation of the visual object (delayed presentation)

(A) Top: manual size estimates (ME) of the perceived size of the small and large spheres at the near or far distances under three conditions: full viewing (*Full*), restricted viewing without auditory cue (*Restricted_noA*), and restricted viewing with auditory cue (*Restricted_withA*). Bottom: maximum grip aperture (MGA) during grasping in all conditions.

(B) The normalized size constancy disruption indices of the three viewing conditions for both tasks.

(C) Contribution of auditory cues to size constancy in manual size estimation and grasping, which was defined as $DI_{noA} - DI_{withA}$.

(D) The reaching error was defined as the difference between location at the end of the reach and the location of the target. The reaching distance was defined as the distance from the midpoint between the index finger and thumb to the participant on the table surface when the velocity dropped below 10% of the peak velocity.[21]

(E) The reaction time on grasping trials when auditory cue was provided (i.e., withA) or not provided (i.e., noA). Asterisks (*) and (***) indicate that the value is significantly different from zero or the values in two conditions are significantly different at $p < 0.05$ and $p < 0.001$, respectively. Error bars show the standard error of mean.

The "n.s." indicates no significant difference between two conditions.

$p < 0.001$, $\eta_p^2 = 0.775$) (Figure 3A and 3B). The contribution of auditory cues to estimation was significantly greater than 0 ($t_{(25)} = 2.516$, $p = 0.019$, Cohen's $d = 0.493$, two-tailed) suggesting that the auditory information makes significant contribution to perceptual size constancy. In contrast, the contribution of auditory information to grasping was not ($t_{(25)} = 0.105$, $p = 0.918$, Cohen's $d = 0.021$, two-tailed) significant suggesting that auditory information has no influence whatsoever on size constancy in grasping. The difference between the contribution to the two tasks were not significant ($t_{(25)} = 1.534$, $p = 0.138$, Cohen's $d = 0.301$, two-tailed) (Figure 3C) suggesting that that the auditory information makes only a modest contribution to perceptual size constancy.

To examine the effect of the asynchronous presentation of the auditory and visual information, we directly compared the contribution of auditory cues to estimation and grasping in experiment 1 (simultaneous presentation) and experiment 2 (delayed presentation). A 2 × 2 mixed-design ANOVA with presentation condition (simultaneous vs. delayed) as a between-subject factor and task (estimation vs. grasping) as a within-subject factor on the contribution of auditory cue showed that the main effect of presentation condition was not significant (F (1, 46) = 0.51, $p = 0.479$, $\eta_p^2 = 0.011$) and the interaction between presentation condition and task was not significant either (F (1, 46) = 2.112, $p = 0.153$, $\eta_p^2 = 0.044$) suggesting that the results of the two experiments were generally consistent and size constancy mechanisms did not benefit from either synchronous or asynchronous presentation of the auditory and visual signals. The main effect of Task was significant (F (1, 46) = 9.014, $p = 0.004$, $\eta_p^2 = 0.164$) demonstrating as a higher contribution to estimation than to grasping.

Contribution of the auditory cue to reaching distances

The auditory cue also did not improve the reaching performance during grasping for the near distance or the far distance (withA

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

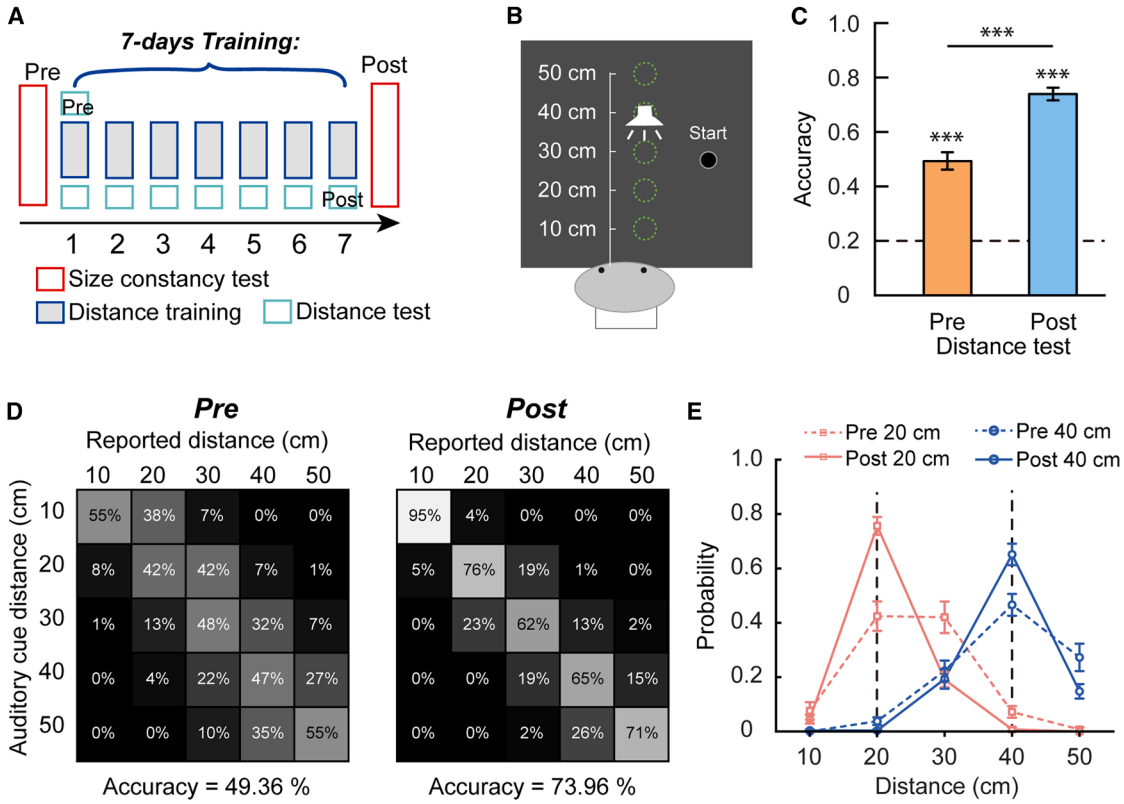


Figure 4. Training protocol and results of the auditory distance discrimination test

(A) Participants performed the size constancy test both before and after distance training. Moreover, before training and each day following the distance training, they completed a distance discrimination test without visual feedback to assess the effect of training.

(B) Participants were trained to judge the location of a speaker, which could be at one of five possible distances. During training, participants were allowed to see the speaker (i.e., learning with visual feedback), but during testing, they could no longer see the speaker.

(C) Accuracy for distance discrimination on the first (i.e., Pre) and last day (i.e., Post). The chance level was set at 0.2.

(D) The confusion matrices on the first (i.e., Pre) and last day (i.e., Post). The number in each cell indicates the percentage of reporting the auditory cue distance as each of the five distances.

(E) A graphic demonstration of the probability of judging the 20 cm (Near) and 40 cm (Far) as other distances before and after training. Asterisks (***) indicate that the accuracy was significantly larger than the chance level 20% at $p < 0.001$. Error bars show the standard error of mean.

vs. noA: Near, $t_{(25)} = 1.657$, $p = 0.110$, Cohen's $d = 0.325$, two-tailed; Far: $t_{(25)} = 1.999$, $p = 0.057$, Cohen's $d = 0.392$, two-tailed) suggesting that the auditory cue had no contribution to the transport component of grasping (Figure 3D).

Contribution of the auditory cue to reaction time in grasping

In contrast to experiment 1, in which the addition of auditory cue (*Restricted_noA* vs. *Restricted_withA*) increased the reaction time for grasping when auditory cue and visual images were presented simultaneously, the analyses of the reaction time in experiment 2 showed an opposite pattern: the addition of auditory cue significantly reduced the reaction time for grasping (mean ± std: *Restricted_noA*, 764 ± 183 ms, *Restricted_withA*: 633 ± 211 ms; main effect of Cue condition: $F_{(2, 50)} = 31.298$, $p < 0.001$, $\eta_p^2 = 0.556$, post-hoc comparison between *Restricted_noA* and *Restricted_withA*, $t_{(25)} = 4.816$, $p < 0.001$) (Figure 3E). This was probably due to the fact that participants had already extracted any distance information from the auditory cue before they viewed the glowing sphere. Nonetheless, the auditory information did not help them scale their grip aperture.

Training protocol and results of the auditory distance discrimination test

Effects of training on distance discrimination

Another possibility of the absence of size constancy in grasping is simply that participants had only limited experience with the auditory cue during grasping. In experiment 3, we trained 22 of the participants of experiment 2 in auditory distance discrimination over seven days (Figure 4A) and tested their perception and grasping after training. During training, a burst of white noise, identical to that used in the size constancy tasks, was played from five different locations (10, 20, 30, 40, and 50 cm; Figure 4B). Participants listened to the sound while visually observing the speaker, allowing them to use visual feedback to calibrate their distance judgments when pointing to the target. They were also asked to complete an auditory distance discrimination test (i.e., select from one of the five possible locations

where the sound came from) on the first day before training and each day after training, during which they pointed to the target without visual feedback.

After the seven days of training, participants' accuracy in discriminating distance based on auditory cues significantly improved (mean accuracy ±SD: day 1 before training, 0.494 ± 0.151; day 7, 0.740 ± 0.110; Accuracy: pre vs. post, t (21) = 6.545, $p < 0.001$, Cohen's $d = 1.395$, two-tailed, Figure 4C), suggesting that the auditory cues had become more reliable indicators of distance. Importantly, even on day 1, the accuracy was significantly larger than chance level (20%) (t (21) = 9.144, $p < 0.001$, Cohen's $d = 1.950$, two-tailed, Figure 4C) suggesting that the auditory cue could provide valid distance information even before training. This can also be seen from Figure 4D, which shows the confusion matrices indicating the percentage of reporting the auditory cue distance at each of the five distances, and Figure 4E which shows the probability of reporting the 20 cm (Near) and 40 cm (Far) as other distances. These results again show that the null contribution of auditory distance cues to grasping observed in experiments 1 and 2 could not be attributed to the possibility that the distance information provided by the auditory cue was simply invalid.

Results of size constancy in estimation and grasping after distance training in the restricted-noA and the restricted-withA condition

We then tested whether training enhanced the contribution of the auditory information to size constancy and whether both tasks could equally benefit from the training. Unsurprisingly, we observed no changes in size-distance integration under the *restricted-noA condition* for either task after training (Figure 5A; DIpre vs. DIpost: Estimation, $t_{(21)} = -0.384$, $p = 0.705$, Cohen's $d = -0.082$, two-tailed; Grasping, $t_{(21)} = -0.824$, $p = 0.419$, Cohen's $d = -0.176$, two-tailed).

In contrast, in the *restricted-withA condition*, size constancy improved for both perception and action (Figure 5B; DIpre vs. DIpost: Estimation, $t_{(21)} = 3.776$, $p = 0.001$, Cohen's $d = 0.805$, two-tailed; Grasping, $t_{(21)} = 2.304$, $p = 0.032$, Cohen's $d = 0.491$, two-tailed). The contribution of auditory cues, as defined by DI_{noA}-DI_{withA}, was significantly increased after training for both tasks (Estimation: $t_{(21)} = 3.204$, $p = 0.004$, Cohen's $d = 0.683$, two-tailed; Grasping: $t_{(21)} = 2.537$, $p = 0.019$, Cohen's $d = 0.541$, two-tailed, Figure 5C). Moreover, the size of this learning effect defined as the improvement in the contribution of auditory cues after training did not significantly differ between the two tasks ($t_{(21)} = -0.575$, $p = 0.572$, Cohen's $d = -0.123$, two-tailed, Figure 5D). Notably, because no learning was observed in the *restricted-noA condition*, the learning effect could not be attributed to multiple repetitions of the size constancy test or to discovering the underlying statistical structure of the experiment (i.e., that the 2.5 and 5 cm object at the 20 and 40 cm distance was presented more frequently than any other combination of size and distance). Taken together, these results suggest that, after extensive training, auditory information became a more reliable distance cue and improved size constancy processes for both perception and action, although size constancy in both tasks was never fully restored.

In addition to improvement in size constancy in grasping, improvement in reaching movements during grasping were also observed, as shown in Figure 5E for the Restricted_withA condition. After training, the reaching error significantly decreased in the *restricted-withA condition* compared to before training at both distances (Near: $t_{(21)} = 3.962$, $p < 0.001$, Cohen's $d = 0.845$, two-tailed; Far: $t_{(21)} = 2.644$, $p = 0.015$, Cohen's $d = 0.564$, two-tailed), but the reaching error decreased only for the near distance in the restricted-noA condition (Near: $t_{(21)} = 2.925$, $p = 0.008$, Cohen's $d = 0.624$, two-tailed; Far: $t_{(21)} = -0.545$, $p = 0.592$, Cohen's $d = -0.116$, two-tailed). Moreover, after training, the error was significantly smaller in the restricted_withA condition than in the restricted_noA condition (Near: $t_{(21)} = 3.327$, $p = 0.003$, Cohen's $d = 0.709$, two-tailed; Far: $t_{(21)} = 5.127$, $p < 0.001$, Cohen's $d = 1.093$, two-tailed).

After training, the reaction time of grasping was significantly different between noA and withA conditions (Restricted_noA: 759 ± 208 ms, Restricted_withA: 653 ± 236 ms; main effect of Cue: ($F_{(1, 21)} = 10.06$, $p = 0.005$, $\eta_p^2 = 0.324$; the interaction between Cue and Test or Task was not significant, all $p > 0.329$). However, when we compared the reaction time before and after training, the main effect of Test (pre vs. post) was not significant ($F_{(1, 21)} = 0.109$, $p = 0.754$, $\eta_p^2 = 0.005$). The reaction time data before and after training with auditory cue or without auditory cue were shown in Figure 5F.

Summary of the contribution of various cues to size constancy in estimation and grasping

Thus far, we have shown that, unlike proprioception, which restored size constancy for grasping completely, the auditory distance cue made no contribution to size constancy in grasping—at least before training. Proprioception and auditory distance cues had only a modest effect on size constancy for perceptual judgments. This was the case for both simultaneous presentation of sound and visual target (experiment 1) and delayed presentation of the visual target after the end of the sound (experiment 2).

In Figure 6A, we summarize the contribution of various distance cues, including proprioception,[15] sematic distance knowledge,[20] and auditory cue, where visual target was presented simultaneously (i.e., auditory [simultaneous], experiment 1) or delayed (i.e., auditory [delayed_pre], experiment 2), or delayed after training (i.e., auditory [delayed_post], experiment 3 after training), to size constancy in both grasping and perception. Because these cues had different reliability,[1–3,25,26] we also normalized the contribution to grasping relative to the contribution to estimation. This normalization process sets the contribution of various cues to estimation at one, and the contribution of a cue to grasping as a ratio of its contribution to estimation (Figure 6B). It can be clearly seen that among all the cues, the contribution of proprioception to grasping is unique because it was the only cue that showed more contribution to grasping than to estimation—and it almost completely restored size constancy in grasping. Moreover, the pattern of the contribution of auditory cue to estimation and grasping seems similar to that of the contribution of semantic knowledge. This is especially the case for auditory cues after training (auditory (delayed_post); Figure 6B, red and purple lines almost overlapped). These findings imply that the integration of proprioception and auditory cues with diminished visual input engage quite different mechanisms during the programming of grasping. The contribution of

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

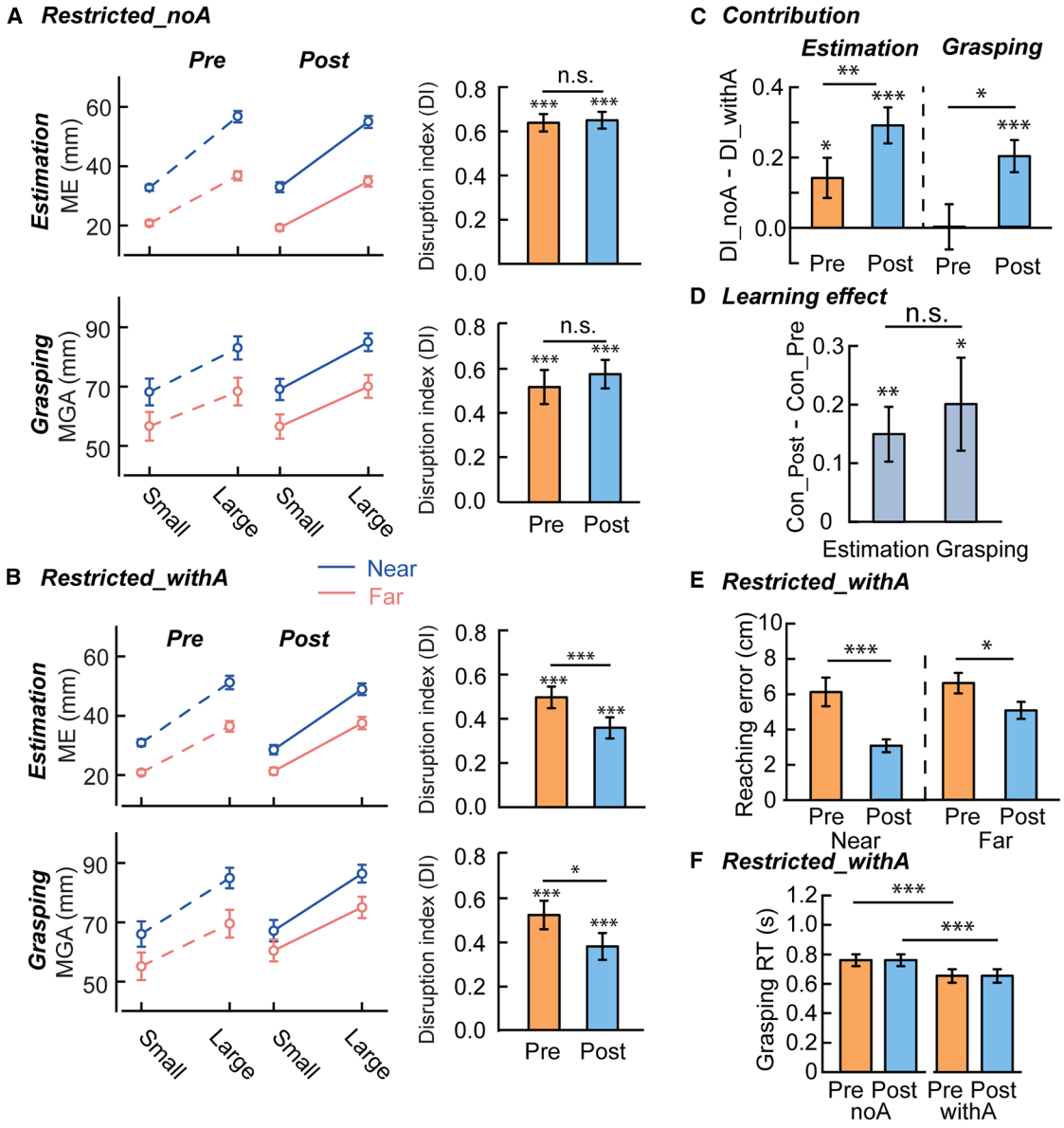


Figure 5. Results of size constancy in estimation and grasping after distance training in the restricted-noA and the restricted-withA condition

(A) Results in the in the *restricted-noA* condition. Left: manual estimations (ME) and maximum grip apertures (MGA) of small and large spheres at near and far distances before and after distance discrimination training. Right: size constancy disruption index (DI).

(B) The same as (A) but in the *Restricted_withA* condition.

(C) Contribution of auditory cues to size constancy, which was defined as DI_{noA}-DI_{withA} in estimation and in grasping.

(D) Effects of distance learning in the restricted withA condition, defined as the difference in the contribution of the auditory cue before and after training (i.e., Con_Post – Con_Pre) on size constancy for estimation and grasping.

(E) Reaching errors, which were defined as the difference between reach distance and target distance, at the two target distances in the Restricted_withA condition before and learning.

(F) The grasping reaction time in the Restricted_noA and Restricted_withA conditions before and after training. Note: the results before training were the results of experiment 2. Asterisks (*), (**), and (***) indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. Error bars show the standard error of mean.

The "n.s." indicates no significant difference between two conditions.

auditory cue to size constancy in grasping is more like a cognitive process similar to the contribution of semantic knowledge, whereas the contribution of proprioception may be an implicit and automatic process at the sensorimotor level.

DISCUSSION

One remarkable ability of our brain is multisensory integration: combining information from multiple sensory modalities to

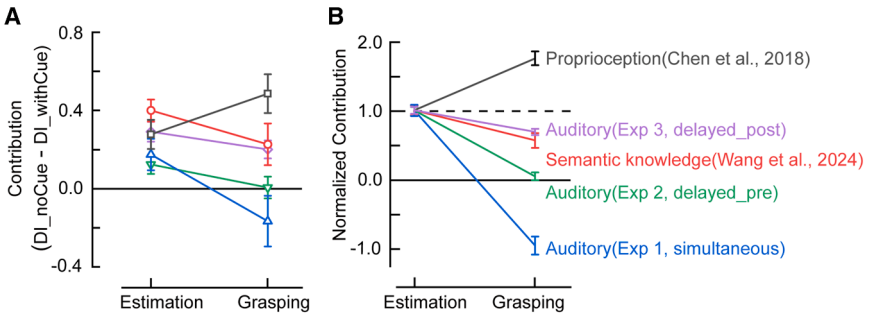


Figure 6. Summary of the contribution of various cues to size constancy in estimation and grasping

(A) Raw contribution values were calculated based on the difference in size constancy disruption index between noCue and withCue conditions (i.e., DI_noCue-DI_withCue).

(B) The contribution values of grasping were normalized relative to the contribution to estimation.

control behavior. Importantly, when information from one sensory source is compromised, information from another sensory modality can be reweighted and used to compensate for the loss. In a previous study,[15] we showed that the weighting of different sources of sensory information varies as a function of the control system being used which suggests that not only on the reliability of the information itself,[6,8] but also on the task, and thus the "consumer" of that information (i.e., the action system or the perceptual system) determines the weighting of cues. Nevertheless, it remained unclear whether the impact of task demands on the weighting of multisensory integration is confined to proprioception or also occurs with other sensory signals. To address this question, we used the same design as in our previous study with proprioceptive distance cues, but this time examined the contribution of auditory distance cues to size constancy for both grasping and perceptual judgments of size. In sharp contrast to what we had found earlier with proprioception, auditory distance cues had no effect on size constancy in grasping in the absence of reliable visual cues to distance. Instead, participants appeared to rely mainly on the retinal image size of the glowing sphere during grasping and this was the case no matter whether the glowing sphere was presented simultaneously with the auditory cue (experiment 1) or immediately afterward (experiment 2). However, the same auditory distance cue reliably improved the size constancy in perception (experiments 1 and 2) and the explicit distance judgment test (experiment 3) suggesting that the null effect of auditory cue to grasping could not be simply attributed to the poor auditory distance cue.

A small but reliable improvement in size constancy in perceptual judgments of size also occurred when we provided proprioceptive and verbal semantic information about the distance of the target sphere in our earlier experiments.[15,20] In normal circumstances, our perception of size, particularly with unfamiliar objects, depends almost exclusively on integrating retinal image size with visual distance cues. (With familiar objects, of course, we can use memory—and we can even use the retinal image size of familiar objects as a distance cue.) It seems likely that the neural circuits mediating the direct perception of size are not able to make use distance cues from other modalities, such as proprioception and audition, for perceiving the size of objects in the same seamless way that they make use of visual distance cues. One possibility is that when visual distance cues are severely restricted and cues from other modalities or even semantic information about distance are all that is available, participants resort to using cognitive operations to work out the size of the object perceptually. Future studies could test this possibility by examining the time course of the contribution auditory cues to size constancy in perception with electroencephalography (EEG)[27] or magnetoencephalography (MEG). If a cognitive operation is involved, then the modulation of auditory cues on size perception should appear at later stages of visual processing.

So, what happens with grasping? The provision of semantic information or auditory distance cues, even after extensive training, still results in only modest improvement in size constancy of grasping, akin to what we observed with perceptual reports of size. In contrast, proprioceptive cues to distance restore near-perfect size constancy in grasping, even without explicit training. The reason for proprioception being the "odd man out" could be due to the fact that proprioception works only in peripersonal space, the very space in which grasping normally operates. Moreover, the execution of grasping unavoidably recruits the proprioception system, and proprioceptive cues are encoded in somatosensory and premotor cortex,[28] which not only play an essential role in grasping,[29,30] but are interconnected with visuomotor areas in posterior parietal and premotor cortex. In contrast, the grasping system does not often use auditory or semantic cues to predict the distance and size of the object to be grasped, and certainly not in the same automatic fashion as it uses proprioception especially in the peripersonal space.

We propose then that the integration of distance cues and retinal image size can take place in two ways: one involves implicit and automatic integration at the sensory level; the other requires a higher order cognitive process. A prototypical example of the former is what happens with full viewing of objects where visual distance cues and retinal image size are seamlessly integrated into a direct perception of the real-world size of objects. This direct experience is phenomenally similar to what happens with the classic McGurk effect: one really does hear something quite different when listening to the same speech sound accompanied by two different videos of someone speaking. Automatic integration of distance and retinal image size, we would argue, was also evident in our earlier experiment in which proprioceptive distance cues completely restored size constancy in grasping in the absence of visual distance cues.[15] In contrast to this more automatic sensory integration, the use of semantic information in size constancy in the same paradigm may engage more cognitive processes, in which one may first form a perceptual judgment of distance and then combines that information with the perceived size of the retinal image to make an explicit

judgment about the object's real size and to scale one's grip aperture during grasping. The use of auditory distance cues in size constancy in the simultaneous presentation may also engage more cognitive processes given that the contribution of auditory cues is similar to semantic cues in that both are far from perfect. A cognitive process is even more likely for the delayed presentation given that sound and visual images are temporally separated. As addressed previously, further studies can test this possibility with EEG or MEG.

Unlike our account, Linton[31,32] put forth a purely cognitive explanation for size constancy, proposing that size-distance scaling is solely the result of our subjective knowledge about changes in viewing distance. Our work suggests that is not necessarily the case. Although the contribution of auditory distance cues (and direct semantic information about distance) to size constancy is likely to be implemented via subjective/cognitive inference, this is not necessarily true for proprioception. As our earlier experiment showed, proprioceptive cues restore almost perfect size constancy in grasping, but the same cues resulted in only a limited improvement in size constancy in perceptual judgments of size. By extension, one might argue that visual distance cues, which support perfect size constancy in both grasping *and* perceptual judgments of size depend on a similar automatic integration of distance cues and retinal image size—but not a subjective/cognitive process.

It is important to note that the absence of any contribution of the auditory distance cues to size constancy in grip scaling and the tiny contribution to reaching before training could not be simply attributed to poor auditory distance information. First, exactly the same distance information significantly improved size constancy in perception (as measured by manual estimation). Second, the distance discrimination results of experiment 3 showed that the auditory distance cue provides reliable distance information for the distance judgment before training. In an earlier study, we found that, while proprioception makes a greater contribution to size constancy for action than it does for perception, audition makes a larger contribution to size constancy for perception than it does for action. Taken together, these results provide additional evidence for the two-visual-stream theory (vision-for-perception and vision-for-action,[16,17]) based on clear but opposite effects of the weighting of auditory and proprioceptive cues for perceptual judgments and the control of actions—and provide insights into the nature of multisensory integration in the two visual systems.

It should also be noted that after training on distance discrimination, improvements in the accuracy of both reaching and size constancy in grip aperture were observed, suggesting that training can facilitate the weighting of a specific sensory modality in multisensory integration. Previous studies showed that severe visual loss leads to enhancement of auditory spatial localization, especially for signals located in peripheral space.[33,34] Our findings indicate that size perception and grip aperture during grasping might also become more accurate.

Taken together, we found that before auditory distance discrimination training, auditory cues contributed to perception, but not to grasping, and may be integrated with retinal image size via a higher order cognitive process. Training improved the contribution of auditory cues such that people can learn to integrate auditory distance information with restricted visual information about the goal object to improve size constancy. The findings of our current study and the two earlier ones[15,20] suggest that proprioception and vision are the primary distance cues for integration with retinal image size for computing size constancy in grasping. Moreover, the integration of distance cues from different modalities and the retinal size may take place in two ways: one involves implicit and automatic integration at the sensory level; the other requires a higher-order cognitive process similar to the integration of semantic cues.

Limitations of the study

Here, we used white noise, a neutral stimulus, as an auditory cue. One might argue that the reason for the null contribution of auditory cue to grasping observed in experiments 1 and 2 was due to the nature of the auditory stimulus we used. That is, if we used a more realistic sound, such as the sound generated by dropping an object on a table at different distances, then the outcome might have been different. Of course, the sounds that objects make when placed on a surface can reveal something about their size and composition. In fact, some studies have shown that auditory information can be used to infer the size of goal objects and modulate a person's grip aperture during grasping, even when vision is available.[35] Although the distance and direction of the object might be signaled more accurately by such a sound, it is still possible that the information might not be automatically integrated with retinal image size for preserving size constancy in grasping. Future research is needed to test this. The current study represents an initial step, using an unbiased scenario in which a burst of white noise provides distance information. Another reason for the null contribution to grasping could be that auditory cues are not an effective cue for distance in peripersonal space—and thus are poor cues for automatic size constancy in action. At greater distances, well beyond peripersonal space, it has been shown that the perceived size of an object can be modulated by audiovisual asynchronies.[36] Future research can test this possibility by employing a longer distance beyond the peripersonal space.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Juan Chen (juanchen@m.scnu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: Data have been deposited at https://doi.org/10.17632/bccfxy9t37.1.
- Code: All original code has been deposited at https://doi.org/10.17632/bccfxy9t37.1 and is publicly available as of the date of publication.
- Additional information: Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This research was supported by the National Science and Technology Innovation 2030 Major Program (STI2030-Major Projects 2022ZD0204802 to J.C. and

2022ZD0204804 to X.Z.), the China Scholarship Council and the National Natural Science Foundation of China (no. 31970981 and no. 31800908) to J.C. It was also supported by the Young Faculty Research Cultivation Program of South China Normal University (no. 23KJ27) to J.G., and the Research Center for Brain Cognition and Human Development, Guangdong, China (no. 2024B0303390003) to J.G. and J.C.

AUTHOR CONTRIBUTIONS

C.Z., M.A.G., and J.C. designed the study. C.Z. and G.W. performed the research. C.Z., G.W., J.G., and J.C. analyzed the data. C.Z. and J.C. wrote the draft of the manuscript. X.Z., I.S., and M.A.G. revised the manuscript. All authors have read and approved the final version of the manuscript and agree with the order of presentation of the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Apparatus and stimuli
  - Procedure and design
  - Assessment of size constancy
  - Auditory distance discrimination training
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.113341.

Received: September 10, 2024
Revised: January 1, 2025
Accepted: August 8, 2025

REFERENCES

1. Kolarik, A.J., Moore, B.C.J., Zahorik, P., Cirstea, S., and Pardhan, S. (2016). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Atten. Percept. Psychophys. *78*, 373–395. https://doi.org/10.3758/s13414-015-1015-1.
2. Zahorik, P., Brungart, D.S., and Bronkhorst, A.W. (2005). Auditory distance perception in humans: A summary of past and present research. Acta Acust. united Acust *91*, 409–420. https://doi.org/10.3758/s13414-015-1015-1.
3. Wilson, E.T., Wong, J., and Gribble, P.L. (2010). Mapping Proprioception across a 2D Horizontal Workspace. PLoS One *5*, e11851. https://doi.org/10.1371/journal.pone.0011851.
4. Capaday, C., Darling, W.G., Stanek, K., and Van Vreeswijk, C. (2013). Pointing to oneself: active versus passive proprioception revisited and implications for internal models of motor system function. Exp. Brain Res. *229*, 171–180. https://doi.org/10.1007/s00221-013-3603-4.
5. Fuentes, C.T., and Bastian, A.J. (2010). Where is your arm? Variations in proprioception across space and tasks. J. Neurophysiol. *103*, 164–171.
6. Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429–433. https://doi.org/10.1038/415429a.
7. Ernst, M.O., and Bülthoff, H.H. (2004). Merging the senses into a robust percept. Trends Cognit. Sci. *8*, 162–169. https://doi.org/10.1016/j.tics.2004.02.002.
8. Fetsch, C.R., Pouget, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. Nat. Neurosci. *15*, 146–154. https://doi.org/10.1038/nn.2983.
9. Battaglia, P.W., Di Luca, M., Ernst, M.O., Schrater, P.R., Machulla, T., and Kersten, D. (2010). Within- and Cross-Modal Distance Information Disambiguate Visual Size-Change Perception. PLoS Comput. Biol. *6*, e1000697. https://doi.org/10.1371/journal.pcbi.1000697.
10. Sperandio, I., Kaderali, S., Chouinard, P.A., Frey, J., and Goodale, M.A. (2013). Perceived size change induced by nonvisual signals in darkness: the relative contribution of vergence and proprioception. J. Neurosci. *33*, 16915–16923. https://doi.org/10.1523/JNEUROSCI.0977-13.2013.
11. Sperandio, I., and Chouinard, P.A. (2015). The mechanisms of size constancy. Multisens. Res. *28*, 253–283. https://doi.org/10.1163/22134808-00002483.
12. Blakemore, C., Garner, E.T., and Sweet, J.A. (1972). The site of size constancy. Perception *1*, 111–119. https://doi.org/10.1068/p010111.
13. Boring, E.G. (1940). Size Constancy and Emmert's Law. Am. J. Psychol. *53*, 293–295. https://doi.org/10.2307/1417427.
14. Whitwell, R.L., Sperandio, I., Buckingham, G., Chouinard, P.A., and Goodale, M.A. (2020). Grip constancy but not perceptual size constancy survives lesions of early visual cortex. Curr. Biol. *30*, 3680–3686.e5. https://doi.org/10.1016/j.cub.2020.07.026.
15. Chen, J., Sperandio, I., and Goodale, M.A. (2018). Proprioceptive distance cues restore perfect size constancy in grasping, but not perception, when vision is limited. Curr. Biol. *28*, 927–932.e4. https://doi.org/10.1016/j.cub.2018.01.076.
16. Goodale, M.A., Milner, A.D., Jakobson, L.S., and Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. Nature *349*, 154–156. https://doi.org/10.1038/349154a0.
17. Goodale, M.A., and Milner, A.D. (1992). Separate visual pathways for perception and action. Trends Neurosci. *15*, 20–25.
18. Leone, L.M., and McCourt, M.E. (2015). Dissociation of perception and action in audiovisual multisensory integration. Eur. J. Neurosci. *42*, 2915–2922.
19. Kentridge, R.W. (2018). Vision: Non-illusory Evidence for Distinct Visual Pathways for Perception and Action. Curr. Biol. *28*, R264–R266.
20. Wang, G., Zheng, C., Wu, X., Deng, Z., Sperandio, I., Goodale, M.A., and Chen, J. (2024). The contribution of semantic distance knowledge to size constancy in perception and grasping when visual cues are limited. Neuropsychologia *196*, 108838. https://doi.org/10.1016/j.neuropsychologia.2024.108838.
21. Fan, A.W.-Y., Guo, L.L., Frost, A., Whitwell, R.L., Niemeier, M., and Cant, J.S. (2021). Grasping of real-world objects is not biased by ensemble perception. Front. Psychol. *12*, 597691.
22. Turella, L., and Lingnau, A. (2014). Neural correlates of grasping. Front. Hum. Neurosci. *8*, 686. https://doi.org/10.3389/fnhum.2014.00686.
23. Begliomini, C., De Sanctis, T., Marangon, M., Tarantino, V., Sartori, L., Miotto, D., Motta, R., Stramare, R., and Castiello, U. (2014). An investigation of the neural circuits underlying reaching and reach-to-grasp movements: from planning to execution. Front. Hum. Neurosci. *8*, 676. https://doi.org/10.3389/fnhum.2014.00676.
24. Whitwell, R.L., Ganel, T., Byrne, C.M., and Goodale, M.A. (2015). Real-time vision, tactile cues, and visual form agnosia: removing haptic feedback from a "natural" grasping task induces pantomime-like grasps. Front. Hum. Neurosci. *9*, 216.
25. Rossetti, Y., Desmurget, M., and Prablanc, C. (1995). Vectorial coding of movement: vision, proprioception, or both? J. Neurophysiol. *74*, 457–463. https://doi.org/10.1152/jn.1995.74.1.457.
26. Wang, C., Gao, J., Deng, Z., Zhang, Y., Zheng, C., Liu, X., Sperandio, I., and Chen, J. (2022). Extracurricular sports activities modify the

proprioceptive map in children aged 5–8 years. Sci. Rep. *12*, 9338. https://doi.org/10.1038/s41598-022-13565-8.

27. Chen, J., Sperandio, I., Henry, M.J., and Goodale, M.A. (2019). Changing the real viewing distance reveals the temporal evolution of size constancy in visual cortex. Current Biology *29*, 2237–2243.
28. Graziano, M.S. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. Proc. Natl. Acad. Sci. *96*, 10418–10421. https://doi.org/10.1073/pnas.96.18.10418.
29. Culham, J.C., Danckert, S.L., DeSouza, J.F.X., Gati, J.S., Menon, R.S., and Goodale, M.A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. Exp. Brain Res. *153*, 180–189. https://doi.org/10.1007/s00221-003-1591-5.
30. Castiello, U. (2005). The neuroscience of grasping. Nat. Rev. Neurosci. *6*, 726–736. https://doi.org/10.1038/nrn1744.
31. Linton, P. (2020). Does vision extract absolute distance from vergence? Atten. Percept. Psychophys. *82*, 3176–3195.
32. Linton, P. (2021). Does vergence affect perceived size? Vision *5*, 33. https://doi.org/10.3390/vision5030033.
33. Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. Nature *395*, 278–280.
34. Simon, H.J., Divenyi, P.L., and Lotze, A. (2002). Lateralization of Narrow-Band Noise by Blind and Sighted Listeners. Perception *31*, 855–873. https://doi.org/10.1068/p3338.
35. Sedda, A., Monaco, S., Bottini, G., and Goodale, M.A. (2011). Integration of visual and auditory information for hand actions: preliminary evidence for the contribution of natural sounds to grasping. Exp. Brain Res. *209*, 365–374. https://doi.org/10.1007/s00221-011-2559-5.
36. Jaekl, P., Soto-Faraco, S., and Harris, L.R. (2012). Perceived size change induced by audiovisual temporal delays. Exp. Brain Res. *216*, 457–462. https://doi.org/10.1007/s00221-011-2948-9.
37. Litovsky, R.Y., and Clifton, R.K. (1992). Use of sound-pressure level in auditory distance discrimination by 6-month-old infants and adults. J. Acoust. Soc. Am. *92*, 794–802. https://doi.org/10.1121/1.403949.
38. Kopčo, N., and Shinn-Cunningham, B.G. (2011). Effect of stimulus spectrum on distance perception for nearby sources. J. Acoust. Soc. Am. *130*, 1530–1541. https://doi.org/10.1121/1.3613705.
39. Aglioti, S., DeSouza, J.F., and Goodale, M.A. (1995). Size-contrast illusions deceive the eye but not the hand. Curr. Biol. *5*, 679–685. https://doi.org/10.1016/s0960-9822(95)00133-3.
40. Chen, J., Sperandio, I., and Goodale, M.A. (2015). Differences in the effects of crowding on size perception and grip scaling in densely cluttered 3-D scenes. Psychol. Sci. *26*, 58–69. https://doi.org/10.1177/0956797614556776.
41. Goodale, M.A., and Milner, A. (2013). Sight Unseen: An Exploration of Conscious and Unconscious Vision (Oxford University Press). https://doi.org/10.1093/acprof:oso/9780199596966.001.0001.
42. Chen, J., Jayawardena, S., and Goodale, M.A. (2015). The effects of shape crowding on grasping. J. Vis. *15*, 6. https://doi.org/10.1167/15.3.6.

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data (behavioral)	Mendeley repository	https://data.mendeley.com/preview/bccfxy9t37?a=f8d41942-f1a4-4a6d-8691-9a6793086e2e

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Twenty-two participants (5 males and 17 females; age: 20.40 ± 1.60) took part in Experiment 1, in which the auditory cue (a sound burst) and the visual stimulus (a glowing sphere) were presented simultaneously. Another twenty-six participants (10 males and 16 females; age: 21.0 ± 2.77) took part in Experiment 2, which had a similar design to Experiment 1, but the visual stimulus was presented immediately after the end of the presentation of the auditory cue. In both experiments, participants made perceptual judgements of size and also grasped the sphere. Twenty-two of the participants from Experiment 2 (7 males and 15 females; age: 21.18 ± 2.87) also participated in Experiment 3, in which they received explicit training on the discrimination of the auditory distance cue and then performed the same size constancy tests used in Experiment 2. The sample size was comparable or larger than previous studies on similar research question.[15,20] All participants reported normal hearing and normal or corrected-to-normal vision. None of the participants had a history of neurological or psychiatric disorders. All participants were naive to the purpose of the study. The experiments were conducted following the ethical standards laid down in the 1964 Declaration of Helsinki, as well as the ethical guidelines laid down by the South China Normal University (Study Approval Number: 2019-4-031, Human Research Ethics Committee for Non-Clinical Studies, The School of Psychology, South China Normal University). All participants gave informed consent to participate in the study before beginning an experimental session. All the participants were recruited from South China Normal University and were all of Asian (Chinese) ethnicity.

METHOD DETAILS

Apparatus and stimuli

To measure size constancy, spheres of two sizes (small: 2.5 cm, large: 5 cm in diameter) were placed at two distances (near: 20 cm, far: 40 cm). The near-small and far-large stimuli generated the same image size on the retina ($\sim 7^{\circ}$ of visual angle) (Figure 1A). To increase the variability of size and distance information and to enhance participant engagement, we included three additional spheres (1.25 cm, 3.75 cm, and 6.25 cm in diameter) and a third distance (30 cm; Figure 1B). However, only four conditions were considered as critical for the analysis of size constancy (i.e., spheres with a diameter of 2.5 cm or 5 cm placed at 20 cm or 40 cm, see Figure 1A).

All the spheres were 3D-printed, hollow white spheres painted with glow-in-the-dark paint. Each of the five spheres was mounted on a pedestal using a movable stand with different heights to ensure that the center of each sphere was consistently aligned across all size conditions. The pedestal had the same diameter in all cases. The stand and pedestal were black, so the participants could not see them in the dark. A start button was located 30 cm from the participant and 14.5 cm to the right of the middle target location (30 cm). The distances from the start button to the Near (20 cm) and Far (40 cm) target locations were identical to minimize the influence of movement distance on perceptual judgements and grasping (Figure 1B).

To provide auditory distance cues, a small speaker was attached to the pedestal with its mouth passing through the rod of the pedestal so that the sound is right beneath the center of the object (Figure 1D). The auditory distance cue consisted of a 100-ms silent segment followed by a 900-ms white noise segment. The intensity of the white noise was set at 80 dB sound pressure level. The same auditory stimulus was used in the auditory distance discrimination training and testing.

When the sound was presented at different distances, the overall level/intensity and interaural intensity and time differences as cues of the sound served as a distance cue.[1,37] Previous studies have shown that listeners' judgment of distance is more accurate for lateral sources than for frontal sources, due to the added benefit of interaural-level and interaural-time differences.[38] Therefore, we moved the stimulus locations (green circles in Figure 1B) to the side of the observer (i.e., aligned with the observer's right eye). This positioning also provided interaural intensity and time differences as cues to distance when the speaker was placed at different distances from the observer.

Participants wore liquid crystal goggles (PLATO goggles; Translucent Technologies, Toronto, ON, Canada) to control the availability of vision. In the full-viewing condition ("*Full*"), participants viewed the spheres binocularly with the room lights on. Under the restricted-viewing condition ("*Restricted-noA*" or "*Restricted-withA*"), the left lens of the goggles was always closed. A piece of black paper with a 1-mm hole was positioned behind the right lens of the goggles. When that right lens of goggles was opened,

Please cite this article in press as: Zheng et al., Contribution of auditory distance cues to size constancy in perception and grasping in restricted viewing, iScience (2025), https://doi.org/10.1016/j.isci.2025.113341

participants were able to see only the glowing sphere in complete darkness through the 1-mm hole with their right eye (Figure 1D). In this case, almost all the distance cues (pictorial cues, stereo, and vergence) were unavailable and participants had to rely entirely on retinal-image size when attempting to compute the real-world size of the object.[15,18] The position of the fingers during grasping and manual size estimation were recorded by infrared light emitting diodes (IREDs) attached to the inside of the tip of the right thumb and the outside of the tip of the index finger using an OPTOTRAK system (Northern Digital, Waterloo, ON, Canada) with a sampling frequency of 200 Hz.

Procedure and design

In all three experiments, participants performed the size constancy tests under the *Restricted-noA, and Restricted-withA* cue conditions. In Experiment 1, the full-viewing condition was not included because previous studies had demonstrated perfect size constancy under this condition. This was further confirmed in Experiments 2, where the full-viewing condition ("*Full*") was included (Figure 1D).

In Experiment 1, the sound was played simultaneously with the presentation of the sphere in the *Restricted-with A* condition. In contrast, in Experiments 2 and 3, the sound was played before the presentation of the sphere. In Experiment 3, participants underwent distance discrimination training before performing the size constancy tasks (Figure 4A). All participants in Experiment 3 had previously participated in Experiment 2, allowing a direct comparison between their size constancy performance before training (i.e., in Experiment 2, Pre) and after training (i.e., in Experiment 3, post).

Assessment of size constancy

Participants completed two size-constancy tasks, size estimation and grasping, to examine the contribution of auditory cues to perception and action, respectively. Before the experiment begun, participants were asked to sit in front of a black table with their chin on a chinrest. In the Full condition, participants viewed the target binocularly with the room lights on. In the two restricted viewing conditions, participants viewed the glowing target sphere monocularly through a 1 mm hole entirely in the dark. In these two conditions, they were also asked to adjust their head position before each testing session to ensure that they could see the near-large object through the 1-mm pinhole. They were then instructed to keep their head as still as possible to prevent changes in head position during testing. In case they could not see the whole sphere during the task, they were asked to adjust their head position again and redo the trial. The goggles were closed before the start of each trial.

Participants began each trial by pressing down a start button with their thumb and index finger pinched together. Then, the experimenter placed the stimulus at a specified distance, and pressed a button to open the goggles. In the *restricted-withA* condition of Experiment 1, a 1-s white noise, played three times with 100-ms silence intervals, was synchronized with the opening of the goggles while participants viewed the sphere. In other words, the auditory and visual information were presented simultaneously (i.e., simultaneous presentation). In the *restricted-withA* condition of Experiments 2 and 3, the white noise was played immediately before the opening of the goggles (i.e., delayed presentation). After the goggles were opened, participants performed the task.

In the size estimation task, participants were asked to indicate the perceived size of the target sphere by opening their thumb and index finger a matching amount. When participants reported being satisfied with their estimate, the OPTOTRAK was triggered to record the position of their fingers for 500 ms. After that, the experimenter placed the sphere into the hand of the participant so that they had the same tactile feedback about the size of the sphere as they did in the grasping task.

In the grasping task, participants were asked to grasp the sphere with their thumb and index finger as naturally and accurately as possible as soon as they could see the object. The positions of their fingers were recorded for 3 s after the goggles were opened. There were instances where participants reached out and opened their fingers wider but still failed to pick up the sphere, especially in the restricted-viewing conditions. In such cases, the experimenter would place the sphere into the participants' hand to ensure consistent haptic feedback across conditions. For both tasks, the goggles were turned off once participants released the start button to perform the task. Therefore, participants performed the tasks without visual feedback (i.e., open loop).

It is important to note that, although participants also moved their hand during the size estimation task, it was the maintained distance between the thumb and finger indicating the size of the sphere, rather than the movement of the hand, that was used as an independent variable. Previous studies have commonly used manual estimation and grasping tasks to examine potential differences between perception and action, respectively.[15,16,39,40] Patients with lesions in the lateral occipital cortex (ventral stream) or occipitoparietal cortex (dorsal stream) exhibit a strong dissociation between the manual estimation and grasping tasks. For example, patient D.F., who has a lesion in the lateral occipital cortex, cannot manually estimate the size of an object but continues to scale her grasp according to its size.[16] In contrast, patient R.V., who has bilateral lesions of the occipitoparietal cortex, is able to manually indicate the size of objects but cannot scale her grasp when attempting to pick them up.[41]

Theoretically, when perfect size constancy is maintained for both perception and action, the target will be perceived as the same size and grasped with the same maximum grip aperture, regardless of viewing distance[15,20] (Figure 1C, Perfect). When visual information about distance is limited, size constancy is disrupted, leading participants to rely more on retinal-image size for both perception and grasping. As a result, when an object is presented close to the observer, it will be perceived as larger and grasped with a larger grip aperture than when the same object is presented further away (Figure 1C, Disrupted). This was confirmed in our previous studies showing that, when distance information was largely compromised in the restricted-viewing condition, size constancy for both perception and action was disrupted.[15,20] Here, we tested whether providing auditory distance cues can contribute to size

constancy mechanisms and if the reliability of the auditory signals can change as a function of the consumer system, namely perception vs. action systems. To put it simply, we tested whether or not the gap in manual estimates and/or grip apertures between near and far distances would be reduced with the addition of auditory distance cues (Figure 1C, Partially restored).

The size constancy tasks consisted of six blocks in total: one block for each of the combinations between the three cue conditions (*Full, Restricted-noA, and Restricted-withA*) and the two tasks (*Estimation and Grasping*). In each block, the 2.5 m and 5 cm spheres were presented at the distance of 20 cm and 40 cm. Each size and distance combination had 8 repetitions for a total of 32 trials. Three additional spheres (1.25 cm, 3.75 cm, and 6.25 cm in diameter) were presented twice at 20 and 40 cm of viewing distance. Finally, all five spheres were presented once at the middle position, 30 cm away. The additional set of 11 trials was excluded from the analysis. The order of the blocks as well as the conditions within each block were randomized.

Auditory distance discrimination training

In Experiment 3, before administering the two size constancy tasks, participants were trained for seven days to discriminate distance from the auditory cues. Auditory distance discrimination tests were performed prior to training and after each training session to monitor the effects of training on their ability to discriminate distance.

During training, the room light was on. The goggles were closed at the beginning of each trial, and the participants rested their right index finger at the starting position. The experimenter then placed the pedestal with the attached speaker at one of the five testing distances (10 cm, 20 cm, 30 cm, 40 cm, or 50 cm) and played the auditory stimulus. The auditory cue was identical to that used in the size constancy tasks and was presented for 3 s. Then the goggles were opened while the pedestal with the speaker remained in place. At this point, participants used their right index finger to point to one of the five possible locations with five visual markers, to indicate where they believed the speaker was located. Therefore, it is a five-alternatives forced choice tasks. During the training phase, participants could see the pedestal with the speaker while pointing, allowing them to use visual feedback to calibrate their auditory perception. However, during the testing phase, the pedestal with the speaker was removed before the goggles were opened, and participants were still asked to perform the same five-alternatives-forced-choice pointing task. This phase measured how accurately participants could locate the source of the previously presented auditory cue without any visual information.

Participants completed 200 trials in total (40 repetitions for 5 distances) each day during training and performed 60 trials in total (12 repetitions for 5 distances) for both the pre-test and post-test of auditory localization without visual feedback. The order of trials for the five distances was randomized.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the size estimation task, participants used their thumb and index finger to match the perceived size of objects (manual estimate, ME). The finger positions were recorded after participants reported being satisfied with the matching. The Euclidean distance between the two fingers was used as the independent measure for the perception task. For the grasping task, participants' hand opened wider and wider during the approach to the sphere and reached a peak well before they made contact with the target object.[42] The maximum grip aperture (MGA), which is typically achieved well before participants make contact with the object and scales with object size,[30] was used as the dependent measure for grip scaling. In our experiments, we were able to record MGAs even in trials where participants failed to contact the object because the MGA was recorded before they closed down their fingers to contact the object. On average, the percentage of successful grasping trials was 31.1% in the Restricted_noA and 33.5% in the Restricted_withA conditions (withA vs. noA: $t_{(21)} = 0.717$, $p = 0.481$, Cohen's $d = 0.153$, two-tailed) in Experiment 1. In Experiment 2 success rate was 84.7% in the full viewing condition, 30.2% in the Restricted_noA, and 38.5% in the Restricted_withA (withA vs. noA: $t_{(25)} = 2.934$; $p = 0.007$, Cohen's $d = 0.575$, two-tailed). In Experiment 3, success rates were 45% in the Restricted_noA and 60.7% in the Restricted_withA (noA vs. withA: $t_{(21)} = 5.728$, $p < 0.001$, Cohen's $d = 1.221$, two-tailed). Note that even in the full-viewing condition, the percentage of successful grasping was not perfect, likely because the goggles were closed immediately following the release of the start button.

No participants were excluded from the study. In Experiment 1, the percentage of missing trials due to signal loss was 0 as we carefully checked the data during recording and recollected any trials with severe signal loss. In Experiment 2, 9 trials were excluded in total across all participants and all conditions, which corresponds to 1.08% of the total trials. In Experiment 3, 5 trials were excluded in total across all participants and all conditions, representing 0.71% of all trials.

Repeated measures ANOVAs with Distance and Size as main factors were performed on the mean MEs and MGAs to determine whether or not size constancy was maintained (i.e., non-significant main effect of Distance) or disrupted (i.e., significant main effect of Distance).

To evaluate the extent to which size constancy was disrupted by the removal of visual cues, we first calculated a size-constancy disruption index (DI), which captures the differences in ME or MGA between the near and far distances.[15,20] If size constancy were perfect, then the ME and MGA would remain constant regardless of viewing distance, resulting in a size disruption index of 0. When visual cues are compromised, participants might perceive an object at the near distance as larger than when it is at the far distance and/or grasp it with a larger grip aperture. In this case, size constancy would be disrupted and the index would be positive. It should be noted, however, that the slopes for MGAs as a function of object size are typically shallower than those for MEs,[15,20] such that a 1-mm difference in object size results in a smaller increase in MGA compared to a 1-mm difference in ME. Therefore, to directly

compare the results between manual estimation and grasping, each index for ME and MGA was divided by the averaged slope across distances as a function of physical size.

Mathematically, since the denominators for the slope calculations were identical for the two tasks (slope of estimation = $[(ME_{Large}-ME_{Small})/(Large-Small)]_{Averaged\ Across\ distances}$ and slope of grasping = $[(MGA_{Large}-MGA_{Small})/(Large-Small)]_{Averaged\ Across\ distances}$), the disruption index can simply be divided by $(ME_{Large}-ME_{Small})_{Averaged\ Across\ distances}$ for estimation, and by $(MGA_{Large}-MGA_{Small})_{Averaged\ Across\ distances}$ for grasping. Therefore, the DI for the two tasks were defined as follows:

$$DI_{estimation} = (ME_{near} - ME_{far}) / (ME_{Large}-ME_{Small})_{Averaged\ Across\ distances}$$

and

$$DI_{grasping} = (MGA_{near} - MGA_{far}) / (MGA_{Large}-MGA_{Small})_{Averaged\ Across\ distances.}$$

The DI was calculated for both sizes and was then averaged across sizes.

The contribution of auditory cues to size constancy in perception and grasping was quantified as the difference between the DI in the *Restricted-noA* condition and the DI in the *Restricted-withA* condition. The effect of training on size constancy was quantified as the difference in DI before and after training. All the analyses above were done for each participant. Paired sample t-tests were used to assess whether the contribution of auditory cues differed between the two tasks, while one-sample t-tests were used to determine if the contribution of auditory cues was significantly different from 0.

The reaction time of the grasping movement was defined as the interval between stimulus onset (i.e., goggles open) and the movement onset. The movement onset, was operationally defined as the first of 20 consecutive sample frames (100 ms) in which the velocity of the IRED attached to the index finger exceeded a threshold of 50 mm/s.[24,25] Repeated measures ANOVAs were performed with Cue condition, Size, and Distance as within-subject factors.

To quantify the contribution of auditory cues to reaching movements during grasping, the endpoint of each reach, which was defined as the distance from the midpoint between the index finger and thumb to the participant on the table surface when the velocity dropped below 10% of the peak velocity,[21,24]was extracted from the Optotrak data for every trial. Then, reaching errors, which were defined as the distance between reaching endpoint and target distance, were calculated in each condition for each participant. Paired t-tests were performed to compare the reaching error in the withA and noA conditions for each distance.

Finally, to evaluate the effect of training in Experiment 3, one sample t-test was conducted to analyze whether or not the accuracy was significantly different from the chance level 20%. Paired t-test was conducted to compare the accuracy before and after training.

The statistical information of all figures is provided in a supplementary excel file.