Changing the real viewing distance reveals the temporal evolution of size constancy in visual cortex

Juan Chen[1,2*], Irene Sperandio[3*], Molly J. Henry[2], Melvyn A Goodale[2,4],

[1]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

[2]The Brain and Mind Institute, The University of Western Ontario, London, ON, N6A 5B7 Canada

[3]The School of Psychology, University of East Anglia, Norwich, NR4 7TJ, United Kingdom

[4]Department of Psychology, The University of Western Ontario, London, ON, N6A 5C2

*Correspondence should be addressed to Juan Chen (juanchen@m.scnu.edu.cn) or Irene Sperandio (I.Sperandio@uea.ac.uk)

Lead contact: Juan Chen

Short title

Temporal evolution of size constancy

Key words (<10)

EEG, size constancy, real distance, retinal size, physical size, perceived size, representational similarity analysis

Summary

Our visual system provides a distance-invariant percept of object size by integrating retinal image size with viewing distance (size constancy). Single-unit studies with animals have shown that real distance cues, especially oculomotor cues such as vergence and accommodation can modulate the signals in the thalamus or V1 at the initial processing stage [1-7]. Accordingly, one might predict that size constancy emerges much earlier in time [8-10], even as visual signals are being processed in the thalamus. So far, the studies that have looked directly at size coding have either used fMRI (poor temporal resolution [11-13]) or relied on inadequate stimuli (pictorial illusions presented on a monitor at a fixed distance [11, 12, 14, 15]). Here, we physically moved the monitor to different distances, a more ecologically valid paradigm that emulates what happens in everyday life and is an example of the increasing trend of "bringing the real world into the lab". Using this paradigm in combination with electroencephalography (EEG), we examined the computation of size constancy in real time with real world viewing conditions. Our study provides strong evidence that, even though oculomotor distance cues have been shown to modulate the spiking rate of neurons in the thalamus and in V1, the integration of viewing distance cues and retinal image size takes at least 150 ms to unfold, which suggests that the size-constancy related activation patterns in V1 reported in previous fMRI studies (e.g. [12, 13]) reflect the later processing within V1 and/or top-down input from other high-level visual areas.

Results and Discussion

Experiment 1: Full-viewing condition

To investigate the influence of *real* distance on size coding, we physically placed the entire visual display at different distances from the observer (**Figure 1A**). In this more natural viewing paradigm, all distance cues including oculomotor adjustments (vergence, accommodation), binocular disparity, and pictorial cues, such as relative size, familiar size, occlusion, texture gradient, perspective, etc., were available and congruent with one another when participants viewed the stimuli binocularly with the room lights on (i.e., full-viewing condition).

To measure the temporal evolution of the representation of stimulus size (i.e., retinal image size versus physical size and perceived size) with the change of viewing distance, four conditions were examined: near-small (NS), near-large (NL), far-small (FS), and far-large (FL) (**Figure 1B**). Crucially, the stimuli in the NS and FL conditions had the same retinal image size, while those in the NS and FS conditions had the same physical size, as did those in the NL and FL conditions. The similarity between the different conditions in retinal image size and in physical size are reflected in the two "similarity matrices" shown in **Figure 1C,** which by definition were the same for all participants. Unlike retinal size or physical size, however, the perceived size of each stimulus varies between individuals and could be largely influenced by the availability and weighting of distance cues [16-18]. A continuous measure of perceived size was used only in Experiments 1a and 2. Therefore, similarity matrices for perceived size could be calculated in these two experiments (see **Figure 1C, right column** for an example of such a matrix in Experiment 2, in which distance cues were restricted). In Experiment 1, participants simply identified whether the stimulus was the small one or the large one by pushing one of two keys.

Importantly, to minimize the influence of any dynamic visual or oculomotor adjustments that would occur during the actual movement of the monitor on the EEG signals induced by the test stimulus, the stimulus was not presented until 1.5~2.5 s after the monitor had been moved and set in place at the far or near position. This interval between the placement of the monitor and the onset of the stimulus ensured that all the distance cues were processed and any visual and oculomotor signals evoked by the movement of the monitor had stabilized well before the stimulus was presented.

Participants all reported stimuli in both NS and FS as "small" and those in both NL and FL as "large". In other words, they all perceived the size of the stimulus according to its physical size regardless of viewing distance, suggesting that they had size constancy in the full-viewing condition. [In the behavioural part of Experiment 1a (see **Supplemental Information** for details), participants were asked to indicate the perceived size of each stimulus at each viewing distance by opening their thumb and index finger a matching amount (i.e., manual estimation) [16, 19, 20]. The results again confirmed that participants showed size constancy in the full-viewing condition (**Figure S1**)].

Figure 2A shows the event-related potentials averaged across all six electrodes of interest (CP3, CPZ, CP4, P3, PZ, and P4) [21-23]) for each of the four conditions. The first visually evoked component C1, especially the initial portion between 56-70 ms after stimulus onset, is thought to reflect the feedforward signals in V1 [24-27]. Any feedback from higher-level visual areas will appear later in the event-related potentials (ERPs). The C1 component in the current experiment had a peak latency of 56 ms on average, reflecting initial processing in V1 without any trial-specific top-down influences. If size constancy occurs at the initial stages of visual processing in V1 or even earlier in thalamus, then stimuli of the same physical size would be expected to evoke similar C1 amplitudes. However, we found that only the NL stimulus, which had the largest retinal image size, evoked a significant C1 ($t(1,15) = -3.86$; $p = 0.002$), and the amplitude of C1 evoked by the NL stimulus was significantly larger than the one evoked by the FL stimulus, which had the same physical and perceived (but not retinal) size as the NL stimulus ($t(1,16) = -3.08$, $p = 0.008$), suggesting that C1 reflected the retinal image size, but not the physical or perceived size of the stimulus.

As the ERP continued to unfold, the waveform appeared to cluster in a way that reflected the physical size of the stimuli rather than their retinal image size. Thus, as can be seen in **Figure 2A**, the waveforms for the NL and the FL conditions (blue lines) began to overlap one another as did the waveforms for the NS and FS (pink lines). To examine exactly when the transition from the representation of retinal image size to the representation of the physical size occurred, we calculated the difference in the amplitude of the ERPs between conditions that had the same retinal image size (FL-NS) and conditions that had the same physical size (FS-NS and FL-NL). The difference scores (**Figure 2B**) revealed that waveforms for the stimuli with the same retinal image size (FL and NS) overlapped completely until 148 ms after stimulus onset at which point they began to separate, suggesting that before this time point the activity in visual cortex reflected only the retinal image size [$p_{corrected} < 0.05$, corrected using a cluster-based test statistic (Monte Carlo) method embedded in Fieldtrip toolbox [28]; the same criterion was used for all time-course-related comparisons hereafter]. In contrast, the difference scores showed that the waveforms for the two small stimuli (FS and NS) began to overlap at 150 ms after stimulus onset

and the waveforms for the two large stimuli (FL and NL) at 144 ms (**Table S1**), suggesting that after these time points, the activity in visual cortex began to reflect the physical size of the visual stimuli.

We also performed a representational similarity analysis (RSA) based on the *patterns* of signals from all six electrodes within a 20-ms sliding time window. Each element of the similarity matrix for neural signals was the Pearson's correlation between the EEG signal patterns of each pair of conditions (see **Methods** for details). If the visual signals were representing retinal image size, then the similarity matrix for the EEG signal patterns (neural model) should have a higher correlation with the similarity matrix for the retinal image size (retinal model, **Figure 1C left**) than with the similarity matrix for the physical size (physical model, **Figure 1C middle**). Consistent with our prediction, the RSA revealed that the neural model was significantly correlated with the retinal model before about 150 ms (**Figure 2C**, see **Table S2** for details. Note: numbers in **Table S2** show the *start* point of the 20-ms-sliding window), and was significantly correlated with the physical model after about 124 ms. Importantly, the neural model was more strongly correlated with the retinal model at 50~150 ms and was more strongly correlated with the physical model at a later time window, although the latter difference did not survive correction for multiple comparisons (**Figure 2C**). Taken together, these results provide converging evidence that during the early stages of visual processing (within the first ~150 ms) the observed activity is locked to the retinal image size but later on it begins to reflect the real-world size of a visual stimulus.

One might argue that the post-150 ms overlap in the waveforms for stimuli of the same real-world size in Experiment 1 might be due to nothing more than the fact that participants had only two choices in their behavioral response: small or large. To rule this out, in Experiment 1a, we replicated the EEG protocol of Experiment 1, but asked participants to detect the onset of an open circle that was randomly interleaved with the experimental stimuli (solid circles) during the EEG recording. The results were consistent with those in Experiment 1 (**Supplemental Information, Figure S2**), which suggests that size-distance integration is to some extent automatic and independent of the task the participants were performing. Moreover, because each participant gave an estimate of the perceived size of the stimulus in each condition, we were able to compute the similarity matrix for perceived size for each participant. The RSA results showed that the correlation of the neural model with the physical-size model and the correlation of the neural model with the perceived-size model overlapped almost perfectly (**Figure S2C**), which is not surprising given that almost all the participants showed size constancy.

One may also argue that the late convergence of ERP components between conditions with the same physical size reflects the white-black pattern because the ratio of the black stimulus area to the white background area correlates with the physical size of the stimulus regardless of viewing distance. This is unlikely because the ERPs were time-locked to the onset of the stimulus. Importantly, our Experiment 2 also shows that the later ERP components reflect the perceived size of the stimulus, not the white-black patterns (see below).

Experiment 2: Restricted-viewing condition

In Experiment 2, we removed most of the cues to viewing distance, which would be expected to disrupt size constancy [16, 17]. If size constancy emerges in the grouping of the EEG

components after 150 ms, as our earlier results with full viewing suggested, then under restricted viewing we expected to see disruption in that grouping.

The stimuli were white solid circles presented on a black background. Participants were asked to view the stimulus with their non-dominant eye through a 1-mm pinhole in an otherwise completely dark room [16, 17] (i.e. restricted-viewing condition, **Figure 3A**), while performing a size-irrelevant detection task (as in Experiment 1a) during the EEG recording. In this situation, no binocular distance cues (i.e., vergence, binocular disparity) were available and pictorial cues were dramatically reduced as the background merged with the edges of the pinhole in the darkened room. In addition, the small pinhole prevented participants from using accommodation as a reliable cue to distance [29]. As a result, participants would have to rely mainly on retinal image size to judge object size; thus, a stimulus at the near distance would be perceived as larger than the same stimulus at the far distance because the stimulus would subtend a larger retinal image size at the near distance [16, 17].

However, because participants still knew whether the monitor was at the near or the far position, presumably on the basis of cues from the moving monitor when its position was changing and from other cues, such as retinal illuminance, size constancy was not affected by the restricted-viewing condition to the same extent across participants. Given that the purpose of this experiment was to explore the neural correlates of perceived size when size constancy was disrupted, we performed a behavioral screening test before the real experiment to select participants. 15 out of the 32 participants whose size constancy was disrupted to some degree and one participant who showed perfect size constancy in the restricted-viewing condition were selected and performed both the behavioral and the EEG portions of the main experiment. Their behavioral results are shown in **Figure 3B**.

The peak of C1 in Experiment 2 occurred approximately 20 ms later than it did in Experiment 1 , probably because only one eye was being stimulated in this experiment [30]. Nevertheless, consistent with Experiment 1, the NL stimulus, which had the largest retinal size, evoked the strongest C1 component (compared with the amplitude of the other three conditions, paired t-test, all $t < 3.13$, $p < 0.006$; **Figure 4A, middle**), again suggesting that retinal image size, not physical size, was driving the activity of the early ERP components. The waveforms for those conditions in which the stimulus subtended the same retinal image size (NS and FL) began to depart from each other around 144 ms after stimulus onset (**Figure 4B, Table S1**), just as they did in Experiment 1, but overall the waveforms did not show the same clear groupings according to physical size as they did in Experiment 1. Instead, the waveform evoked by the NL stimulus began to separate from the FL stimulus approximately 154 ms after stimulus onset and never showed any overlap with FL, even though they had the same physical size. This pattern is consistent with the fact that, under restricted viewing condition, the NL stimulus was perceived to be the largest stimulus of the four (**Figure 3B**).

Given that there was considerable variability in size constancy across participants (**Figure 3B**), we then tested whether this variability in size constancy would also be reflected in the later components of the EEG waveforms. To this end, we calculated a behavioral index (BI) of disruption in size constancy and an EEG index (EI) of disruption in size constancy for the late component of the ERPs (blue shaded area from 154 ms to 350 ms in **Figure 4A, middle**) for each participant (see Methods for details), and then calculated the correlation between them across participants. We found that there was indeed a significant correlation between BI and EI

across participants ($r = 0.55$, $p = 0.03$; **Figure 4A, right**). We also calculated a similar correlation between BI and EI for the early C1 component (the orange shaded area in **Figure 4A, middle**), but the correlation was not significant ($r = -0.30$, $p = 0.28$; **Figure 4A, left**), suggesting that the variability in perceived size across participants is reflected in the later ERP components but not in C1.

RSA was again performed to reveal the time course of the representation of size (retinal size, physical or perceived size). For the similarity matrix of perceived size, the manual estimates of perceived size provided by the participants were used just as in Experiment 1a (**see Method Details**). As predicted, although the retinal model and the perceived model were both highly correlated with the neural model from about 80 ms after stimulus onset (see **Table S2** for details), we found a trend in favor of the retinal model at the early stage (**Figure 4C**, orange is above green) and a trend in favor of the perceived model at the later stage (**Figure 4C**, green is above orange, see **Table S2** for statistical results). This again provides convincing evidence that the integration of viewing distance with retinal size does not occur until the later stage of visual processing.

Because white circles, instead of black circles, were used in this experiment, one might argue that the retinal illuminance and pupil size would have varied with viewing distance, which might affect the ERP signals. But those effects would likely be smaller compared to changes in retinal size and in any case would likely influence the early components. Our RSA results also confirmed that the ERPs after 150 ms did represent the perceived size. In addition, in Experiment 2, all the participants saw was a white disk (the black background merged completely with the edge of the pinhole in the dark). Therefore, there was no possibility that the ERP activity could reflect differences in the pattern or black-white-ratio of the display.

It is important to note that we changed the physical distance of the stimulus display from trial to trial, so that in the full-viewing condition, a large range of distance cues was available and entirely congruent with one another. A previous study showed that when real distance was manipulated, the size-distance scaling was much stronger than when only pictorial cues were provided [13]. Moreover, the long interval after the monitor had been set in place provided enough time for the distance cues to be well processed before the onset of the stimulus, so that the distance information could theoretically be integrated with the retinal information about the test stimulus as soon as it was presented. For all these reasons, the time (i.e., 150 ms after stimulus onset) we identified as the transition point from the coding of retinal image size to the coding of perceived size is probably the earliest possible time point at which the integration of retinal image size and viewing distance information can take place.

The 150 ms required for the size-distance integration is consistent with the time that is typically required (80 to 150 ms after stimulus onset) for the feedback from higher-order visual areas to V1 or recurrent processing within V1 [31]. Therefore, our results suggest that although the activation related to size constancy was observed in early visual area V1 in previous fMRI studies [10-13], the key integration does not happen at the initial visual processing in V1.

Recurrent feedback to V1 has been shown to be critical for feature binding [32, 33]. In a similar fashion, such feedback could be used to integrate distance information with retinal image size to calculate the real-world size of objects, and subsequently, integrate real-world size with other object features, such as shape, colour, and visual texture. Indeed, it is worth noting that accounts of feature integration have almost entirely ignored object size, perhaps because only images

presented on a display at a fixed distance rather than real objects presented at different distances have been employed in these studies.

On the face of it, the 150 ms required for size-distance integration in perception seems surprisingly late given that cues like vergence and accommodation modulate the spiking rate of neurons in LGN, SC, and the initial response in V1 [1-7, 34]. But it is likely that, although the integration of retinal image size and distance information takes at least 150 ms for perception, some oculomotor distance information could be conveyed rapidly to visuomotor networks in the dorsal stream [27, 35] to mediate action. It has been suggested that efference copy information from vergence (and theoretically accommodation) is conveyed from the superior colliculus (via thalamic nuclei) to the frontal eye fields and to visuomotor areas in the posterior parietal cortex, completely by-passing the geniculostriate pathway altogether [36-38]. Additional support for this idea comes from studies showing that patients with lesions of V1 can scale the opening of their grasping hand to the size and orientation of goal objects [39-42], even though they do not perceive those objects.

Acknowledgments

We thank Amratha Chandrakumar and Jason Kim for their help with data collection. This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (No. RGPIN-2017-04088 to MAG), a grant from the Canadian Institute for Advanced Research (to MAG), and a grant from the National Natural Science Foundation of China (No. 31800908 to JC).

Author Contributions

J.C., I.S., and M.A.G. designed the study. J.C. performed the research. J.C. and M.J.H. analyzed the data. All authors contributed to the writing of the manuscript.

Declaration of Interests

The authors declare no competing financial interests.

Figure Legends

Figure 1 The setup, design, and the "similarity" matrices between conditions. (A) In Experiment 1 and the control experiment (Experiment 1a), participants viewed the stimuli binocularly with room lights on (i.e., full-viewing condition). The stimulus was a black solid circle on a white background, and therefore the changes in the retinal illuminance with distance were minimized. The monitor was placed on a movable track so that it could be moved to different distances from the observer. (B) Solid circles of two sizes (Small = 4 cm and Large = 8 cm) were presented at two distances (Near = 28.5 cm and Far = 57 cm). (C) The retinal-image size similarity matrix, the physical-size similarity matrix, and the perceived-size similarity matrix for all conditions. The retinal-size and physical-size matrices consisted of values of "0" s (i.e. 0s indicate "different") or "1"s (1s indicate the "same"). The elements of the perceived size similarity matrix were calculated for each participant based on the "similarity" of the reported

perceived size between each pair of conditions. "Similarity" was operationally defined as the difference in perceived size between each pair of conditions multiplied by -1. The matrix on the right shows an example of "similarity" in perceived size in Experiment 2 in which distance cues were restricted. For Experiment 1, no continuous estimates of perceived were collected, and therefore only the retinal-size model and physical-size model were tested. For Experiment 1a, all the participants showed excellent size constancy, so the similarity matrix for perceived size (not shown in this figure) was essentially identical to the similarity matrix for physical size.

Figure 2 ERP results of Experiment 1. (A) ERP curves that were first averaged across all six electrodes of interest for each participant and then averaged across participants for each condition. (B) The difference in amplitude between conditions that had the same retinal size (i.e., between NS and FL), and between conditions that had the same physical size (i.e., between FS and NS, and between FL and NL). The gray arrow points to approximately when the representation of retinal image size ended and when the signals began to change to represent the physical size (see Table S1 for statistical results). (C) The results of the representational similarity analysis (RSA). Each curve shows the time course of correlation between the similarity matrix of the neural model obtained from the ERP amplitude pattern and the similarity matrix of each of the size models (Retinal Size model and Physical Size model). The horizontal axis shows the *start point* of the 20-ms sliding time window. Shaded regions show standard error of the mean. The colored thick bars show when the values on each curve were significantly different from 0. The gray box shows when the two correlations were significantly different (see Table S2 for statistical results). The p values were corrected using a cluster-based test statistic (Monte Carlo) method embedded in FieldTrip toolbox [28]; the same criterion was used for all time-course-related comparisons hereafter. See Figures S1 and S2, and Tables S1 and S2 for the perceived-size results and ERP results of Experiment 1a in which participants viewed the same stimuli in the same full-viewing condition as they did in Experiment 1 but performed a different task.

Figure 3 Restricted-viewing condition and the behavioral results of perceived size in Experiment 2. (A) Participants viewed the stimuli monocularly through a 1 mm pinhole in complete darkness. The stimuli were solid white circles presented on a black screen. Through the 1-mm hole, participants were able to see only part of the monitor (dashed-line circle) but not the borders. Again, the monitor was moved to different distances with the same setup as that in Experiment 1. (B) The perceived size (measured via manual estimation) for each individual (shown as each gray line with symbols) in Experiment 2 during restricted viewing and their average results (black lines with symbols).

Figure 4 Results of Experiment 2. (A) Middle: ERP curves that were first averaged across all six electrodes for each participant and then averaged across participants for each condition. Left: Scatter plot showing the correlation between the amount of size-constancy disruption reflected in the perceived size (i.e., behavioral index) and the amount of size-constancy disruption reflected in the earliest visual-evoked component C1 (i.e., the orange area in the middle figure, EEG index). Right: scatter plot showing the correlation between the behavioral index and the EEG index reflected in the later ERP components (i.e., the blue area in the middle figure). (B) The difference in ERP amplitude between conditions that had the same retinal size or the same physical size (see Table S1 for statistical results). (C) RSA results. Each curve shows the time course of the correlation between the similarity matrix of each size model and the similarity matrix of the neural model obtained from the ERP activation pattern. The horizontal axis shows

the start point of the 20-ms sliding time window. Shaded regions show standard error of the mean. Again, the colored thick bars in (B) and (C) show when the values on each curve were significantly different from 0 and the gray box shows when the difference in the correlation of neural model with Retinal Model and with Perceived Model was statistically significant (see Table S2 for statistical results).

STAR★Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact Juan Chen (juanchen@m.scnu.edu.cn).

Experimental Model and Subject Details

Seventeen participants took part in Experiment 1. One participant's data were discarded because of strong noise in his EEG signals. The ages of the remaining 16 participants (6 males, 10 females) ranged between 21 and 27 (M = 24.4, SD = 1.86). Six of the participants of Experiment 1 and ten naïve participants (16 in total, 5 males and 11 females with ages ranging between 19 and 27, M = 23.06, SD = 2.69) took part in the EEG portion of Experiment 1a, but only 14 of them took part in the behavioral portion of the experiment where participants were asked to manually estimate the perceived size of the stimulus. Two participants were unable to complete the behavioral portion because they had to leave the testing session before it was finished. Sixteen participants took part in both the EEG portion and the behavioral size estimation task of Experiment 2 (6 males and 10 females). One of them also took part in Experiment 1 and another also took part in Experiment 1a. Their ages ranged between 19 and 52 (M = 26.69, SD = 9.34). All participants were right handed and had no history of neurological impairments. Participants in Experiments 1 and 1a had either normal or corrected-to-normal visual acuity. All participants in Experiment 2 had normal visual acuity. Informed consent was obtained from all subjects according to procedures and protocols approved by the Health Sciences Research Ethics Board at The University of Western Ontario.

Method Details

Stimuli and setup

In Experiments 1 and 1a, the stimuli were black (luminance: 0.74 cd/m^2) solid circles with a diameter of 4 cm (i.e. 'Small' or 'S') or 8 cm (i.e. 'Large' or 'L') (**Figure 2B**). They were presented in the center of a screen with a white (luminance: 79.13 cd/m^2) background. The stimulus was presented on a 19 inch monitor (ViewSonic, width: 37.5 cm, height: 30 cm). The display monitor was mounted on a movable track so that the experimenter could move it to a

near (28.5 cm, 'N') or a far viewing distance (57 cm, 'F') (**Figure 2A**). We used black circles on a white background, instead of white circles on a black background as stimuli, so that the changes in retinal illuminance with distance should be minimized. We used solid circles, instead of gratings or other complex objects as stimuli, to avoid any confound of differences in spatial frequency at different viewing distances. There was a fixation point (a red dot) on the center of the screen throughout the experiments. Participants were seated in front of the screen with their chin on a chinrest. This experiment was performed with the room lights on and under binocular viewing conditions (i.e., full-viewing condition).

In Experiment 2, the same design as described above (2 sizes × 2 distances) was adopted. The room was completely dark and participants looked at the stimuli through a 1 mm hole on the pin-hole glasses with their non-dominant eye (i.e., restricted-viewing condition). The stimuli were *white* (luminance: 79.13 cd/m^2) solid circles presented on a *black* (luminance: 0.74 cd/m^2) background. The reason for using white circles as stimuli was that if black circles were presented on a white background in Experiment 2, participants would be able to see the boundary of the circular field of view clearly when they wore pin-hole glasses. The relative size between the circular stimuli and the area they could see through the pin-hole would have provided them with information regarding the size of the stimuli, which would have made it impossible to disrupt size constancy.

Procedure

In Experiment 1, participants were asked to indicate whether a solid circle was small or large regardless of distance by pressing two keys ("1" for small and "2" for large) during EEG recording. At the beginning of each trial, the experimenter was cued with a small letter, either 'N' or 'F', that appeared at the corner of the screen to indicate whether the viewing distance of a specific trial would be near or far (note: the participants could not see the letter in their far periphery). The experimenter who sat beside the monitor would move the monitor to the near or far position, accordingly. 1.5 ~2.5 s after the screen was moved to the right position, the experimenter pushed a key to trigger the presentation of the stimulus. The stimulus was presented on the screen for 0.2 s. Participants were asked to maintain fixation at the fixation point throughout the experiment. There were 100 trials in each run, with 25 trials for each condition.

In Experiment 1a, the protocol of the EEG trials was the same as that described for Experiment 1 with two exceptions. First, during EEG recording in each run, there were 10 additional trials in which the stimulus was an open circle of a middle size, rather than a solid circle. Participants were asked to push a key ("0") as soon as they saw the open circle (i.e., size-irrelevant detection task). Second, in addition to the EEG trials, 14 out of the 16 participants also performed a behavioral task in which they were asked to open their thumb and index finger to indicate the perceived size of the stimulus (manual estimation task) [16, 19, 20]. The distance between the finger and thumb was then measured with a measuring tape. This psychophysical measure was taken after the EEG session. Participants completed 4-5 psychophysical blocks depending on the time available, with 2 manual estimates for each of the four conditions in each block. [Note that it is unlikely that the six of the 16 participants who performed both Experiments 1 and 1a would also be implicitly categorizing the two "main" stimuli as "Small" or "Large" in Experiment 1a because the target stimulus in the detection task of Experiment 1a was different in size from the

other two. Moreover, the most obvious difference between the target stimulus and the other two stimuli was that it was an open rather than a solid circle.]

In Experiment 2, the same EEG protocol was used as reported above. Participants performed the same size-irrelevant detection task as in Experiment 1a during EEG recording and also performed a separate behavioral testing session as in Experiment 1a. Unlike Experiment 1a, the psychophysical blocks were performed before any EEG recordings and after every three or four EEG runs, in case the perceptual experience of size changed over EEG runs.

In all experiments, the order of the four conditions was randomized on a trial-by-trial basis. Participants completed between 8 and 14 runs of EEG recording depending on the time available, for a total of 200-300 repetitions for each condition. Each experiment lasted between 3 and 4 hours.

It should be noted that size constancy was not affected by the restricted-viewing condition to the same extent across participants, probably because of individual differences in their ability to use residual depth cues (e.g. vibration or auditory cues provided by the movement of the monitor, or changes in the retinal illuminance of the white stimulus) to enable size constancy. (In another study from our lab in which we moved a sphere, rather than a monitor, to different locations on a table, we were able to successfully disrupt size constancy in all participants using the same restricted-viewing condition [16]). To investigate if the early or the late components of ERPs reflect perceived size, we did a behavioral screening to select participants. Fifteen out of the 32 participants we screened showed size constancy disruption to some degrees. These 15 participants and an additional participant whose size constancy was perfect in the restricted-viewing condition were included in Experiment 2.

EEG measurements

Scalp EEG was collected using NeuroScan Acquire 4.3 recording system (Compumedics) from 32 Ag/AgCl electrodes positioned according to the extended international 10 – 20 EEG system. Vertical electro-oculogram (VEOG) was recorded from two electrodes placed above or below the left eye. Horizontal EOG (HEOG) was recorded from two electrodes placed at the outer canthus of the left and the right eyes. Because we were interested in the six electrodes at the parietal and occipital part of the scalp (i.e., CP3, CPZ, CP4, P3, PZ, and P4) that have been reported to reflect visual processing [21-23], we always kept the impedance of these six electrodes below 10 kΩ. We also tried to keep the impedance of the other electrodes as low as possible, but this revealed to be impossible for all participants due to the long duration of the EEG session (> 3 hours). EEG was amplified with a gain of 500 K, band pass filtered at 0.05 – 100 Hz, and digitized at a sampling rate of 500 Hz. The signals on these electrodes were referenced online to the electrode on the nose.

Quantification and Statistical Analysis

ERP data Preprocessing

Offline data analysis was performed with NeuroScan Edit 4.3 (Compumedics) and MATLAB R2014 (Mathwork). The EEG data was first low-pass filtered at 30 Hz, and then epoched starting at 100 ms before the stimulus onset and ending 400 ms after stimulus onset. Each epoch was baseline-corrected against the mean voltage of the 100 ms pre-stimulus interval. The epochs

contaminated by eye blinks, eye movements, or muscle potentials exceeding ± 50 μV at any electrode were excluded from the average.

Amplitude and latency analyses of ERP components

For the event-related potential (ERP) analysis, the remaining epochs after artifact rejection were averaged for each condition. Preliminary analyses revealed that the activity pattern of the four conditions in all 6 electrodes (i.e., CP3, CPZ, CP4, P3, PZ, and P4) were similar. Therefore, only the ERP amplitude and latency results that were averaged across these six electrodes were reported. The peak amplitude and latency of each component were acquired for each condition and each participant.

Representational similarity analysis (RSA)

To examine at what time the brain activity was representing the retinal size, physical size or perceived size, we calculated the correlation between the similarity matrix revealed in neural signals (i.e., ERP amplitude) and similarity matrices for the retinal size, physical size and the perceived size, respectively, for each sliding window (10 data points, i.e., 20 ms) with the first point of the window moving from -100 ms to 382 ms. The element of the similarity matrix for the neural model (i.e., EEG signals) was set as the Fisher-Z correlation coefficient between the EEG patterns for each pair of conditions at a specific time window. Each EEG patterns included 60 elements (10 data points × 6 electrodes).

The similarity matrices for the retinal size and the physical size are shown in **Figure 1C left and middle, respectively**. The similarity between two conditions was set as 1 if the retinal size or the physical size was the same, but was set as 0 if the retinal size or the physical size was different. These matrices were fixed across participants. The similarity matrix for perceived size was calculated for each individual in Experiments 1a and 2 (see **Figure 1C, right** for an example in Experiment 2). Each element of the matrix was obtained by first calculating the perceived size difference between two conditions, and then multiplying the obtained value by -1. For Experiment 1, no perceived size data was collected for each individual, and therefore only retinal-size model and physical-size model were tested. For Experiment 1a, all the participants showed excellent size constancy, so the similarity matrix for perceived size (not shown in this figure) was essentially identical to the similarity matrix for physical size.

To obtain an unbiased measurement of the correlation between the neural model and the size model, we used a procedure similar to the n-folded cross-validation that was commonly used in pattern recognition analysis [43]. Specifically, we first randomly sampled half group of trials from the whole set of ERP trials for each condition, then we averaged the ERPs of the sampled trials. The averaged ERPs were used to calculate the correlation coefficients between the EEG patterns of each pair of conditions (i.e., the elements of the neural model) at each sliding time window and to calculate the correlation between the obtained neural model and size model. This procedure was repeated 50 times. The 50 correlation coefficients between the neural model and size model were first converted to Fisher-Z scores, and were then averaged to obtain the reported correlation results.

Correlation between size constancy disruption index calculated in perceptual judgments and in ERP components

In Experiment 2, to test which ERP component reflected the individual variability in size-constancy disruption, we calculated the correlation between the amounts of size-constancy disruption measured behaviourally and the amount of size-constancy disruption measured in the ERP components across individuals.

The behavioral size-constancy disruption index (BI) was defined as the difference in perceived size between the NL and the FL conditions normalized by the perceived size in the FL condition, i.e.,

$$\mathrm{BI} = \frac{ME_{NL} - ME_{FL}}{ME_{FL}}, \quad (1)$$

where ME indicates manual estimate of perceived size.

The EEG size constancy disruption index (EI) was defined as the area between the ERP waveforms for the NL and FL conditions normalized by the area under the FL waveform in an interval, i.e.,

$$\mathrm{EI} = -\frac{Area_{NL} - Area_{FL}}{Area_{FL}}, \quad (2)$$

where "Area" stands for the numerical integration under the curve in a specific interval. For C1, this interval was when the C1 amplitudes was significant in the NL condition. Practically, this interval were when C1 amplitudes were significantly higher than the 25% of the peak amplitude of the C1 in the same condition. In the current case, the interval was between 78-90 ms after stimulus onset (the orange shaded area in **Figure 4A, middle**). For the late EEG component, the interval was when the amplitude of NL was significantly different from the FL condition (blue shaded area from 154 ms to 350 ms in **Figure 4A, middle**). The large size, but not the small size, was used to calculate the behavioral and EEG size-constancy disruption indices because the size constancy disruption (i.e., the difference in perceived size or in ERP amplitude between near and far distances) was more evident and reliable in the large size condition than in the small size condition in both the behavioral and EEG results. Pearson correlation was calculated to test whether or not the correlation between behavioral performance and neural signals was significant. For C1, one outlier (beyond +/-5 SD) was excluded.

Statistical Analysis

To examine whether or not there was size constancy, repeated ANOVAs with size and distance as main factors were carried out to reveal specifically whether or not the main effect of distance was significant. To compare the amplitude of C1 component evoked by different conditions, paired sample t-tests were performed on the peak value of the C1 amplitude. To search intervals when there were significant differences between each time course and 0 or between two time courses, paired sample t-tests were conducted point-by-point, and they were then corrected for multiple comparisons using the cluster-based test statistic embedded in FieldTrip toolbox [28] (Monte Carlo method, $p < 0.05$). For the RSA results and the correlation between BI and EI results, all statistical comparisons were conducted on the Fisher Z scores of the Pearson correlation coefficients.

Data and Software availability

The primary data of this study can be found at http://bmi.ssc.uwo.ca/Chen_CB2019/

References

1. Lal, R., and Friedlander, M.J. (1990). Effect of passive eye position changes on retinogeniculate transmission in the cat. J Neurophysiol *63*, 502-522.
2. Weyand, T.G., and Malpeli, J.G. (1993). Responses of neurons in primary visual cortex are modulated by eye position. J Neurophysiol *69*, 2258-2260.
3. Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., and Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. Science *257*, 1279-1281.
4. Trotter, Y., and Celebrini, S. (1999). Gaze direction controls response gain in primary visual-cortex neurons. Nature *398*, 239-242.
5. Dobbins, A.C., Jeo, R.M., Fiser, J., and Allman, J.M. (1998). Distance Modulation of Neural Activity in the Visual Cortex. Science *281*, 552-555.
6. Masson, G.S., Busettini, C., and Miles, F.A. (1997). Vergence eye movements in response to binocular disparity without depth perception. Nature *389*, 283-286.
7. Rosenbluth, D., and Allman, J.M. (2002). The Effect of Gaze Angle and Fixation Distance on the Responses of Neurons in V1, V2, and V4. Neuron *33*, 143-149.
8. Richards, W. (1968). Spatial remapping in the primate visual system. Kybernetik *4*, 146-156.
9. Richards, W. (1971). Size-distance transformations. In Zeichenerkennung durch biologische und technische Systeme/Pattern Recognition in Biological and Technical Systems. (Springer), pp. 276-287.
10. He, D., Mo, C., Wang, Y., and Fang, F. (2015). Position shifts of fMRI-based population receptive fields in human visual cortex induced by Ponzo illusion. Exp Brain Res *233*, 3535-3541.
11. Fang, F., Boyaci, H., Kersten, D., and Murray, S.O. (2008). Attention-Dependent Representation of a Size Illusion in Human V1. Curr Biol *18*, 1707-1712.
12. Murray, S.O., Boyaci, H., and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. Nat Neurosci *9*, 429-434.
13. Sperandio, I., Chouinard, P.A., and Goodale, M.A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. Nat Neurosci *15*, 540-542.
14. Liu, Q., Wu, Y., Yang, Q., Campos, J.L., Zhang, Q., and Sun, H.J. (2009). Neural correlates of size illusions: an event-related potential study. Neuroreport *20*, 809-814.
15. Ni, Amy M., Murray, Scott O., and Horwitz, Gregory D. (2014). Object-Centered Shifts of Receptive Field Positions in Monkey Primary Visual Cortex. Curr Biol *24*, 1653-1658.
16. 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*, 1-6.
17. Holway, A.H., and Boring, E.G. (1941). Determinants of apparent visual size with distance variant. Am. J. Psychol, 21-37.
18. Sperandio, I., and Chouinard, P.A. (2015). The mechanisms of size constancy. Multisens Res *28*, 253-283.
19. Chen, J., Jayawardena, S., and Goodale, M.A. (2015). The effects of shape crowding on grasping. J Vis *15*, 1-9.
20. 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.
21. Luck, S.J. (2005). An Introduction to the Event-Related Potential Technique, (Cambridge, MA: Massachusetts Institute of Technology).
22. Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., and Fang, F. (2014). Attention-Dependent Early Cortical Suppression Contributes to Crowding. J Neurosci *34*, 10465-10474.

23. Chen, J., Yu, Q., Zhu, Z., Peng, Y., and Fang, F. (2016). Spatial summation revealed in the earliest visual evoked component C1 and the effect of attention on its linearity. J Neurophysiol *115*, 500-509.
24. Clark, V.P., Fan, S., and Hillyard, S.A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. Hum. Brain Mapp. *2*, 170-187.
25. Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., and Hillyard, S.A. (2002). Cortical sources of the early components of the visual evoked potential. Hum. Brain Mapp. *15*, 95-111.
26. Bao, M., Yang, L., Rios, C., He, B., and Engel, S.A. (2010). Perceptual Learning Increases the Strength of the Earliest Signals in Visual Cortex. J Neurosci *30*, 15080-15084.
27. Foxe, J.J., and Simpson, G.V. (2002). Flow of activation from V1 to frontal cortex in humans. Exp Brain Res *142*, 139-150.
28. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput intel neurosc *2011*, 9.
29. Hennessy, R.T., Iida, T., Shiina, K., and Leibowitz, H. (1976). The effect of pupil size on accommodation. Vision Res. *16*, 587-589.
30. Mirzajani, A., and Jafari, A. (2014). The effect of binocular summation on time domain transient VEP wave's components. Razi Journal of Medical Sciences *21*, 29-35.
31. Wyatte, D., Jilk, D.J., and O'Reilly, R.C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. Front Psychol *5*, 674.
32. Bouvier, S., and Treisman, A. (2010). Visual Feature Binding Requires Reentry. Psychol. Sci. *21*, 200-204.
33. Koivisto, M., and Silvanto, J. (2011). Relationship between visual binding, reentry and awareness. Conscious Cogn *20*, 1293-1303.
34. Marg, E., and Adams, J. (1970). Evidence for a neurological zoom system in vision from angular changes in some receptive fields of single neurons with changes in fixation distance in the human visual cortex. Cell. Mol. Life Sci. *26*, 270-271.
35. Chen, C.-M., Lakatos, P., Shah, A.S., Mehta, A.D., Givre, S.J., Javitt, D.C., and Schroeder, C.E. (2007). Functional Anatomy and Interaction of Fast and Slow Visual Pathways in Macaque Monkeys. Cereb Cortex *17*, 1561-1569.
36. Sommer, M.A., and Wurtz, R.H. (2008). Brain circuits for the internal monitoring of movements. Annu. Rev. Neurosci. *31*, 317-338.
37. Grieve, K.L., Acuña, C., and Cudeiro, J. (2000). The primate pulvinar nuclei: vision and action. Trends Neurosci. *23*, 35-39.
38. Hanslmayr, S., Volberg, G., Wimber, M., Dalal, Sarang S., and Greenlee, Mark W. (2013). Prestimulus Oscillatory Phase at 7 Hz Gates Cortical Information Flow and Visual Perception. Curr Biol *23*, 2273-2278.
39. Prentiss, E.K., Schneider, C.L., Williams, Z.R., Sahin, B., and Mahon, B.Z. (2018). Spontaneous in-flight accommodation of hand orientation to unseen grasp targets: A case of action blindsight. Cogn Neuropsychol, 1-9.
40. Whitwell, R.L., Striemer, C.L., Nicolle, D.A., and Goodale, M.A. (2011). Grasping the non-conscious: Preserved grip scaling to unseen objects for immediate but not delayed grasping following a unilateral lesion to primary visual cortex. Vision Res. *51*, 908-924.
41. Carey, D.P., Harvey, M., and Milner, A.D. (1996). Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia. Neuropsychologia *34*, 329-337.
42. Carey, D.P., Dijkerman, H.C., and Milner, A.D. (1998). Perception and action in depth. Conscious Cogn *7*, 438-453.
43. Bishop, C.M. (2006). Pattern Recognition and Machine Learning, (Springer).