Published in final edited form as:
J Vis. ; 10(1): 5.1–513. doi:10.1167/10.1.5.

Cue combination for 3D location judgements

Ellen Svarverud,
School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK, & Department of Optometry and Visual Science, Buskerud University College, Kongsberg, Norway

Stuart J. Gilson, and
Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK

Andrew Glennerster
School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

Abstract

Cue combination rules have often been applied to the perception of surface shape but not to judgements of object location. Here, we used immersive virtual reality to explore the relationship between different cues to distance. Participants viewed a virtual scene and judged the change in distance of an object presented in two intervals, where the scene changed in size between intervals (by a factor of between 0.25 and 4). We measured thresholds for detecting a change in object distance when there were only 'physical' (stereo and motion parallax) or 'texture-based' cues (independent of the scale of the scene) and used these to predict biases in a distance matching task. Under a range of conditions, in which the viewing distance and position of the target relative to other objects was varied, the ratio of 'physical' to 'texture-based' thresholds was a good predictor of biases in the distance matching task. The cue combination approach, which successfully accounts for our data, relies on quite different principles from those underlying traditional models of 3D reconstruction.

Keywords

virtual reality; 3D representation; cue combination; motion parallax; binocular stereopsis; perceived distance

Introduction

Traditionally, the view of 3D representation in human vision is a geometric one, in which images from the two eyes, or changes in a monocular image, are used to deduce the most likely 3D scene that could generate those views. In computer vision, this process is known as photogrammetry (Hartley & Zisserman, 2000; Longuet-Higgins, 1981). The output of this process is a description of the 3D location of points in the scene up to an unknown scale factor which can be provided by knowledge of the interocular separation or the distance traveled by the observer. In general, it is assumed that these scale factors are available to, and used by, the visual system in judging distances. This view of visual reconstruction predominates in the literature, even though there is debate about the extent to which the representation of space is distorted (Foley, 1980; Gogel, 1990; Johnston, Cumming, &

Corresponding author: Andrew Glennerster. a.glennerster@reading.ac.uk. Address: Earley Gate, Reading RG6 6AL, UK..

Commercial relationships: none.

Europe PMC Funders Author Manuscripts

Parker, 1993; Luneburg, 1950). There have been suggestions that there may be no single visual representation of a 3D scene that can account for performance in all tasks (Brenner & van Damme, 1999; Glennerster, Rogers, & Bradshaw, 1996). Instead, observers' responses may depend on separable 'modules' (Landy, Maloney, Johnston, & Young, 1995) and may not require a globally consistent map of space (Foo, Warren, Duchon, & Tarr, 2005; Glennerster, Hansard, & Fitzgibbon, 2001, 2009). In a very different context, cue combination approaches have been applied to judgements of surface slant (Hillis, Ernst, Banks, & Landy, 2002; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003) and object shape (Ernst & Banks, 2002; Johnston et al., 1993). As we shall see, the predictions of a cue combination model for object location can be quite different from those of a 3D model of the scene (whether that model is distorted or not).

In the experiments reported here, we used an expanding virtual room (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006; Rauschecker, Solomon, & Glennerster, 2006) to explore whether biases from a distance matching task could be predicted from the thresholds for two types of cue. The first type we called 'physical', which signals the distance of the object relative to the observer, including stereo and motion parallax information. Specifically, binocular disparities and motion parallax provide information that allows a 3D reconstruction of the scene to be computed, where the overall scale of the reconstruction depends on the estimated length of the 'baseline', i.e. the interocular separation or, in the case of motion parallax, the distance the eye has translated between two views (Hartley & Zisserman, 2000; Longuet-Higgins, 1981). The second type of cue comprises all those sources of information about viewing distance that are unaffected by the expansion of the room. For example, when the virtual room contracts, the textures depicting the bricks on the walls and the tiles on the floor contract with the room (see Figure 1), so any strategy that is reliant on the size or distance of the bricks or tiles will be insensitive to changes in the scale of the room. This category of cues includes other types of information such as eye height above the ground plane, but for the sake of brevity we have grouped them all under the label 'texture-based' cues. The two types of cue are quite distinct. For example, 'texture-based' cues to distance are available to anyone watching a video of the images that a participant in the experiment receives whereas 'physical' cues are not. The latter requires a binocular view or proprioceptive information about the distance the participant has walked in order to provide information about the overall scale of the scene. Of course, certain cues to distance fall outside these two categories, e.g. those that are constant in all conditions, such as accommodation which is fixed in a head mounted display, but in this paper we have concentrated on the consequences for distance perception when 'physical' and 'texture-based' cues are varying. Only using virtual reality can these two types of cue to object location be effectively separated. Experiments in natural environments have been used to measure distortions of perceived space (Battro, Netto, & Rozestraten, 1976; Koenderink, van Doorn, & Lappin, 2000; Loomis, Da Silva, Philbeck, & Fukusima, 1996), but it requires the flexibility of immersive virtual reality to manipulate 'physical' and 'texture-based' cues independently.

While the most striking result in the expanding room is that people fail to notice a large change in size of the room, there is evidence that stereo and motion parallax cues nevertheless contribute to judgements of object size (Glennerster et al., 2006; Rauschecker et al., 2006). The size judgements participants make are somewhere between the predictions based on purely 'physical' and purely 'texture-based' cues. It has been suggested that a cue combination model, based on the relative reliability of each type of cue, may provide a good prediction of the size judgements at different viewing distances (Glennerster et al., 2006; Rauschecker et al., 2006). In the current paper, we have focussed on judgements of distance rather than size, although we assume that the two are related. Here, we have measured thresholds for both 'physical' and 'texture-based' cues and used them to predict biases in

Europe PMC Funders Author Manuscripts

distance judgements. The predictions were close to the data under a variety of conditions. In the Discussion, we consider the implications this has for models of 3D vision based on reconstruction.

General methods

Participants

Five observers participated (age 16 to 38). One was one of the authors (S1) and four were naïve to the purpose of the experiment (S2–S5). Participants had normal or corrected-to-normal vision (6/6 or better) and normal stereopsis (TNO 60 arcsec or better). Observers' participation in the experimental studies was approved by the University of Reading Research Ethics Committee.

Equipment

The virtual reality system comprised a head mounted display, a head tracker and a computer that generated appropriate binocular images according to the location and pose of the head. The Datavisor 80 (nVision Industries Inc, Gaithersburg, Maryland, USA) head mounted display unit presented separate 1280 × 1024 pixel images (interlaced) to each eye using CRT displays. In the experiments, each eye's image was 73 deg horizontally by 69 deg vertically with a binocular overlap of 38 deg giving a total horizontal field of view of 108 deg (horizontal pixel size 3.4 arcmin).

The location and pose of the head was tracked using a Vicon real time optical tracker with seven MX3 cameras (Vicon Motion Systems Ltd, Oxford, UK). The system provided 6 degrees of freedom information on the location and a nominal spatial accuracy of 0.1 mm. This information was then used to render images for the appropriate optic center location and display frustum of each eye's display, which was determined according to a calibration procedure described by Gilson, Fitzgibbon, and Glennerster (2008).

The image generator computer was a dual processor workstation with dual graphic cards which sent images simultaneously to the headset and to the operator's display console. The computer generated appropriate images for each eye at 60 Hz and the system had a total latency of two frames (34 ms).

Stimulus and task

The participant was surrounded by a virtual room with brick textured wall, black and white checker board floor and a uniform gray ceiling (see Figure 1). In all the experiments, the participant's task was to judge whether a comparison square displayed in the second interval was closer or farther away than a reference square displayed in the first interval. While the room was always the same in the first interval, it generally had a different size in the second. The participant was asked to make judgements in terms of either the perceived 'absolute' distance (Experiments 1a and 2a) or the distance of the two squares relative to the room (Experiments 1b and 2b). The participants knew that the reference square was always presented at the same distance and that the room could change size between intervals.

Each run began with the participant in a virtual wire frame room that had physical dimensions of 3.0 × 3.5 × 3.1 m (width similar × depth × height), to the real room. When the participant pressed the button to initiate the trial, the brick room illustrated in Figure 1 appeared (Interval 1). The location of their cyclopean point was recorded at this moment, setting the floor height in the first interval appropriate for that participant and setting the height at which the square targets were displayed (eye height).

Both the reference and the comparison squares were red, uniformly lit and had a constant angular size (5.7 deg). Target position was defined relative to a point, T_0, at which participants entered a small 'trigger' zone (a tall invisible cylinder of 10 cm radius positioned half way between the side walls of the room). In the first interval, the reference square was placed directly in front of T_0 at a distance of 1, 3 or 5 m on a line through T_0 and perpendicular to the back wall. The comparison square was presented in the second interval at a distance assigned by a staircase procedure (see below). Additionally, the comparison square was given a random lateral jitter (±3, 6, 9 and 0 deg) to avoid the participant being able to solve the task from one single monocular view.

The participant was instructed to move from side to side in order to gain an impression of the distance to the square. The square was visible only when the participant's head (cyclopean point) was within a viewing zone of ±1.25 m laterally and ±0.5 m in depth with respect to T_0. So, the participant had to pass through a small trigger zone to initiate the display of the square and then keep within a larger viewing zone for the square to remain visible. However, a table in front of the participant provided a physical restriction. The participant was asked to keep close to the table during experiments so that in practice the range of forward and backward movement with respect to T_0 was small (see below).

In the first interval, the room always had the same dimensions (standard sized room, 5.0 × 6.4 × 3.1 m). When the room changed size between intervals, it did so in such a way that there was no information about the scale change as viewed from the cyclopean point. The cyclopean viewpoint, relative to the virtual room, was identical at the start of the first interval and the start of the second interval, independent of the scale of the room in the second interval. The expansion/contraction occurred in all directions so that the room became wider, deeper and taller for expansion factors greater than 1. For example, the dimensions of a room expanded two-fold were 10.0 × 12.8 × 6.2 m. The texture of the room was scaled with the room, so that there was the same number of bricks on the walls and tiles on the floor in both intervals. The key constraint on the expansion was as follows. The first trial began when the participant entered the trigger zone and the second trial began when the participant re-entered the trigger zone after the ISI. On any given trial, the difference in real-world location of the observer at the start of the first interval and the start of the second interval may in theory have been as large as 20 cm but, critically, we arranged that the alignment of the real and virtual worlds was such that the cyclopean viewpoint in the virtual world was identical at both instants.

These spatial constraints imposed limitations on the timing of the intervals. The first interval lasted for at least 2 s, ending when the participant's head entered the trigger zone. The ISI was at least 500 ms, likewise ending when the participant re-entered the trigger zone. The second interval lasted for exactly 2 s. In all the experiments, participants developed a rhythm when moving from side to side throughout the trials, so in practice the intervals were close to the minimum periods. Within the constraints of the viewing zone, participants were allowed to move with an amplitude and frequency that they found comfortable. Recordings from two participants showed that the typical lateral movement during trials was around 0.65 m. The standard deviation from the mean position in a lateral direction was 0.21 and 0.28 m for S1 and S2, respectively, and 0.02 and 0.04 m in the forward-back direction. The frequency of the movement was typically around 0.4 and 0.45 Hz for S1 and S2, respectively. Interval 1 lasted for 2 s plus an additional time period of 0.5 ± 0.11 s and 0.2 ± 0.04 s for S1 and S2, respectively, while the ISI lasted for 0.5 s plus an additional 0.7 ± 0.09 s and 0.6 ± 0.04 s.

Psychometric procedure

The reference square was presented at three different viewing distances in separate runs (1, 3 and 5 m). In Experiment 1, measuring biases for distance matches, the expansion factor of the scene between intervals took one of five values (0.25, 0.5, 1, 2 and 4) making five independent psychometric functions that were randomly interleaved in one run of trials. Each psychometric function consisted of 100 trials, giving a total of 500 trials for each run. In Experiment 2, measuring thresholds for the detection of distance changes, the expansion of the scene was chosen using a staircase procedure with a single psychometric function of 200 trials in one run. Runs could be spread over several sessions and observers were encouraged to have breaks around every 100–150 trials.

The distance of the comparison square presented in the second interval was chosen from a standard staircase procedure. The staircase used was based on Cornsweet's method (Glennerster et al., 2006; Johnston et al., 1993; Rauschecker et al., 2006). The initial staircases were set so as to include both matches for physical distance and a texture-based match. In addition, the staircases were clamped so that the comparison square was never shown behind the back wall.

The proportion of trials on which the comparison was judged as 'farther away' was plotted as a function of disparity. The resulting psychometric function was fitted with a cumulative Gaussian by probit (Finney, 1971). Thresholds were defined as $1/\sqrt{2}$ times the standard deviation of the fitted cumulative Gaussian because we used a 2IFC paradigm (Green & Swets, 1974). In the distance matching task, distance matches show the bias (point of subjective equality, PSE, or 50% point). In each case, error bars show the standard error of the mean. Both thresholds and biases are given in arcmin.

The fact that we are using disparity as a measure of the difference in distance between the reference and comparison squares does not imply that binocular disparity is the only, or even the most important, cue in the experiment. It can also be used as a measure of the motion parallax cue assuming (simply for the purpose of comparison with binocular disparity) a baseline of 6.5 cm, as we do here. Thus, for example, biases of ±30 arcmin for a reference at 1 m correspond to matched distances of 0.9 and 1.2 m; for a reference at 3 m they correspond to matches at 2.1 and 5 m and for a reference at 5 m the ±30 arcmin matches are at 3 and 15 m.

Results

In Experiment 1 we measured biases in a distance matching task. In Experiment 2 we measured 'physical' and 'texture-based' thresholds (see Table 1) and determined whether these could be used to predict the biases in Experiment 1.

Experiment 1: Measurements of biases when a room changes size

In Experiment 1 we measured biases in a distance matching task. Here, the participant was asked to make judgements in terms of the perceived 'absolute' distance (Experiment 1a) or the distance of the two squares relative to the room (Experiment 1b). While the room was always the same size in Interval 1, the expansion factor of the scene between intervals took one of five values (0.25, 0.5, 1, 2 and 4).

Experiment 1a: Matching perceived absolute distance

In this experiment, the participant's task was to compare the perceived absolute distance of the reference and comparison squares. Figure 2 shows biases in the matching task plotted against expansion factor of the room for an individual participant (Figure 2a) and for three

other participants (Figure 2b). The horizontal line represents a pure 'physical' match as specified by stereo and motion parallax cues. The dashed curves show the prediction of a strategy purely based on texture-based cues: it plots the vergence angle change that would occur if participants fixated on the reference square before and after a rigid expansion of the room (including the reference square). Biases are given in arcmin (see General methods), where positive values correspond to matches that are farther from the participant than the reference distance.

As expected (Rauschecker et al., 2006), matched distances were closer to the participant for small expansion factors and farther away for large expansion factors, reflecting an influence of the expanding room. The prediction of a strategy purely based on texture-based cues is shown by the dashed curves in Figure 2. The data generally fall between the predicted pattern for using 'physical' cues (dashed horizontal line) and that for 'texture-based' cues (dashed curve). The data can be modeled by a linear combination of these two cues: $bias = k\,f(g) + c$, where g is the expansion factor, $f(g)$ is the bias predicted for a pure texture-based match as shown by the dashed curve, and k and c are free parameters. A weighted linear least squares fit was used to find the values of k and c and their covariance. The standard deviation on k was taken to be the corresponding value from the covariance matrix. Figure 2 shows the fits derived in this way.

As discussed in the General methods, the staircase was clamped so that stimuli were never presented beyond the back wall, raising the question of whether participants' true biases have been measured. However, in practice, apart from one participant in one condition (S2 at 5 m), the back wall was always outside the 95% confidence interval of participants' PSE.

Experiment 1b: Matching perceived distance relative to the room

In the above experiment, the participant's task was to judge the change in the perceived 'absolute' distance of the target. Here, the participant was asked to judge whether the comparison square in the second interval was closer or farther away relative to the surrounding room. The experiment was otherwise exactly the same as in Experiment 1a. Figure 3a shows data for the same participant as shown in Figure 2a. For this participant, at 1 and 3 m viewing distances, the pattern of biases are clearly altered towards a more texture-based match. The results for participant S1, who is an author, are not typical. Other participants' distance matches were barely altered by the change in task as shown in Figure 3b.

Figure 3b plots the fitted values of k for judgements of perceived distance relative to the room (k_{rel}) against values of k for judgements of perceived absolute distance (k_{abs}). For both axes, data close to zero imply the use of physical cues, and data close to 1 the use of texture-based cues. Data are clustered by viewing distance (1, 3 and 5 m), where k_{abs} is about 0.1, 0.5 and 0.9, respectively. In almost all cases, the data lie above the line of equality, meaning that there was a shift towards using texture-based cues for the 'relative' task, although for most participants the difference is small. Participant S1, whose data are shown by circles, is more able to make a texture-based match in the 'relative' task, as we saw in Figure 3a.

Experiment 1c: Matching perceived distance while varying the 'texture-based' cue

To explore how proximity to other objects influenced the distance judgements, the target was positioned so as to abut the left wall. The added texture-based cues were expected to have the greatest impact for the nearest viewing distance so here the reference square was always placed at 1 m. An object placed adjacent to the wall should provide the greatest propensity to use the room as a reference frame, while the stereo and motion parallax cues should be of the same magnitude as in the previous experiments. To facilitate the use of

texture-based cues, both reference and comparison objects were modified to be horizontal rectangles piercing the wall on the left side.

The room was shifted so that the target and the center of the viewing zone were 0.5 m from the left hand wall. The rectangle was generated by creating a square stimulus (for either reference or comparison) in the same way as for the previous experiment but extending the square at the same depth until it reached the target wall. Thus, the height of the rectangle was 5.7 deg, the right hand edge of the target rectangle behaved in the same way as in the previous experiment and the left hand edge pierced the wall. In this case, the retinal width of the rectangle always co-varies with target distance relative to the room but not with expansion factor. Between intervals, the structure of the bricks was randomly changed so that it would not be possible to use a particular brick as a landmark to judge the change in distance to the target rectangle.

Figure 3c plots values for best fit for the 'absolute' and 'relative' tasks (k_{abs} and k_{rel}) where a value of $k = 0$ corresponds to the 'physical' cue prediction and $k = 1$ corresponds to the 'texture-based' prediction. It shows values of k when the target was placed close to the wall (open symbols) and away from the wall (closed symbols). All participants shifted their biases toward a more texture-based match when the target was close to the wall for both the absolute and relative matching tasks. The mean values of k_{abs} shifted from 0.08 ± 0.002 to 0.42 ± 0.005, while the mean values for k_{rel} shifted from 0.22 ± 0.005 to 0.83 ± 0.008.

Experiment 2: Measurements of thresholds

In Experiment 2 we measured 'physical' and 'texture-based' thresholds (see Table 1 for definitions) and determined whether these could be used to predict the biases in Experiment 1. Here, the expansion of the scene was chosen using a staircase procedure.

Experiment 2a: Measurements of 'physical' thresholds

The goal was to measure thresholds for detecting the expansion or contraction of the room from 'physical' cues alone. There were no 'texture-based' cues because the square and room were rigidly connected, expanding and contracting together. Hence, there were no cues to distance that could be determined purely from the images the participant received unless the participant combined these with information about the interocular separation or the distance moved. The top left panel of Figure 4 shows this relationship, where every point on the walls and the floor of the room moves farther away from the observer in the second interval (solid line). Just as in Experiment 1a, participants were asked to judge whether the square in the second interval (comparison square) appeared physically closer or farther away than the reference, i.e. ignoring the relation of the squares to the room.

Figure 4, left column, shows thresholds for four participants for each of the three viewing distances (closed symbols). Thresholds are given as a Weber fraction, i.e. threshold, in arcmin, as a proportion of the vergence angle when fixating the reference square. Two participants also carried out the experiment without the surrounding brick room. Their data show a similar effect of viewing distance, albeit with higher thresholds (open symbols). Similar data for a vertical single line stimulus can be found in Appendix A.

The dashed line shows the threshold when participants were asked to judge whether the room had become larger or smaller between intervals, without any square present. It is perhaps surprising that the threshold in this condition does not provide a ceiling to the other thresholds since the information available in this condition is always present. However, the nature of the task is rather different in the two cases, with participants' attention focused on

Europe PMC Funders Author Manuscripts

the perceived distance of the square in one case and on the overall size of the room in the other.

Experiment 2b: Measuring 'texture-based' thresholds

In order to measure 'texture-based' thresholds for discriminating the distance of the square relative to the room (Experiment 2b), we kept the distance of the square fixed on every trial so that the physical cues for the square remained constant throughout the run (see top middle panel Figure 4). However, the change in size of the room between intervals provided information about the target square relative to the room. We refer to this information as a texture-based cue because it is independent of the overall scale of the room.

The magnitude of the thresholds for the 'texture-based' cue needs to be determined in relation to a constant sized room (since this cue is independent of the overall scale of the room). To achieve this, thresholds derived from changes in room size, ΔD_2 (see Figure 4), were scaled by the ratio of the distance to the reference square and the distance to the back wall. Figure 4, middle column, shows these thresholds for detecting 'texture-based' cues. The results show the opposite effect compared to the 'physical' thresholds, with poorest thresholds for 1 m. This is likely to be because the texture-based cues were more reliable as the square moved closer to the back wall (see the Discussion).

Experiment 2c: Measuring thresholds while varying the 'texture-based' cue

The purpose here was to measure thresholds for predicting biases in distance judgements for a target placed close to the wall and so the virtual environment was similar to the scene in Experiment 1c.

Figure 5 shows how the thresholds for physical and texture-based cues changed as the target was moved to the position close to the wall for four participants at 1 m (S1–S3, S5). The horizontal axis re-plots thresholds when the target was placed away from the wall and the vertical axis shows thresholds when the target was close to the wall.

As before, thresholds were measured both for detecting 'physical' cues (black symbols) and 'texture-based' cues (gray symbols). Thresholds for discriminating distance for 'texture-based' cues were always lower when the target was presented close to the wall (gray symbols lie below the line of equality). Thresholds based on 'physical' cues were always *worse* when the target was close to the wall (black symbols above the line), which might have been due to the lateral jitter having a detrimental influence in this case.

Predicting biases

The thresholds from the two tasks in Experiment 2 can be used to predict the biases in the distance matching tasks if we assume that information from the two cues is combined optimally (Johnston et al., 1993; Knill & Saunders, 2003; Landy et al., 1995; Young, Landy, & Maloney, 1993). Specifically, perceived distance of the target D is given by:

$$D=w_P P+w_T T, \quad w_p+w_T=1. \quad (1)$$

where P is the estimate of target distance derived from stereo/motion parallax cues, T from texture-based cues and w_P and w_T are the weights applied to these estimates, respectively.

The model assumes that noises on each of the estimates, P and T, are independent and Gaussian with variances σ_p^2 and σ_T^2. Then, the weights can be estimated by the following equations:

$$w_P = \frac{1/\sigma_P^2}{1/\sigma_P^2 + 1/\sigma_T^2}, \quad w_T = \frac{1/\sigma_T^2}{1/\sigma_P^2 + 1/\sigma_T^2}. \quad (2)$$

We take the thresholds for distance judgements with the use of physical and texture-based cues (Experiment 2) as estimates of σ_P and σ_T respectively. Figure 4, right hand column, plots the weights, w_P and w_T, derived from these thresholds for each viewing distance. As we have noted, participants placed more weight on physical cues at 1 m and on texture-based cues at 5 m. At 3 m most participants had a slightly greater weighting for texture-based cues.

Just as for Figure 2, we found the bias, b, by fitting the function $b = k\,\mathrm{f}(g) + c$, with k and c as free parameters. We can now see how well the weights, w_P and w_T, predict the bias data. The equation for the predicted curve is $b = w_T\,\mathrm{f}(g)$. This can be done for both k_{abs} which applies to the absolute task and k_{rel} which applies to the relative task. Figure 6 shows how k_{abs} and k_{rel} relate to the predicted values, w_T, for judgements made in the middle of the room. In both cases there is a strong correlation between the prediction w_T and the best fitting k value (k_{abs}: $r(11) = 0.98$, $p < 0.001$; k_{rel}: $r(11) = 0.94$, $p < 0.001$). The fits, k_{abs} and k_{rel}, span a narrower range than the predicted values, w_T, for which we have no clear explanation. When the target was close to the wall, participants reported using a variety of strategies to carry out the task, leading to variability in performance and a poor correlation between predictions, w_T, and fits, k_{abs} and k_{rel}. For comparison with Figure 3c, the mean value of w_T close to the wall was 0.45 ± 0.28 and away from the wall it was 0.019 ± 0.018.

Discussion

As discussed in the Introduction, it is not obvious, *a priori*, that cue combination rules should apply to the perception of distance but here we have shown that they do. Using only measurements of (i) thresholds for judging the physical distance of a target (independent of neighboring objects) and (ii) thresholds for judging target distance relative to surrounding objects, it is possible to make quite accurate predictions, with no free parameters, about the biases that people will make when judging the perceived distance of an object (Figure 6). We have shown that the ability to predict biases on distance judgements holds over a range of conditions in which the reliability of physical and texture-based cues is varied, including the effect of viewing distance (Figure 4) and proximity to other objects (Figure 5).

Under natural viewing conditions, the two classes of cue that we manipulated provide consistent information about object distance but, using virtual reality, we could investigate their effects independently. ‘Physical’ cues, which include stereo and motion parallax, provide information about the distance of an object independent of other objects in the scene. ‘Texture-based’ cues indicate where an object is relative to others in the scene. In our experiment, when the room changed size between intervals, there were large changes in physical cues while changes in texture-based cues were small or absent. Virtual reality does not, as yet, provide realistic variations in some depth cues, such as accommodation, which is known to have an influence on depth judgements (Watt, Akeley, Ernst, & Banks, 2005). Nevertheless, we have assumed that any other cues to distance that remain fixed throughout the experiment will not affect our conclusions about the relative weights applied to ‘physical’ and ‘texture-based’ cues.

There is a debate about whether information from these two types of cue can be accessed independently or whether they are fused ‘mandatorily’ as others have suggested for stereo and texture cues (Hillis et al., 2002). Rauschecker et al. (2006) suggested that there might be

Europe PMC Funders Author Manuscripts

mandatory fusion of physical and texture-based cues because participants were unable to use feedback appropriately to change their responses when the feedback indicated the physical distance of the target object. On the other hand, participants *could* use feedback effectively if it indicated the location of the target object relative to the room. Intuitively, this result seems reasonable, i.e. observers should be able to judge the distance of an object relative to the room (ignoring its absolute distance). However, one of the interesting results from the experiments we report here is that, without feedback, participants do not report the relative location of the target object at all accurately. Indeed, they show almost as much 'mandatory fusion' of cues for this task as when they are asked to report the perceived absolute distance of the target (Figure 6b).

The tendency toward fusion is all the more remarkable in the light of the large conflict between cues. Hillis et al. (2002) found fusion of visual cues when both were close to their discrimination threshold but a breakdown of mandatory fusion when either cue was above this level. In our stimuli, on the other hand, the cue conflict could be dramatic. As results in Figure 4 show, the largest changes in room size (400%) are far greater than the discrimination threshold for the physical cue presented in isolation (6–25% Weber fractions at 1 m), yet still the matching data (Figure 6) are consistent with cue fusion. We found one exception to this rule. Participant S1 (Figures 3 and 6) showed evidence of being able to attend to the texture-based distance information when the task demanded it, in line with the conclusions of Rauschecker et al. However, this participant was one of the authors and clearly knew the purpose of the experiments. It is perhaps more remarkable that she was unable to make responses that were close to the correct, texture-based match.

It might seem contradictory that the thresholds reported here are quite low while at the same time participants 'failed to notice' a change of 400% in room size in the distance matching task. In fact, there is no contradiction. Participants can carry out the threshold task in a number of ways without necessarily perceiving a change in the size of the room, for example by noticing how slowly they appear to move through the room when its size is increased. Nevertheless, participants do report using perceived distance of the target square as a cue to help them carry out the threshold task for detecting physical cues, particularly when the target was at 1 m where thresholds are lowest. Clearly, failing to notice the change in size of the room is partly a question of attention akin to the phenomenon of 'change blindness' (Rensink, 2002), or even blindsight (Stoerig & Cowey, 2007). Our measurements of thresholds show that large and small rooms are not indistinguishable metamers (Backus, 2002), but it is still remarkable how participants report no apparent change in room size when the actual change is several times the threshold level measured from a forced-choice judgement.

Although we have not found, in the literature, directly comparable thresholds for perception of distance from motion parallax, there are related results for binocular stimuli. For example, the thresholds for an absolute disparity judgement reported by Westheimer (1979), if expressed as a fraction of the vergence angle, were about 1% in the worst case, which is still considerably better than we found (Figure 4). Differences in the type of display, the observer being static or moving and the ISI (200 ms in Westheimer, 1979) are all likely to contribute to the difference in performance. Similarly, our task for measuring the texture-based thresholds can be compared to a stereo task in which target depth was judged relative to a reference at a quite different disparity (McKee, Levi, & Bowne, 1990). They found a linear increase as the disparity of the pedestal increased. The thresholds we found for texture-based cues followed a similar pattern when the distance between the target and the back wall was increased (Figure 4) although, again, the magnitude of the thresholds in our study was considerably higher. As we have discussed, all virtual reality setups fall short of natural viewing conditions and some concerns, notably over biases in distance judgements, have

been discussed extensively (Bingham, Bradley, Bailey, & Vinner, 2001; Knapp & Loomis, 2004; Willemsen, Colton, Creem-Regehr, & Thompson, 2004). In fact, such biases are much less pronounced in our lab (Tcheang, Gilson, & Glennerster, 2005) where we have a wide field of view and accurate calibration (Gilson et al., 2008). In general, although we recognize that virtual reality brings some inherent limitations to performance, we have based our conclusions in this paper on comparisons of data collected under similar conditions rather than making a direct comparison with natural viewing conditions.

The cue combination that we have described here implies a very different approach to scene representation than one based on geometric reconstruction. For example, the texture-based cue depends on a comparison of one visual stimulus with another viewed several seconds ago, leading to inevitable biases in this task. This strong dependence on the past is not a feature of all tasks, and one would expect tasks that can be done purely on the basis of current visual information to suffer less, or not at all, from the expansion or contraction of the room. An explanation based on 3D reconstruction, on the other hand, seeks to explain performance in all tasks on the basis of a single internal model, albeit one that may be distorted.

Conclusions

We have shown that a cue combination approach can be applied to judgements of distance such that, under a range of conditions, changes in thresholds of different cues cause concomitant changes in the perceived distance of objects.

Acknowledgments

This research was supported by the Wellcome Trust, the University of Reading and Buskerud University College. We are grateful to the anonymous reviewers for helpful comments and to Lyndsey Pickup for statistical advice.

Appendix A

Thresholds for a single line target

Figure A1 shows thresholds for a single line target, collected under identical conditions to those for a single square target (Figure 4, left hand column, open symbols). The line was one pixel wide regardless of viewing distance, effectively infinitely tall (the top and bottom of the line were not visible) and placed directly in front of the observer in Interval 1. The participant’s task was to judge whether

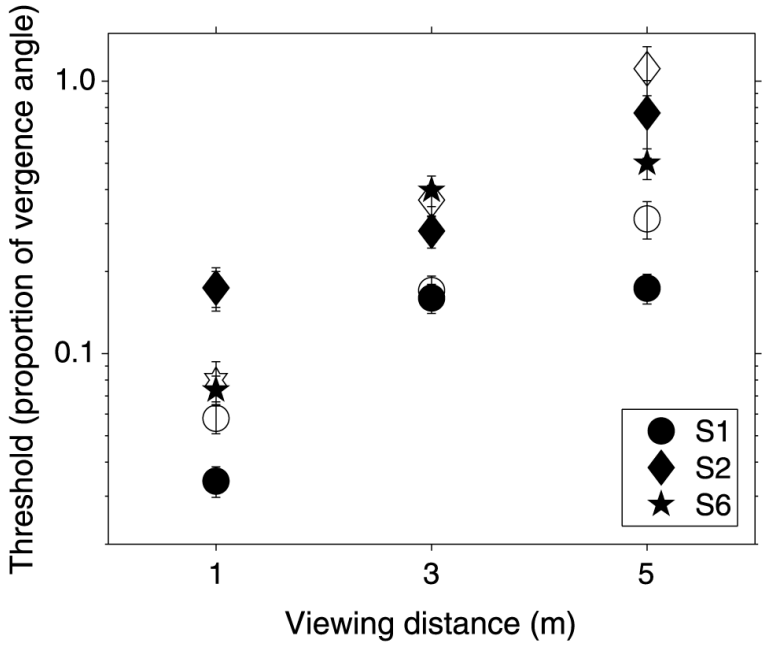

Figure A1.
Thresholds for detection of change in distance of a single target without a surrounding scene. Closed symbols show thresholds for three participants for discriminating the distance of a vertical line at viewing distances of 1, 3 and 5 m. Open symbols show data for a square presented on its own. For participants S1 and S2, the open symbol data are re-plotted from the left hand column in Figure 4.
the line in Interval 2 was closer or farther away (random lateral jitter ±1 deg). In this condition, the participant could not use image deformation of the square as a potential cue to determine the distance of the target. Thresholds in the two conditions are very similar, suggesting that any deformation cue did not contribute significantly to the thresholds shown in Figure 4.

References

Backus, BT. Perceptual metamers in stereoscopic vision. In: Dietterich, TG.; Becker, S.; Ghahramani, Z., editors. Advances in neural information processing systems. 14th ed.. MIT Press; Cambridge, USA: 2002. p. 1223-1230.

Battro AM, Netto S. d. P. Rozestraten RJA. Riemannian geometries of variable curvature in visual space: Visual alleys, horopters, and triangles in big open fields. Perception. 1976; 5:9–23. [PubMed: 958853]

Bingham GP, Bradley A, Bailey M, Vinner R. Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. Journal of Experimental Psychology: Human Perception and Performance. 2001; 27:1314–1334. [PubMed: 11766927]

Brenner E, van Damme WJ. Perceived distance, shape and size. Vision Research. 1999; 39:975–986. [PubMed: 10341949]

Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002; 415:429–433. [PubMed: 11807554]

Finney, DJ. Probit analysis. 3rd ed.. CUP; Cambridge: 1971.

Foley JM. Binocular distance perception. Psychological Review. 1980; 87:411–434. [PubMed: 7413886]

Foo P, Warren WH, Duchon A, Tarr MJ. Do humans integrate routes into a cognitive map? Map-versus landmark-based navigation of novel shortcuts. Journal of Experimental Psychology: Human Learning and Memory. 2005; 31:195–215. [PubMed: 15755239]

Gilson SJ, Fitzgibbon AW, Glennerster A. Spatial calibration of an optical see-through head-mounted display. Journal of Neuroscience Methods. 2008; 173:140–146. [PubMed: 18599125]

Glennerster A, Hansard ME, Fitzgibbon AW. Fixation could simplify, not complicate, the interpretation of retinal flow. Vision Research. 2001; 41:815–834. [PubMed: 11248268]

Glennerster A, Hansard ME, Fitzgibbon AW. View-based approaches to spatial representation in human vision. Lecture Notes in Computer Science. 2009; 5064:193–208.

Glennerster A, Rogers BJ, Bradshaw MF. Stereoscopic depth constancy depends on the subject's task. Vision Research. 1996; 36:3441–3456. [PubMed: 8977011]

Glennerster A, Tcheang L, Gilson SJ, Fitzgibbon AW, Parker AJ. Humans ignore motion and stereo cues in favor of a fictional stable world. Current Biology. 2006; 16:428–432. [PubMed: 16488879]

Gogel WC. A theory of phenomenal geometry and its applications. Perception & Psychophysics. 1990; 48:105–123. [PubMed: 2385484]

Green, DM.; Swets, JA. Signal detection theory and psychophysics. Robert E. Krieger Publishing Company; Huntington, New York: 1974.

Hartley, R.; Zisserman, A. Multiple view geometry in computer vision. Cambridge University Press; Cambridge, UK: 2000.

Hillis JM, Ernst MO, Banks MS, Landy MS. Combining sensory information: Mandatory fusion within, but not between, senses. Science. 2002; 298:1627–1630. [PubMed: 12446912]

Hillis JM, Watt SJ, Landy MS, Banks MS. Slant from texture and disparity cues: Optimal cue combination. Journal of Vision. 2004; 4(12):1, 967–992. http://journalofvision.org/4/12/1/, doi: 10.1167/4.12.1. [PubMed: 14995894]

Johnston EB, Cumming BG, Parker AJ. Integration of depth modules: Stereopsis and texture. Vision Research. 1993; 33:813–826. [PubMed: 8351852]

Knapp JM, Loomis JM. Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. Presence-Teleoperators and Virtual Environments. 2004; 13:572–577.

Knill DC, Saunders JA. Do humans optimally integrate stereo and texture information for judgments of surface slant? Vision Research. 2003; 43:2539–2558. [PubMed: 13129541]

Koenderink JJ, van Doorn AJ, Lappin JS. Direct measurement of the curvature of visual space. Perception. 2000; 29:69–79. [PubMed: 10820592]

Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: In defense of weak fusion. Vision Research. 1995; 35:389–412. [PubMed: 7892735]

Longuet-Higgins HC. A computer algorithm for reconstructing a scene from two projections. Nature. 1981; 293:133–135.

Loomis JM, Da Silva JA, Philbeck JW, Fukusima SS. Visual perception of location and distance. Current Directions in Psychological Science. 1996; 5:72–77.

Luneburg R. The metric of binocular visual space. Journal of the Optical Society of America. 1950; 40:627–642.

McKee SP, Levi DM, Bowne SF. The imprecision of stereopsis. Vision Research. 1990; 30:1763–1779. [PubMed: 2288089]

Rauschecker AM, Solomon SG, Glennerster A. Stereo and motion parallax cues in human 3D vision: Can they vanish without a trace? Journal of Vision. 2006; 6(12):12, 1471–1485. http://journalofvision.org/6/12/12/, doi:10.1167/6.12.12. [PubMed: 17209749]

Rensink RA. Change detection. Annual Review of Psychology. 2002; 53:245–277.

Stoerig P, Cowey A. Blindsight. Current Biology. 2007; 17:R822–R824. [PubMed: 17925204]

Tcheang L, Gilson SJ, Glennerster A. Systematic distortions of perceptual stability investigated using immersive virtual reality. Vision Research. 2005; 45:2177–2189. [PubMed: 15845248]

Watt SJ, Akeley K, Ernst MO, Banks MS. Focus cues affect perceived depth. Journal of Vision. 2005; 5(10):7, 834–862. http://journalofvision.org/5/10/7/, doi:10.1167/5.10.7. [PubMed: 16441189]

Westheimer G. Cooperative neural processes involved in stereoscopic acuity. Experimental Brain Research. 1979; 36:585–597. [PubMed: 477784]

Willemsen, P.; Colton, MB.; Creem-Regehr, SH.; Thompson, WB. The effects of head-mounted display mechanics on distance judgments in virtual environments; Proceedings of the First Symposium on Applied Perception in Graphics and Visualization; 2004; p. 35-48.

Young MJ, Landy MS, Maloney LT. A perturbation analysis of depth perception from combinations of texture and motion cues. Vision Research. 1993; 33:2685–2696. [PubMed: 8296465]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

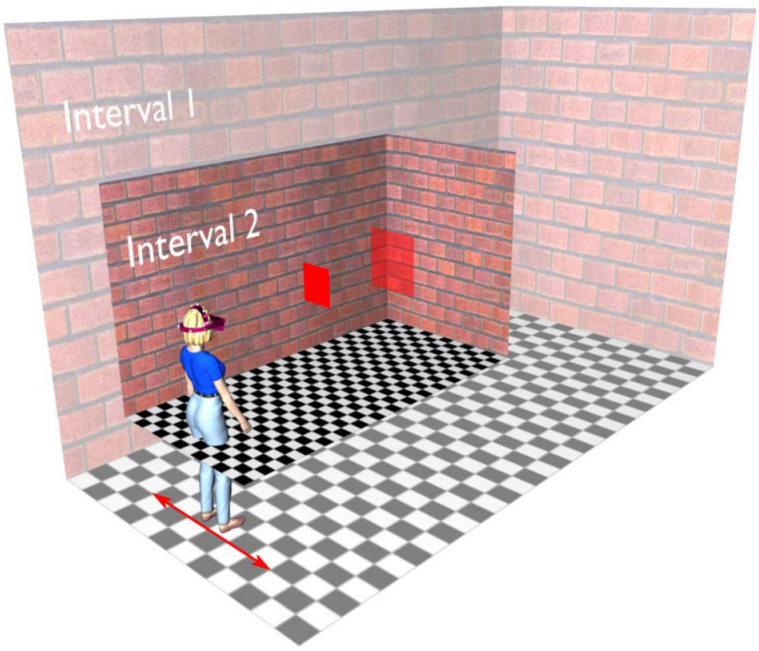

Figure 1.
The virtual scene. The semi-transparent and opaque parts of the figure represent the scene in Intervals 1 and 2, respectively. The participants were required to move from side to side to generate motion parallax (red arrow). In a 2IFC task, participants compared the distance to two squares presented in separate intervals.

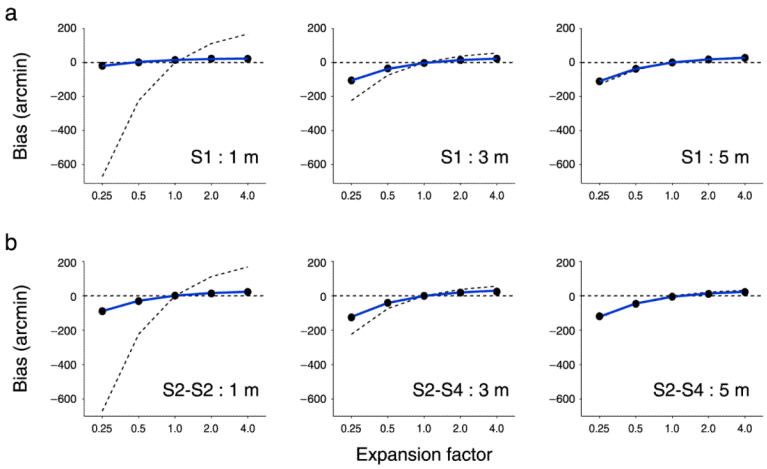

Figure 2.

Distance matching results from Experiment 1a. Distance matches (PSE) are plotted against the expansion factor of the virtual room for three reference distances (1, 3 and 5 m, in left, middle and right hand column, respectively). Results are shown for one participant (S1) in a) and a mean for the other participants (S2–S4) in b). Matches are plotted as biases, in arcmin, relative to a correct distance match (see General methods). Standard errors were smaller than the size of the symbols; on average they were around 3 arcmin. The horizontal line represents a pure physical match as specified by stereo and motion parallax cues and positive values correspond to matches that are farther away from the reference distance than the participant. The dashed curves show the prediction of a strategy purely based on texture-based cues. The blue line shows the linear combination of the above two predictions that best fits the data in each case.

Europe PMC Funders Author Manuscripts

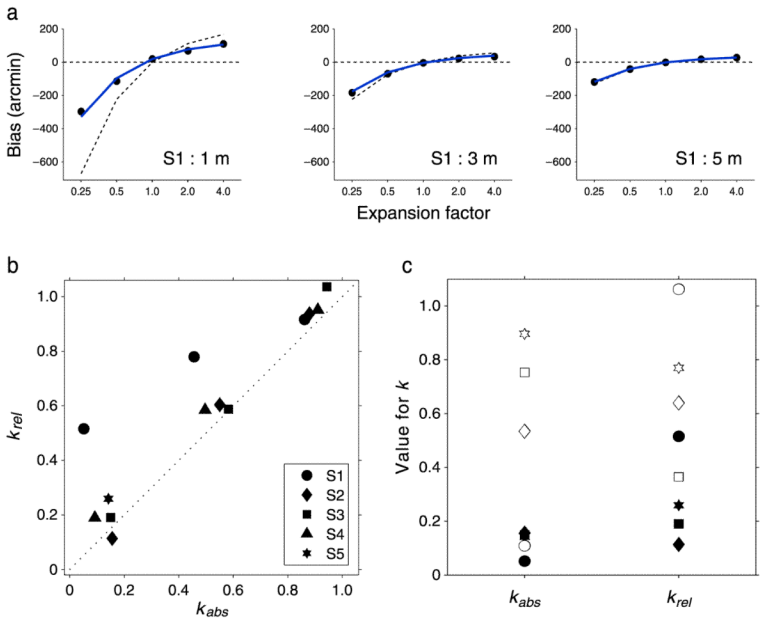

Figure 3.

Distance matching from the 'relative' task in Experiment 1b. The experiment was the same as for Figure 2, except that participants were asked to judge the distance of the reference and comparison squares in relation to the room. a) Data for the same participant (S1) as shown in Figure 2a. For this participant, the change in task makes a substantial difference to the matches. b) Values of the weighting parameter for the 'relative' task (k_{rel}) plotted against those for the 'absolute' task (k_{abs}), shown for the four participants in Figure 2 and a fifth participant for a reference distance of 1 m. Participant S1 (circles) is atypical in showing a large effect of task. c) Data for the condition in which reference and comparison squares were presented close to the wall. Open symbols show the weighting parameter k_{abs} derived from the absolute task and k_{rel} for the relative task. Closed symbols show k_{abs} and k_{rel} re-plotted from b).

Europe PMC Funders Author Manuscripts

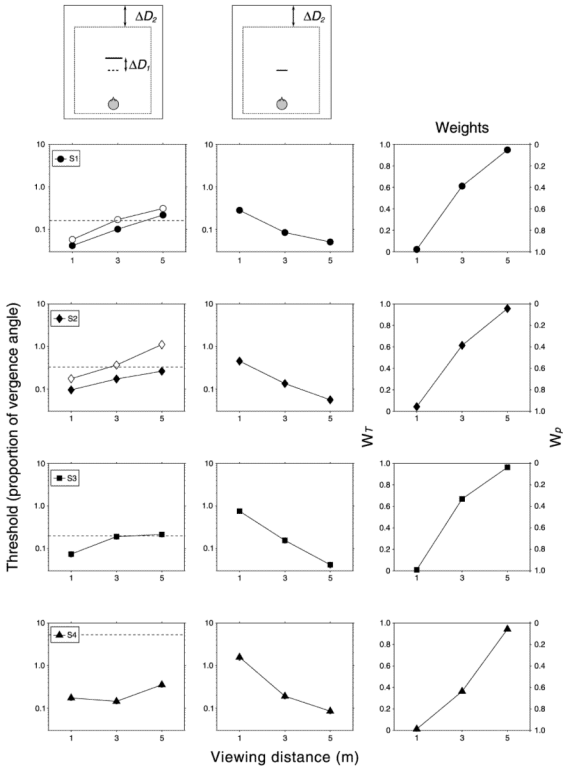

Figure 4.

Thresholds for detection of change in ‘physical’ and ‘texture-based’ cues. The left hand column shows thresholds for detecting a change in distance to the square when there is no change in texture-based cues (referred to as ‘physical’ threshold, see Table 1). The icon above shows the change in distance to the square (ΔD_1) and the proportional change in the room (ΔD_2 for the back wall). Thresholds are given as a Weber fraction defined as a proportional change in the vergence angle compared to fixating the reference square. Closed symbols indicate conditions in which the target was surrounded by the brick room. For two participants, open symbols show data for the target presented on its own. The dashed line shows the threshold for detecting a change in room size alone (without any square). The middle column shows thresholds when there was no change in location of the target square between intervals, only a change in its distance relative to the surrounding room (referred to as ‘texture-based’ threshold) as shown in the icon above. The right hand column shows weights for ‘physical’ and ‘texture-based’ cues derived from these thresholds (w_P and w_T respectively).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

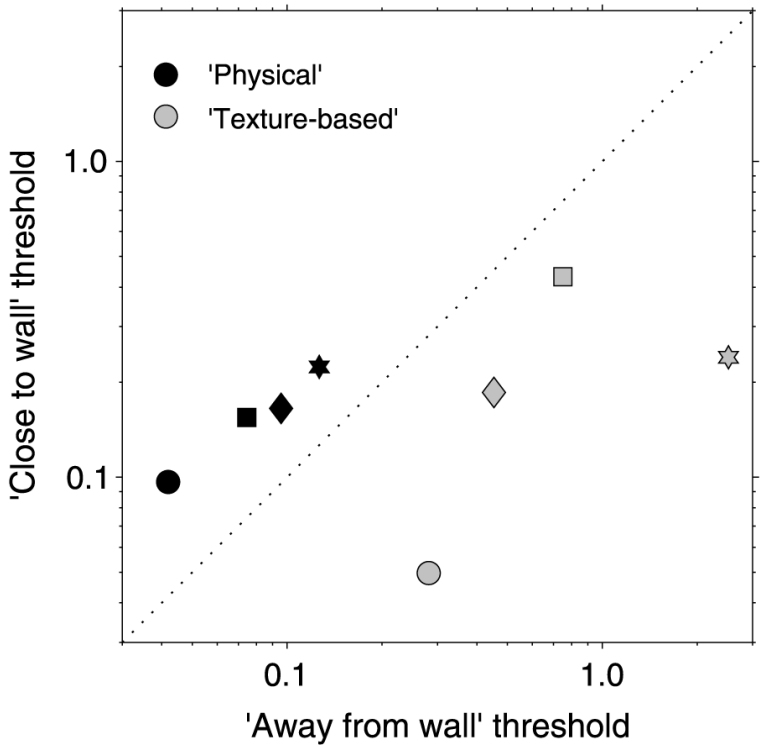

Figure 5.

The effect of target location. Thresholds measured when the target square was close to the wall are plotted against thresholds measured when the target was presented in the center of the room (re-plotted from Figure 4) for a reference distance of 1 m. As in Figure 4, thresholds are given as a proportion of vergence angle. Black symbols indicate thresholds for detecting a change in the ‘physical’ distance of the target and gray symbols indicate ‘texture-based’ thresholds. Symbols for the individual participants correspond to those in Figures 4 and 6.

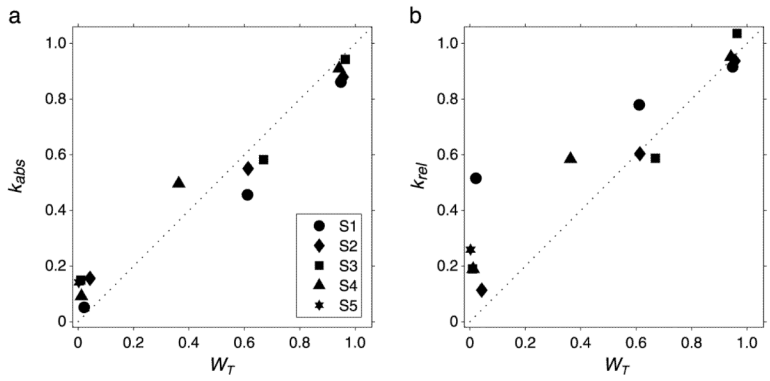

Figure 6.

Predictions and fits for distance matching results when the target was in the middle of the room. The fits to the matching data for both tasks (k_{abs} and k_{rel}, re-plotted from Figure 3b) are plotted against the predicted weight for texture-based cues, w_T (re-plotted from Figure 4). a) k_{abs} is the weighting determined by a best fit to the matching data when participants judged the perceived absolute distance of the target. b) k_{rel} is the weighting derived from judgements of target distance relative to the room.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Nomenclature used in the paper

Variable	Definition
σ_P, σ_T	Thresholds for detection of change in 'physical' cues (σ_P, Figure 4, column 1) or 'texture-based'cues (σ_T, Figure 4, column 2). The 'physical' cue signals the distance of the object independent of other objects in the scene (stereo and motion parallax). The 'texture-based' cue signals the distance to the object relative to others and is independent of the overall scale of the room.
W_P, W_T	Predicted weights for 'physical' and 'texture-based' cues derived from σ_P and σ_T, respectively (Equation 2).
k_{abs}, k_{rel}	Fitted parameters for the data on matching perceived absolute distance (k_{abs}, Figure 2) and perceived distance relative to the room (k_{rel}, Figure 3).

J Vis. Author manuscript; available in PMC 2010 July 01.