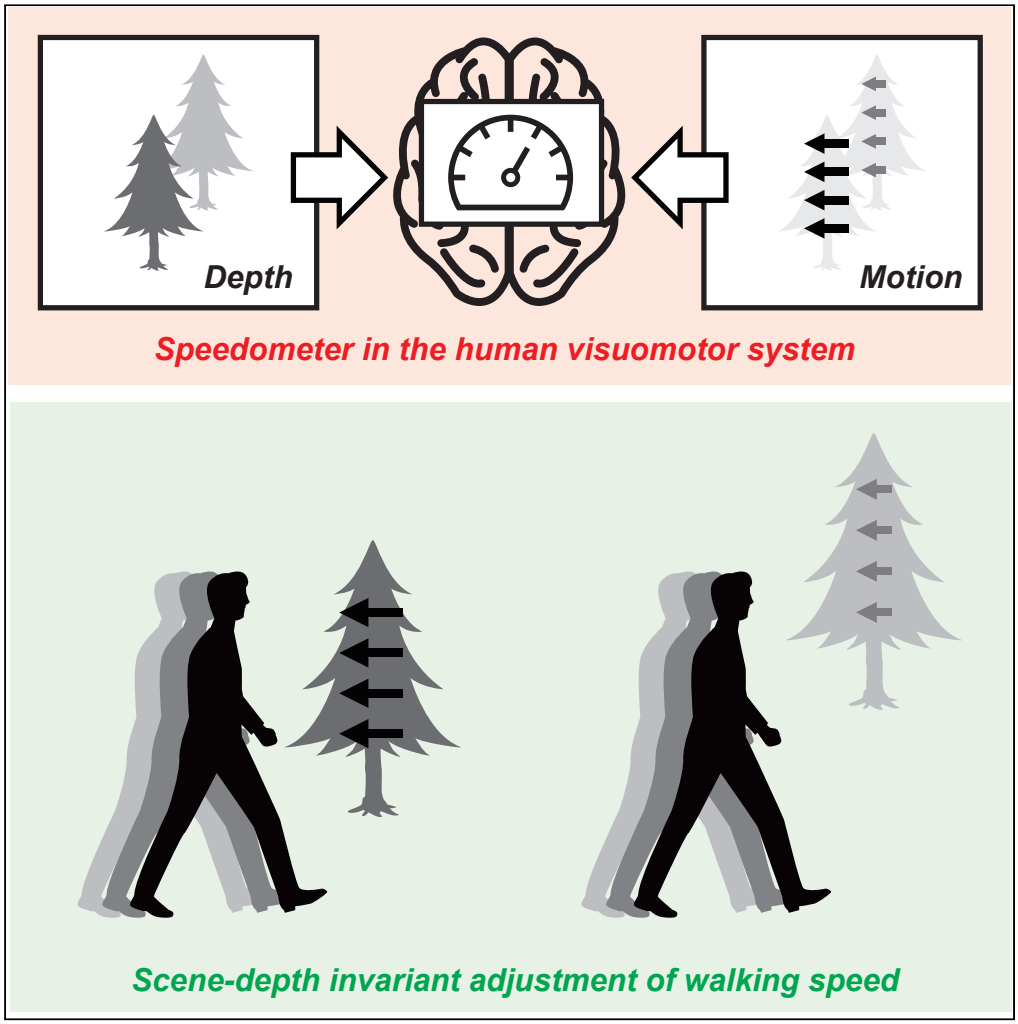
iScience

CellPress
OPEN ACCESS

Article

Vision-based speedometer regulates human walking



Shinya Takamuku, Hiroaki Gomi

shinya.takamuku.ka@hco.ntt.co.jp

Highlights

Changes in optic flow speed triggers implicit adjustments of walking speed

The response is invariant with respect to the depths of objects in the scene

The invariance is not explained by temporal-frequency-based speed coding

Both binocular and monocular depth cues contribute to the invariance

Takamuku & Gomi, iScience *24*, 103390
December 17, 2021
https://doi.org/10.1016/j.isci.2021.103390

iScience

CellPress
OPEN ACCESS

Article

Vision-based speedometer regulates human walking

Shinya Takamuku[1,2,*] and Hiroaki Gomi[1]

SUMMARY

Can we recover self-motion from vision? This basic issue remains unsolved since, while the human visual system is known to estimate the *direction* of self-motion from optic flow, it remains unclear whether it also estimates the *speed*. Importantly, the latter requires disentangling self-motion speed and depths of objects in the scene as retinal velocity depends on both. Here we show that our automatic regulator of walking speed based on vision, which estimates and maintains the speed to its preferred range by adjusting stride length, is robust to changes in the depths. The robustness was not explained by temporal-frequency-based speed coding previously suggested to underlie depth-invariant object-motion perception. Meanwhile, it broke down, not only when the interocular distance was virtually manipulated but also when monocular depth cues were deceptive. These observations suggest that our visuomotor system embeds a speedometer that calculates self-motion speed from vision by integrating monocular/binocular depth and motion cues.

INTRODUCTION

How does our brain visually estimate the motion of our body relative to the environment? The question relates to fundamental issues such as how we build our internal map and how we navigate within it. Computer vision studies (Fraundorfer and Scaramuzza, 2012; Scaramuzza and Fraundorfer, 2011) have made significant theoretical advances on how self-motion can be estimated. Our understanding on how the brain solves the problem has also progressed in recent decades.

Translational motion of our body relative to the environment causes radial optic flow on the retina. The *direction* of self-motion coincides with the focus of expansion which is independent of the 3D structure of the scene. Accordingly, when gaze direction is known and the rotational component of the flow can be canceled out, based for instance on extra-retinal signals, the direction can be specified from the local orientation pattern of the flow on the retina (Warren, 2004). In agreement with this theory, it has been shown that we can accurately perceive our heading direction under various conditions (Crowell et al., 1998; Royden et al., 1992; Warren, 1976; Warren and Hannon, 1988). The dorsal medial superior temporal area (MSTd) in the primate brain was identified to be selective to the local orientation pattern of the flow (Duffy and Wurtz, 1995; Tanaka et al., 1989) and was causally linked to heading-direction judgments based on microstimulation (Britten and van Wezel, 1998). In humans, attending to heading direction increases the activity in the inferior satellite of MT/V5+ (Peuskens et al., 2001), the homolog of MSTd.

The *speed* of translational self-motion, on the other hand, cannot be specified purely from the optic flow pattern as retinal velocity depends both on self-motion speed and depths of objects in the scene (Frenz et al., 2003; e.g., doubling self-motion speed and halving the depths have identical effects on retinal velocity). This raises a question of whether our visual system can recover the self-motion speed independently of the depths. Although several studies have examined speed tuning in self-motion-coding areas of the primate visual system such as MSTd (Duffy and Wurtz, 1997; Tanaka and Saito, 1989) and 7a (Avila et al., 2018; Phinney and Siegel, 2000), these only referred to correlation with retinal-motion velocity (i.e., angular velocity). No studies to our knowledge have revealed a *depth-invariant* coding of self-motion speed based on vision.

In this regard, studies on vection, the illusory sensation of self-motion caused by optic flow, have suggested mechanisms that could contribute to achieving the invariance. Wist et al. (1975) reported that the perceived

[1]NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3-1 Morinosato-Wakamiya, Atsugishi 243-0198, Kanagawa, Japan

[2]Lead contact

*Correspondence: shinya.takamuku.ka@hco.ntt.co.jp

https://doi.org/10.1016/j.isci.2021.103390

velocity of rotary self-motion inside a rotating cylinder increases with the perceived distance of the surround. Similarly, visual motion in the far background is found to dominate vection than motions nearby (Ito and Shibata, 2005; Ohmi et al., 1987; Ohmi and Howard, 1988). These findings suggest that sensitivity to retinal motion is scaled based on the perceived distance of the observed points (Wist et al., 1975). Such a strategy can contribute to the speed estimation as the angular velocity of a fixed point relative to the heading direction will have a one-to-one correspondence with self-motion speed once it is scaled with the distance (see the geometry of optic flow in STAR Methods). Other studies (Palmisano, 1996, 2002; Seya and Shinoda, 2018) have shown that vection is stronger with binocular vision. This suggests that stereo cues on depth and/or motion-in-depth could contribute to self-motion estimation.

It is, however, still unclear whether the visual system actually recovers the translational speed. Perceived velocity of self-motion is often influenced by eye height (Larish and Flach, 1990; Flach et al., 2004; Rudin-Brown, 2006). Similarly, road narrowing and widening are confused with acceleration and deceleration respectively (Festl et al., 2012) even under binocular vision (Ott et al., 2016). Frenz et al. (2003) examined human judgments on the distances traveled by a monocular camera based on its images. They found that the judgments were robust to changes in view height if the changes were explicitly shown. However, in this case, changes in camera speed biased the judgments on traveled distance. Similarly, Redlick et al. (2001) reported a dependency of travel-distance judgment on the acceleration profile of the moving camera. These findings may reflect erroneous attribution of changes in optic-flow speed to changes in depth. Posture response to motion-in-depth optic flow was initially reported to be distance-independent (Dijkstra et al., 1992), but several follow-up studies suggested otherwise (Dijkstra et al., 1994; Moraes et al., 2009).

Why did these studies fail to find the invariance? One possibility is that the provided visual cues were insufficient to fully compensate for the variation in depth (Distler et al., 2000; Tozawa, 2008). Yet another possibility is that the human visual system essentially lacks the mechanism to fully recover the scale. Interestingly, while visual inference of travel distance is explained (Lakshminarasimhan et al., 2018) as Bayesian integration of a visual estimate of self-motion velocity and a slow-speed prior, its accuracy, depending on the reliability of the visual estimate, did not improve by providing additional depth cues or with stereo vision (Frenz et al., 2007). Similarly, while speed sensitivity to stereo-motion is reported (Harris and Watamaniuk, 1995; Wardle and Alais, 2013), no study examined its distance invariance, and contributions of stereo cues to magnitude judgments on motion-in-depth speed are found to be limited (Palmisano et al., 2019; Seya and Shinoda, 2018). Finally, an earlier study that examined depth-invariance of posture responses to optic flow failed to find a significant contribution of stereo vision (Moraes et al., 2009).

These findings recall an earlier discussion on whether and how our brain estimates object-motion velocity independently of viewing distance; the issue of speed constancy (McKee and Smallman, 1998; Wallach, 1939). Many studies, with some exceptions (Distler et al., 2000; Tozawa, 2008), favored the view that the perception is based on temporal-frequency-based speed coding (Mather et al., 2017; McKee and Smallman, 1998); a useful but suboptimal strategy. Considering also the ubiquitous reliance on retinal velocity rather than the translational velocity in other species (Altshuler and Srinivasan, 2018), it was speculated (McKee and Smallman, 1998) that the human visual system may not have a genuine representation of translational velocity.

Here, in search of the invariant coding, we examined a highly optimized motor behavior; our locomotion. Although an earlier study (Warren et al., 2001) revealed that optic flow is used to estimate and adjust our heading direction, other studies (Pailhous et al., 1990; Konczak, 1994; Prokop et al., 1997) have suggested that it may also be used to control our walking speed. These studies showed that participants unintentionally increase their walking speed when optic-flow speed is decreased and vice versa. The change in walking speed was primarily owing to changes in stride length. Importantly, there was a strong linear relationship between optic flow speed and walking speed (Prokop et al., 1997) which suggested that some visual representation of speed regulated the locomotion based on feedback control. Although some studies speculated that this could contribute to minimizing energy consumption by maintaining optimal walking speed (Mohler et al., 2007; O'Connor and Donelan, 2012), it remained unclear whether the response was based on a genuine representation of the translational self-motion speed or some low-level representation of speed such as retinal-motion velocity or temporal frequency.

In this study, we show that the implicit adjustment of walking speed, driven by optic flow under a constraint on stride frequency, is invariant to changes in the depths of objects in the scene and that the invariance is

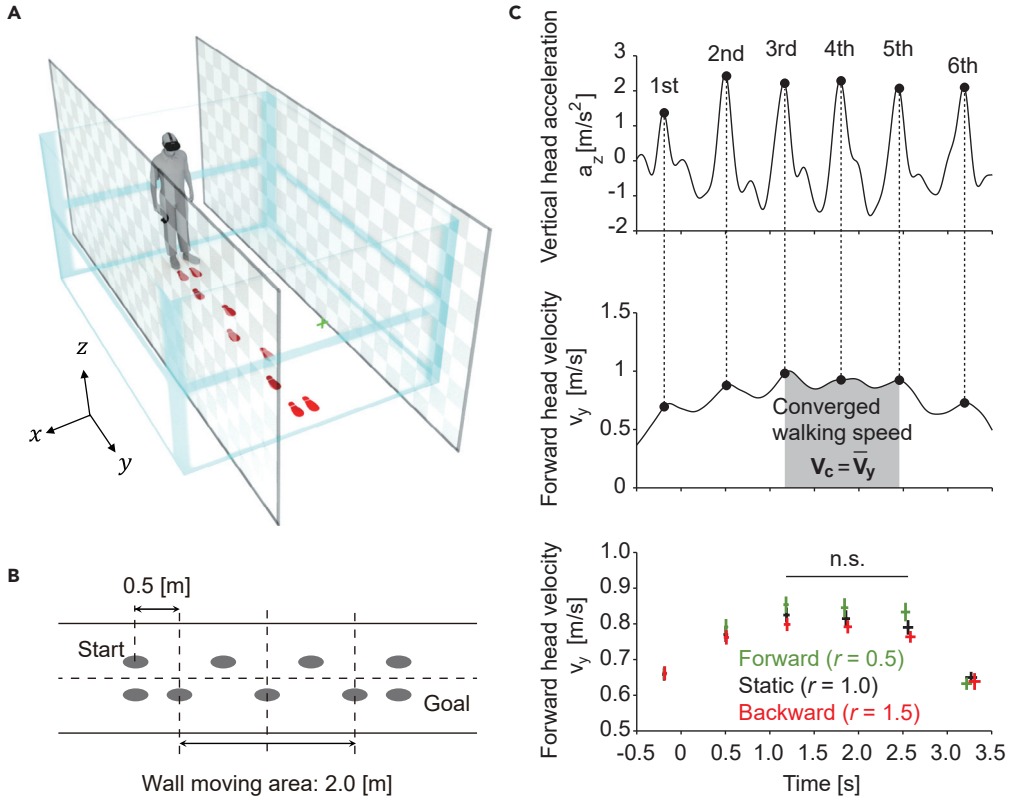


Figure 1. General design of experiments

(A) Setup of the experiment. Participants wore a binocular HMD and walked in a virtual corridor with checkerboard patterns attached to the sidewalls while viewing the fixation cross (shown in green).

(B) Participants walked seven steps starting from the right step. The wall moved when the head was within the indicated wall-moving area.

(C) Peak head velocity for each step was defined by referring to the vertical head acceleration (top panel). We took the average head velocity between the third to fifth steps as the approximation of the converged walking speed in each trial (middle panel). Bottom panal indicate averages and standard errors of peak velocity (vertical bar) and peak time (horizontal bar) across participants.

not explained by the temporal-frequency-based speed coding (Experiments 1 and 2). Then, we show that both binocular (Experiment 3) and monocular (Experiment 4) depth cues contribute to the invariant estimation. Our findings suggest that the implicit visuomotor response involves a visual speedometer that estimates translational self-motion speed from vision on the fly based on binocular/monocular depth and motion cues. Namely, the human visuomotor system solves the geometric problem of estimating self-motion speed from vision.

RESULTS

Based on the earlier studies (Konczak, 1994; Mohler et al., 2007; O'Connor and Donelan, 2012; Pailhous et al., 1990; Prokop et al., 1997), we assumed that changes in optic flow velocity would trigger adjustments of walking speed that correlate with some visual representation of speed. Our interest was on (Q1) which representation of speed underlies our adjustments of walking speed (translational self-motion velocity, retinal motion velocity or temporal frequency) and (Q2) what visual cues contribute to the neural computation of self-motion speed. We addressed these questions by (1) triggering changes in walking speed by moving the walls on the side and (2) examining how the depth and depth-related visual cues influence the changes in the walking speed.

Participants wore a binocular head-mounted display (HMD) and walked inside a virtual corridor with black and gray checkerboard patterns attached to the sidewalls (Figure 1A). They walked seven steps starting from the right foot (Figure 1B) while fixating at a target placed 3 m ahead along the corridor. The rhythm

of the steps was specified by periodic beeps with an interval of 700 ms. The constraint was introduced to ease the comparison of walking speed across participants and conditions with minimum effect on the visuomotor process. Note that implicit adjustment of walking speed is mostly explained by changes in stride length (Prokop et al., 1997). When the walker's head was within the wall-moving area, an area 0.5 m from the starting position and 2.0 m in length, the walls either moved forward, moved backwards, or stayed static, in order to change the optic flow velocity. The walls moved with half the velocity of the participant's head (i.e., closed-loop display) so that the velocity of the head relative to the sidewalls (wall-relative head velocity) would be 0.5 and 1.5 times the usual relative velocity in the forward and backward conditions, respectively. We refer to the wall-motion conditions using these ratios (relative velocity ratio r). r=0.5, 1.0 and 1.5 indicate the sidewalls moving forward, staying static, and moving backwards, respectively. In order to avoid any effects of explicit cognitive strategies, participants were instructed to disregard vision in controlling their walking speed and this was confirmed in a post-study questionnaire.

Figure 1C shows the procedure for the gait analysis. First, we detected the time point when the vertical acceleration of the head peaked for each step. Forward head velocity tended to peak around these time points. Analysis of the head velocities at these time points revealed that the walking speed changed by the wall motions as expected. As there were no significant differences in the velocity among the third, fourth, and fifth steps ($p > .05$), we took the head velocity, averaged across the time window between the time point of the third step and the time point of the fifth step, as an approximation of the converged walking speed for each trial (hereafter referred to as walking speed).

Experiment 1: Visuomotor control of walking speed is depth-invariant

First, in order to address (Q1), we examined whether the adjustment of walking speed depends on the distance of the sidewalls. We put the same walls, fully covered with a checkerboard pattern, either at a distance of 1.0 m (F condition) or 0.5 m (N condition) from the center of the corridor (red and green walls in Figure 2C). Panels surrounded by green and red rectangles in Figures 2A and 2B show the corresponding setups and images sent to the right display, respectively.

Upper graphs of Figure 2D show the translational and angular velocities of the sidewalls relative to the walker for each condition. Red and green points show the values for the far and near walls, respectively. The values are normalized so that the baseline condition (static far wall; F, r=1.0) would have a value of one. Lower graphs show the differences in walking speed relative to the baseline condition predicted for each hypothesis (see model section for details). As we can only predict the relative scale of these differences, the predicted differences in walking speed also have normalized values; the condition with far wall moving forward (F, r=0.5) having a value of 1 (increase in walking speed) and the condition with far wall moving backwards (F, r=1.5) having a value of −1 (decrease in walking speed). Further information on the normalizations are summarized in Method for normalization. If the walking speed is adjusted based on *translational* self-motion velocity (translational-velocity hypothesis), walking speed would depend on the head-relative translational velocity of the sidewalls but not on wall distance as shown in the bottom left graph of Figure 2D (see model subsection in STAR Methods). On the other hand, if the walking speed is adjusted based on the velocities of the motion signals on the retina (retinal-velocity hypothesis), walking speed would depend on the head-relative angular velocity of the walls (and thus on wall distance) as shown in the bottom right graph of Figure 2D as the retinal velocities correlate with the angular velocity.

As our model only predicts the normalized differences in walking speed, we test the hypotheses based on the magnitude relationship of the walking speeds and the relative scale of the wall-motion effects. Specifically, we test the following two important dissociations between the predictions of the two hypotheses. First, the translational-velocity hypothesis predicts that the walking speed would be lower when the wall is far and moving backwards (F, r=1.5) relative to when the wall is near and static (N, r=1.0) as the speed would appear to be higher in the former condition (green arrow in the bottom left panel of Figure 2D). Meanwhile, the retinal velocity hypothesis predicts the opposite relationship (green arrow in the bottom right panel of Figure 2D) as the retinal velocities would be larger when the wall is near and static (N, r=1.0). Second, the former hypothesis predicts that sensitivity to the wall motion (i.e., magnitudes of changes in walking speed) would be similar between the two walls, whereas the latter hypothesis predicts that the sensitivity would be doubled when the wall is near as retinal velocities would be twice as large in this condition compared with the other (F) condition (bottom panels of Figure 2D).

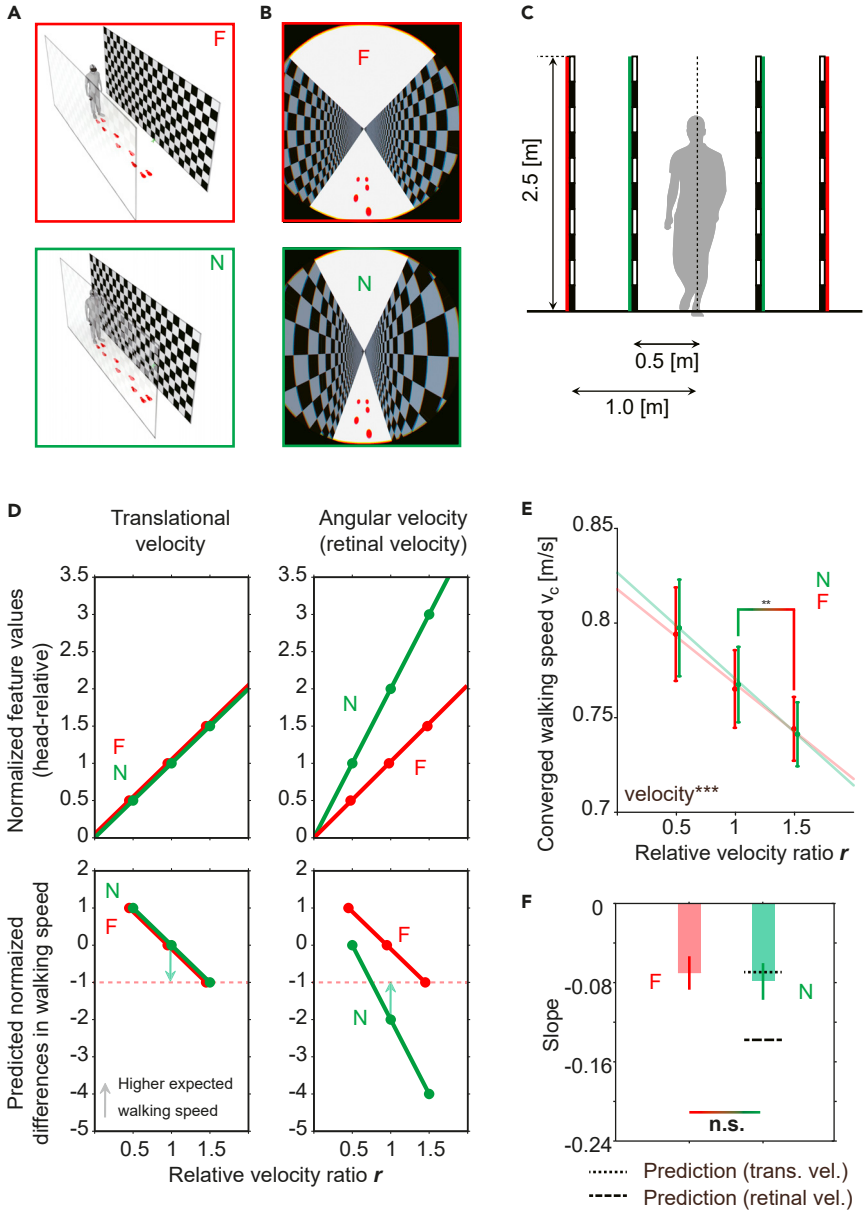


Figure 2. Visuomotor control of walking speed is depth-invariant (Exp. 1; N = 10)

(A) Wall conditions (F: Far, N: near).

(B) Images shown to the right eye when the head faced straight ahead at the center of the corridor (distortion of the image was corrected by the lens attached to the HMD).

(C) Height and location of the sidewalls and attached checkerboard patterns.

(D) Graphs in the upper row show normalized head-relative translational velocity of the sidewalls and head-relative angular velocity of the sidewalls approximated from the geometric relationship (note that the former velocity is defined in the reference frame of the laboratory whereas the later velocity is defined relative to the head). Owing to the normalization, only one value is shown for the angular velocity. Graphs in the lower row show the normalized differences in walking speed relative to the static far wall condition (F, r=1.0) predicted for each hypothesis. Green arrows indicate tested critical magnitude relationships.

(E) Walking speed plotted against the relative velocity ratio. Line with vague color indicates linear function fitted to plots for each sidewall. ** represents $p < .01$ (ANOVA/paired t test).

(F) Slope of linear fit between walking speed and the wall-relative head velocity (i.e., sensitivity to wall motion). Dotted and dashed lines indicate the slope of near-wall predicted from the translational velocity hypothesis and the retinal velocity hypothesis, respectively. In panels (E) and (F), plots and error bars indicate averages and standard errors across participants.

Figure 2E shows the walking speed, averaged across all participants (mean and standard error), for the conditions observed in Experiment 1. Two-factor repeated measures ANOVA revealed the main effect of wall velocity ($F_{2,18}$ = 16.0, p<.001, η_p^2 = 0.64), but the effects of wall distance ($F_{1,9}$ = 0.07, p = 0.80, η_p^2 = 0.01) and the interaction between wall distance and wall velocity ($F_{2,18}$ = 0.87, p = 0.44, η_p^2 = 0.09) were not significant. Importantly, comparison of the critical conditions revealed that walking speed was larger when the wall was near and static than when the wall was far and moving backwards (N, r=1.0 vs F, r=1.5; t_9=3.53, p=.006,d=1.12). This argued against the retinal-velocity hypothesis.

In order to specifically examine how the sensitivity to the wall motion differed between the two wall conditions, we fitted the walking speed with a linear function of the wall-relative head velocity (translational velocity of the head relative to the walls). Figure 2F shows the slopes of the fitted function. Dotted and dashed lines indicate the slope for the near-wall predicted from the slope of the far wall based on the translational velocity hypothesis and the retinal velocity hypothesis, respectively. Consistent with the former hypothesis, the difference in wall distance did not have a statistically significant effect on the slopes (N vs F: t_9=1.25, p=0.24,d=0.40). Furthermore, when we doubled the slopes for the far wall, these were significantly smaller (larger magnitude) than the slopes for the near wall (N vs doubled-F: t_9=3.25, p=0.01,d=1.03).

Although these results favored the translational-velocity hypothesis, Bayesian analysis (Dienes, 2014), applied to the slopes, did not allow us to determine whether the non-significant effect of wall distance represented depth-invariance of the visuomotor response or a lack of statistical power. When the H0 hypothesis assumed no difference between the wall conditions and H1 hypothesis assumed a uniform distribution between 0 and the average slope of the near wall, the Bayesian Factor BF_{01} was 1.9 (see STAR Methods for definition of Bayesian factor).

Experiment 2: Temporal-frequency-based speed coding does not explain the invariance

While our first experiment showed the depth-invariance of the response, it was not clear whether the results represented a genuine estimation of the translational velocity for two major reasons. First, as mentioned above, Bayesian analysis did not allow us to suggest the depth invariance of the response. The area, density, and location of the pattern (and consequently the optic flow) on the retina differed between the two wall conditions in Experiment 1, and the effects of these factors may have disturbed clear observation of the depth-invariance. Second, and more important, it was also possible that the walking speed was adjusted based on the temporal frequency of the visual motion stimuli, for example, the number of edges passing per unit time. Note that the frequency did not differ between the two wall conditions in Experiment 1 (see red and green plots in the top right panel of Figure 3E). Our second study addressed these concerns by running the same study with controlled stimuli.

Here, we used the same far walls (F condition shown in panel with red rectangle in Figure 3A), but used gray-colored walls with half-height and doubled-density checkerboard pattern for the near walls (N condition shown in panel with blue rectangle in Figure 3A; note that wall conditions are renamed for each experiment but colors are used consistently across experiments to specify the used walls). The height of the pattern on the near wall was adjusted for each participant so that the retinal positions of the near- and far-wall patterns would be mostly identical (blue wall in Figure 3C; see STAR Methods section for detail). As shown in panels surrounded by blue and red rectangles in Figure 3B, the density and the location of the checkerboard pattern on the retina were matched between the far and near walls.

The grey-colored part of the near wall provided a cue on the actual distance of the walls, but several monocular depth cues were lost. In order to examine whether the distances of the walls were still distinguishable, we asked the participants to judge the wall distance (far or near) at the beginning of each trial. Figure 3D shows the correct rate, averaged across all participants (mean and standard error), for each condition. Participants were highly accurate in identifying the distance (average correct rate above 98% in all conditions; see Table S1 for details).

Upper and lower graphs on the right side of Figure 3E show the normalized temporal frequency of the sidewalls (the number of edges passing the viewer per unit time) and the associated prediction of changes in walking speed, respectively. If the walking speed is adjusted based on the visual representation of translational self-motion velocity (translational-velocity hypothesis), the speed would depend on the relative translational velocity of the walls but not on the wall distance as shown in the lower left graph of Figure 3E.

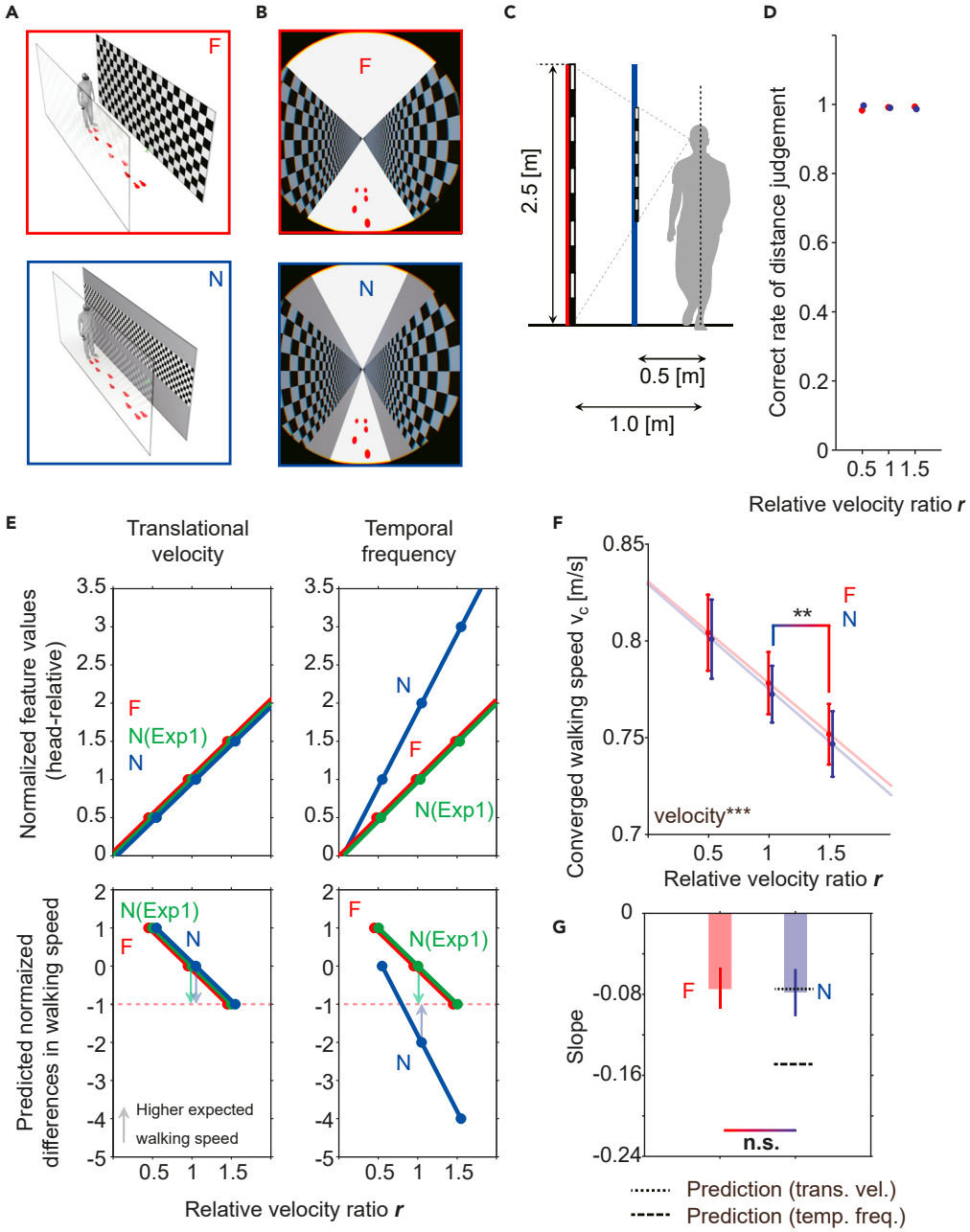


Figure 3. Temporal-frequency-based speed coding does not explain the invariance (Exp. 2; N = 10)

(A) Wall conditions (F: Far, N: near).

(B) Images shown to the right eye.

(C) Height and location of the sidewalls and attached checkerboard patterns.

(D) Correct rate of answering the wall distance.

(E) Graphs in the upper row show normalized head-relative translational velocity of the sidewalls and head-relative temporal frequency of the sidewalls approximated from the geometric relationship (note that the latter is defined relative to the head). Owing to the normalization, only one value is shown for the temporal frequency. Graphs in the lower row show the normalized differences in walking speed relative to the static far wall condition (F, r=1.0) predicted for each hypothesis. Green and blue arrows show tested critical magnitude relationships.

(F) Walking speed plotted against the relative velocity ratio. ** represents $p < .01$ (ANOVA/paired t test).

(G) Slope of linear fit between walking speed and the wall-relative head velocity (i.e., sensitivity to wall motion). Dotted and dashed lines indicate the slope of near-wall predicted from the translational velocity hypothesis and the temporal frequency hypothesis, respectively. In panels (D), (F), and (G), plots and error bars indicate averages and standard errors across participants.

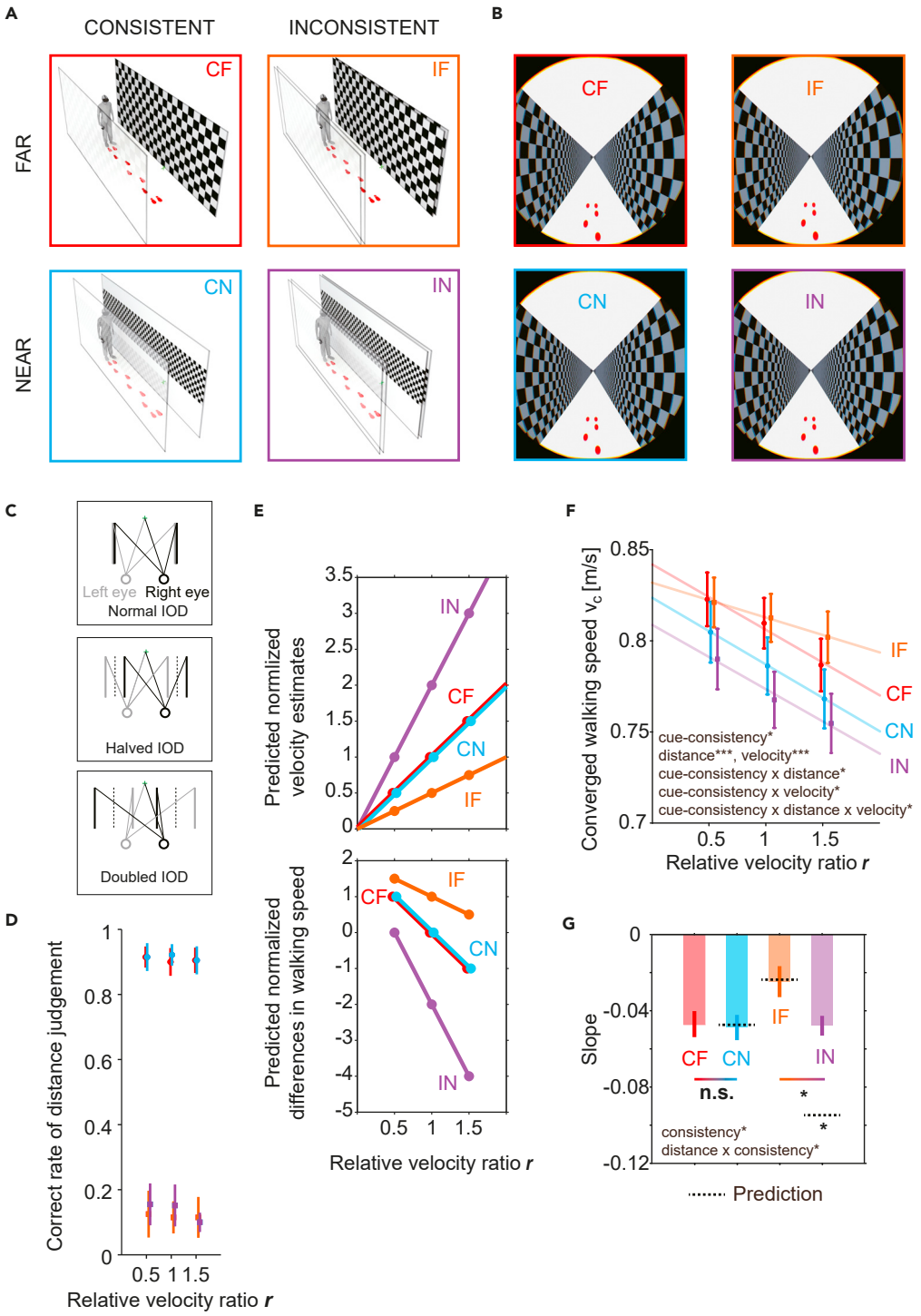


Figure 4. Binocular cues contribute to the invariance (Exp. 3; N = 10)

(A) Wall conditions (CF: Consistent Far, CN: Consistent Near, IF: Inconsistent Far, IN: Inconsistent Near).

(B) Images shown to the right eye.

(C) Method of manipulating IOD. Displayed positions of the sidewalls in VR space were slightly shifted from their original position (dotted line) either to the left or to the right depending on the viewing eye and the IOD condition (normal, halved, or doubled IOD). Light- and dark-colored lines indicate the positions of the sidewalls shown to the left and right eye, respectively. Green cross indicates fixation point.

(D) Correct rate of answering the wall distance.

Figure 4. *Continued*

(E) Upper graph shows the normalized self-motion velocity estimate predicted for each condition based on our model (Equation 14 in STAR Methods). Lower graph shows the normalized differences in walking speed relative to the static far wall condition (F, r=1.0) predicted from the model.

(F) Walking speed plotted against the relative velocity ratio.

(G) Slope of linear fit between walking speed and the wall-relative head velocity (i.e., sensitivity to wall motion). Dotted lines indicate model-predicted slopes. In panels (D), (F), and (G), plots and error bars indicate averages and standard errors across participants. *, **, and *** represent p < 0.5, p < .01, and p < .001, respectively (ANOVA/paired t test).

If the walking speed is adjusted based on the *temporal frequencies* of motion signals on the retina (temporal-frequency hypothesis), the pattern would depend on the temporal frequency (and thus on wall distance) as shown in the lower right graph of Figure 3E (see model in STAR Methods).

Here, the important differences in the predictions of the two hypotheses are as follows. First, the translational-velocity hypothesis predicts that the walking speed would be lower when the wall is far and moving backwards (F, r=1.5) relative to when the wall is near and static (N, r=1.0), whereas the temporal-frequency hypothesis predicts the opposite (blue arrows in the bottom panels of Figure 3E). Second, the former predicts similar sensitivity to the wall motion between the far and near walls (F and N), whereas the latter predicts that the sensitivity would be twice as large when the wall is near (N) as temporal frequency would be doubled in this condition (Figure 3E).

Figure 3F shows the walking speed, averaged across all participants (mean and standard error), for the conditions observed in Experiment 2. Two-factor repeated measures ANOVA revealed the main effect of wall velocity ($F_{2,18}$ = 12.3, p<.001, η_p^2 = 0.58), but there were again no statistically significant effects of wall distance ($F_{1,9}$ = 3.04, p = 0.12, η_p^2 = 0.25) or the interaction between wall distance and wall velocity ($F_{2,18}$ = 0.21, p = 0.81, η_p^2 = 0.02). Importantly, comparison of the critical conditions revealed that walking speed was larger when the wall is near and static compared with when the wall is far and moving backwards (N, r=1.0 vs F, r=1.5; t_9=3.31, p=.009,d=1.05). This denied the temporal-frequency hypothesis.

Figure 3G shows the slopes of the linear function fitted between the walking speed and the wall-relative head velocity. Dotted and dashed lines indicate the slope for the near-wall predicted from the slope of the far wall based on the translational velocity hypothesis and the temporal frequency hypothesis, respectively. Again, consistent with the former hypothesis, the slopes did not differ by wall distance (N vs F: t_9=0.79, p=0.45,d=0.25). Meanwhile, when we doubled the slopes for the far wall, this was significantly smaller (larger magnitude) than the slopes for the near wall (N vs doubled-F: t_9=3.87, p=.004,d=1.23).

Finally, Bayesian Factor BF_{01} was 5.3. As the factor was larger than 3, we considered this as evidence for the H0 hypothesis (i.e., no difference in sensitivity to wall motion by wall distance).

Experiment 3: Binocular cues contribute to the invariance

The results of Experiments 1 and 2 suggested that genuine representation of translational self-motion speed underlies the visuomotor adjustments of our walking speed. This shifted our attention to (Q2); what visual cues contribute to the computation of self-motion speed? In Experiment 3, we examined whether stereo cues contribute to the estimation by virtually varying the interocular distance (IOD) (Palmisano et al., 2019).

Here, we used the same far walls, but used white-colored walls with half-height and doubled-density checkerboard patterns for the near walls. Note that monocular depth cues are removed to maximize the effect of stereo cues. In one case, these walls were shown with normal IOD (CF condition and CN condition shown in panels with red and cyan rectangles in Figures 4A and 4B). In another case, these walls were shown with doubled- and halved IODs, respectively (IF condition and IN condition shown in panels with orange and purple rectangles in Figures 4A and 4B). We refer to the former two conditions as consistent conditions and the latter two conditions as inconsistent conditions as the stereo depth cues would be consistent and inconsistent with the actual wall distance in these conditions, respectively. The IOD was manipulated by slightly shifting the positions of the walls shown to each eye in the VR space (i.e., see Figure 4C). The manipulation made the far wall appear as if it is at a near distance and vice versa (see Figure 4D; see Method details for a geometry-based explanation). Three-factor repeated measures ANOVA, applied to the correct rate for answering the wall distance, revealed a significant effect of only the cue-consistency

(see Table S1 for details). Correct rate, averaged across the consistent conditions, was 91%. The rate for the inconsistent conditions was 12.7%.

If the estimation of the translational velocity involves calibrating the sensitivity to the retinal motion based on the estimated wall distances (geometry for optic flow in STAR Methods), and stereo cues contribute to the distance estimation, this predicts that estimated self-motion velocity (and therefore the walking speed) would be similar between the two consistent conditions (CF and CN), but would differ between the two inconsistent conditions (IF and IN) as the compensation would not work in the latter case. Upper and lower graphs in Figure 4E show the velocity estimates and differences in walking speed predicted from the model which assumes the decisive role of the manipulated cue(s) in estimating the wall distance.

Figure 4F shows the walking speed, averaged across all participants (mean and standard error), for the conditions observed in Experiment 3. Three-factor repeated measures ANOVA revealed main effects of cue-consistency ($F_{1,9}$ = 6.78, p = .029, η_p^2 = 0.43), wall distance ($F_{1,9}$ = 81.8, p<.001, η_p^2 = 0.90), wall velocity ($F_{2,18}$ = 40.1, p<.001, η_p^2 = 0.82), interaction between cue-consistency and wall distance ($F_{1,9}$ = 9.89, p = .012, η_p^2 = 0.52), interaction between cue-consistency and wall velocity ($F_{2,18}$ = 3.77, p = .043, η_p^2 = 0.30) and most importantly, the interaction among the three factors ($F_{2,18}$ = 4.67, p = .023, η_p^2 = 0.34).

Figure 4G shows the slopes. Two-factor repeated measures ANOVA revealed a significant effect of IOD ($F_{1,9}$ = 9.05, p = .015, η_p^2 = 0.50) and the interaction between cue-consistency and wall distance ($F_{1,9}$ = 7.67, p = .022, η_p^2 = 0.46). Importantly, while the slopes did not differ between the two consistent conditions (CF vs CN: t_9=0.28, p=0.78,d=0.09), they differed between the two inconsistent conditions (IN vs IF: t_9=2.60, p=.029,d=0.82), consistently with our prediction. The slope for the near wall was significantly larger than the double of the slope for the far wall in the consistent condition (CN vs doubled-CF: t_9=3.96, p=.003,d=1.25), but not in the inconsistent condition (IN vs doubled-IF: t_9=0.09, p=0.93,d=0.03). This suggested incomplete compensation for the difference in retinal velocity with inconsistent stereo cues. Our model predicts that the slopes of the CN, IF, and IN conditions would be 1.0, 0.5, and 2.0 times the slope of the CF condition if the manipulated cue(s) play a decisive role in estimating the wall distance (dotted line; see Equation 14 in STAR Methods). Although the slopes for the CN and IF conditions were similar to the prediction (two-sided t test between scaled CF slope and slopes of CN and IF conditions; p > .05), the slope for the IN condition was significantly larger (smaller magnitude) than the prediction (two-sided t test between scaled CF slope and IN slope; t_9=3.2, p=.01,d=1.01). This may reflect the contributions of other depth cues such as the accommodation cue.

The Bayesian factor BF_{01} was 4.6 for the consistent conditions, and 0.04 for the inconsistent conditions. This suggested that the sensitivities to the wall motion were identical between the consistent conditions, but differed between the inconsistent conditions. The findings of Experiment 3 suggested that stereo cues contribute to speed estimation.

Experiment 4: Monocular depth cues contribute to the invariance

Our final experiment also addressed (Q2). Here, we tested the contributions of monocular depth cues by manipulating the structure of the scene so that the far walls appear as if it is near and vice versa under monocular vision. In order to maximize the effect, vision was limited to the right eye.

The walls used in this experiment are shown in Figure 5A. For near walls with consistent and inconsistent monocular cues, we used the near walls used in Experiment 2 (CN condition shown in the panel with blue rectangle) and the near walls used in Experiment 3 (IN condition shown in panel with cyan rectangle), respectively. For far walls with consistent monocular cues, we used the far walls used in all the previous experiments (CF condition indicated in panel with red rectangle). Finally, for far walls with inconsistent monocular cues, we used far walls with horizontal gray panels attached to their top and bottom to partly cover the floor and the ceiling (IF condition indicated in panel with magenta rectangle). As shown in Figure 5B, the wall distance appeared erroneously in the conditions with inconsistent monocular cues. Participants consistently misjudged the distance of the walls in these conditions (Figure 5C). Again, three-factor repeated measures ANOVA, applied to the correct rate for answering the wall distance, revealed a significant effect of only the monocular-cue-consistency (see Table S1 for details). Correct rate, averaged across the consistent conditions, was 88.7%. The rate for the inconsistent conditions was 14.2%. This confirmed that the monocular depth cues were successfully manipulated.

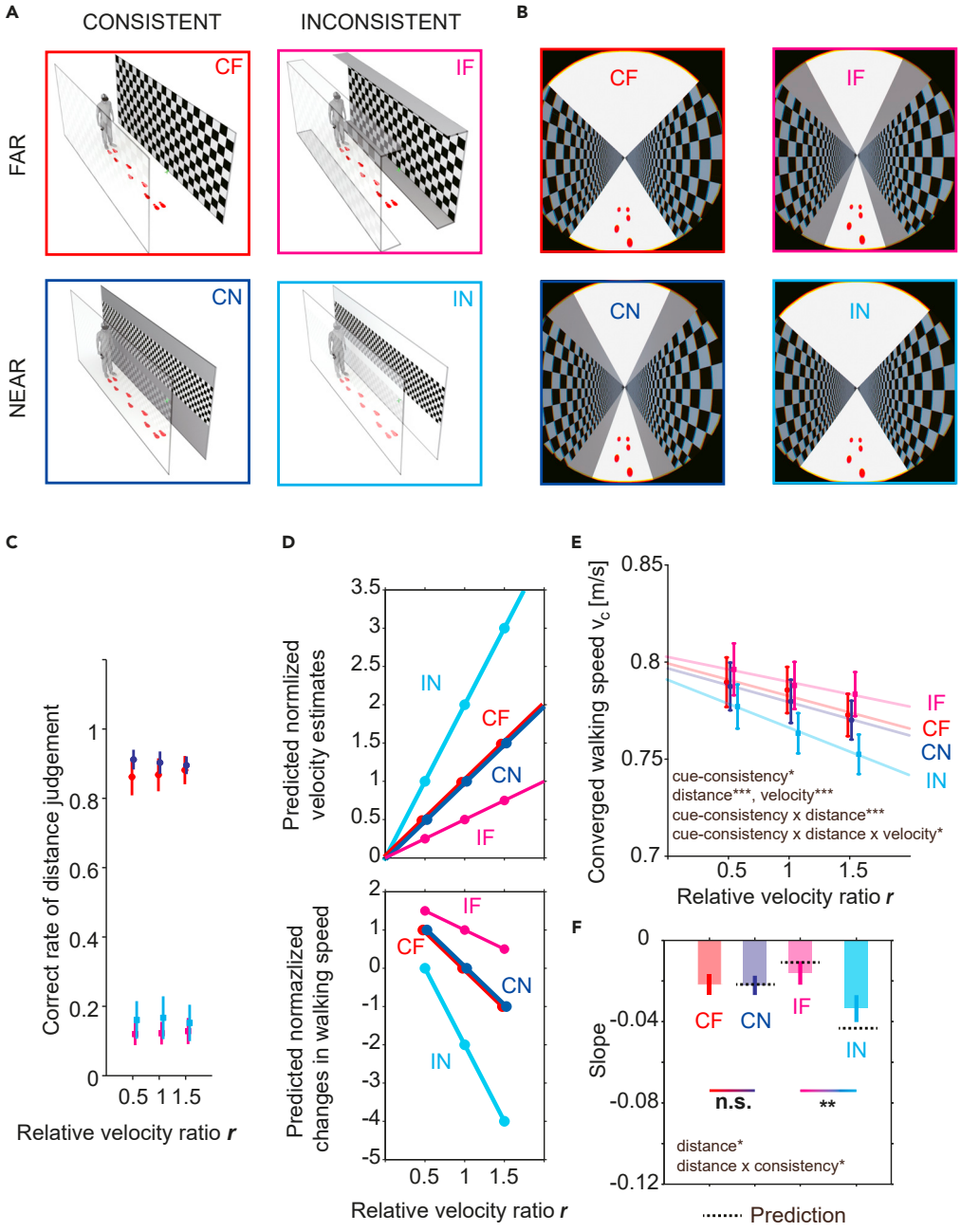


Figure 5. Monocular depth cues contribute to the invariance (Exp. 4; N = 20)

(A) Wall conditions (CF: Consistent Far, CN: Consistent Near, IF: Inconsistent Far, IN: Inconsistent Near).

(B) Images shown to the right eye.

(C) Correct rate of answering the wall distance.

(D) Upper graph shows the normalized self-motion velocity estimate predicted for each condition based on our model (Equation 14 in STAR Methods) Lower graph shows the normalized differences in walking speed relative to the static far wall condition (F, r=1.0) predicted from the model.

(E) Walking speed plotted against the relative velocity ratio.

(F) Slope of linear fit between walking speed and the wall-relative head velocity (i.e., sensitivity to wall motion). Dotted lines indicate model-predicted slopes. In panels (C), (E), and (F), plots and error bars indicate averages and standard errors across participants. *, **, and *** represent $p < 0.5$, $p < .01$, and $p < .001$, respectively (ANOVA/paired t test).

Predictions were similar to those made in Experiment 3. If the estimation of translational velocity involves calibrating the sensitivity to retinal motion based on the estimates of wall distances, and monocular cues contribute to the distance estimation, the estimated velocity (and therefore the walking speed) would be similar between the two consistent conditions, but would differ between the two inconsistent conditions as the compensation would not work in the latter case. Upper and lower graphs in Figure 5D show the velocity estimates and differences in walking speed predicted from the model which assumes the decisive role of the manipulated cue(s) in estimating the wall distance.

Figure 5E shows the walking speed. Three-factor repeated measures ANOVA revealed the main effects of monocular-cue-consistency ($F_{1,19}$ = 6.44, p = .02, η_p^2 = 0.25), wall distance ($F_{1,18}$ = 24.8, p <.001, η_p^2 = 0.56), wall velocity ($F_{2,38}$ = 42.1, p <.001, η_p^2 = 0.68), and the interaction among the three factors ($F_{2,38}$ = 3.58, p = .038, η_p^2 = 0.16).

Figure 5F shows the slopes. Two-factor repeated measures ANOVA revealed a significant effect of wall distance ($F_{1,19}$ =5.56, p =.029, η_p^2 =0.23) and the interaction between monocular-cue-consistency and wall distance ($F_{1,19}$ = 6.57, p = .019, η_p^2 = 0.26). Importantly, while the slopes did not differ between the two conditions with consistent monocular cues (HN-GP vs FF: t_{19}=0.11, p=0.91,d=0.03), they differed between the two conditions with inconsistent monocular cues (HN-WP vs FF-GP: t_{19}=2.98, p=.008,d=0.67), in agreement with our prediction. The slope for the near wall was significantly larger than the double of the slope for the far wall in the consistent condition (CN vs doubled-CF: t_{19}=2.67, p=.015,d=0.60), but not in the inconsistent condition (IN vs doubled-IF: t_{19}=0.15, p=0.87,d=0.03). Our model predicted that the slopes of the CN, IF, and IN conditions would be 1.0, 0.5, and 2.0 times the slope of the CF condition if the manipulated cue(s) play a decisive role in estimating the wall distance (dotted line; see Equation 14 in STAR Methods). Neither of these predictions was significantly different from the data (two-sided t test between scaled CF slope and CN, IF, IN slopes; p > .05).

Bayesian factor BF_{01} was 3.3 for the consistent-cue conditions, and 0.05 for the inconsistent-cue conditions. This suggested that the sensitivities to the wall motion were identical between the consistent conditions, but differed between the two inconsistent conditions. The findings of Experiment 4 suggested the contribution of monocular depth cues.

DISCUSSION

Our study repeatedly showed that visuomotor adjustments of walking speed, driven by optic flow under a constraint on stride frequency, are invariant to changes in depths of objects in the scene (Experiments 1, 2, and consistent conditions of Experiments 3 and 4). Although depth-invariance of postural responses to visual motion can be explained by the minimization of retinal slips (Dijkstra et al., 1992), the observed invariance is clearly inconsistent with feedback control of retinal velocity. The invariance was not explained by temporal-frequency coding (Experiment 2) previously suggested underlying speed constancy (McKee and Smallman, 1998). Meanwhile, it broke down when monocular or binocular depth cues were manipulated (Experiments 3 and 4). Our study thus suggests that the visuomotor system embeds a speedometer that estimates self-motion speed from vision by integrating depth and motion cues.

Our results are inconsistent with several variants of walking-speed adjustments based on retinal velocity or temporal frequency. For instance, the representation of the walking speed could be a linear summation of the translational and retinal velocities. However, when we regressed the walking speeds with both the relative translational velocity and the relative angular velocity of the walls for the experiments in which we tested the invariance (note that this correlates with temporal frequency in Experiment 2), only the weight of the former factor was significant (See Figure S1). One may also speculate that the brain predicts the retinal velocity (or temporal frequency) for the preferred walking speed, and take the prediction error in those dimensions to adjust the walking speed. In fact, such predictive coding has been suggested to account for optic-flow-speed perception (Durgin et al., 2005). Such feedback control, however, is inconsistent with our observation that sensitivity to wall motion did not depend on the wall distance.

Instead of the easy solutions, our results suggested that the visuomotor system solves the geometric problem of estimating self-motion speed from vision by integrating depth and motion cues. The fact that the depth-invariance persisted when monocular and binocular cues were removed (Experiments 3 and 4) suggests major contributions of both cues. This was confirmed by showing that the manipulation of each cue

breaks the invariance. Importantly, earlier evidence on the predominance of peripheral visual area (>30 deg) in controlling posture (Brandt et al., 1973; Lestienne et al., 1977) and locomotion (Stoffregen et al., 1987) does not rule out stereo contributions to the speed estimation as a large portion of peripheral vision (<120 deg) is available binocularly, interocular velocity difference can be calculated across widely separated regions (Greer et al., 2016), and depth and motion-in-depth estimations are based on the integration of binocular and monocular cues (Landy et al., 1995; Regan and Beverley, 1979). In estimating the depth, monocular cues may be beneficial in estimating the depth structure of the whole scene.

Regarding the whole control framework, all our results are consistent with a model which assumes that (H1) self-motion speed is estimated from vision by integrating monocular/binocular depth and motion cues, (H2) estimates of different sensory modalities are integrated based on their reliabilities (optimal integration theory (Campos et al., 2012; Ernst and Banks, 2002)), and (H3) walking speed is adjusted so that the multimodal estimate would match the preferred speed (Figure 6A; see modelsubsection in STAR Methods). Optimal integration theory (H2) predicts that the visual contribution would decrease when its reliability deteriorates. It is worth noting, in this regard, that the estimated sensitivities to the wall motions were relatively smaller in experiments that included inconsistent conditions (Experiments 1 and 2 vs consistent conditions in Experiments 3 and 4), and were even smaller with monocular vision (consistent conditions of Experiment 4). Note that the magnitudes of the slopes would correlate with the ratio between the sensory weightings (Equation 14 in Model subsection of STAR Methods).

Our finding contrasts with the earlier observation of distance-*dependent* self-motion estimations in humans. Several factors could explain the contrast. First, many earlier studies provided visual stimuli in 2D. These stimuli may not have provided enough cues to fully recover the self-motion speed. Second, the fact that participants actually moved may have contributed to the invariant estimation as consistency among multisensory signals could prevent erroneous interpretation of the visual input. Earlier studies showed that locomotion actually impacts optic flow processing (Durgin et al., 2005; Yabe and Taga, 2008). Finally, depth-invariant speed estimation may be a unique feature of the implicit visuomotor system. In our study, post-study questionnaires confirmed that the adjustments were implicit. Many studies have suggested that such implicit visuomotor responses may involve visual-motion processing pathways that differ from those for perception (Bridgemen et al., 1981; Gomi et al., 2006; Goodale et al., 1986). Interestingly, while an earlier study (Palmisano et al., 2019) suggested that virtually doubling the IOD during viewing optic flow causes the perceived motion-in-depth speed to increase, we found that the same manipulation decreases the sensitivity to the wall motion in adjusting the walking speed (comparison of slope between CF and IF: t_9=4.25, p=0.002,d=1.35). Although several factors (e.g., stimuli, fixation) differ between the two studies, the contrast may represent the dissociation between perceptual and motor processes.

To conclude, our study revealed the ability of our visuomotor system to automatically estimate self-motion speed from vision on the fly by integrating depth and motion cues. A similar ability has been tested in various species (Altshuler and Srinivasan, 2018; Creamer et al., 2018; David, 1982; Saleem et al., 2013; Scholtyssek et al., 2014), starting with the seminal work on honeybees (see Srinivasan, 2011 for review). Although some insects showed the ability to estimate speed or traveled distanced from vision independently of contrast and spatial frequency of the optic flow pattern (Fry et al., 2009; Si et al., 2003), robustness to changes in depth has rarely been observed and demonstrates the marvelous efficiency of the human visual system. Self-motion speed estimation from monocular vision has only recently been explored in the field of computer vision (Li et al., 2018; Wang et al., 2017; Zhan et al., 2018) and further study on the primate visual odometry system may contribute to developing advanced algorithms for speed monitoring in artificial systems.

Limitations of the study

While our study suggests the existence of a speedometer in the human visuomotor system, several questions remain to be addressed. First, our observation of depth-invariant adjustment of walking speed based on vision is limited to a situation in which the stride frequency is specified. Although the finding by Prokop et al. (1997) suggests that involuntary adjustments of walking speed based on optic flow are predominantly owing to changes in stride length even without the specification of stride frequency, depth-invariance of the adjustment in unconstrained locomotion remains to be confirmed.

Second, while our setup controlled the use of several cues (HMD: accommodation, fixation: vergence), the contribution of specific binocular and monocule cues remains to be clarified. For example, the advantage

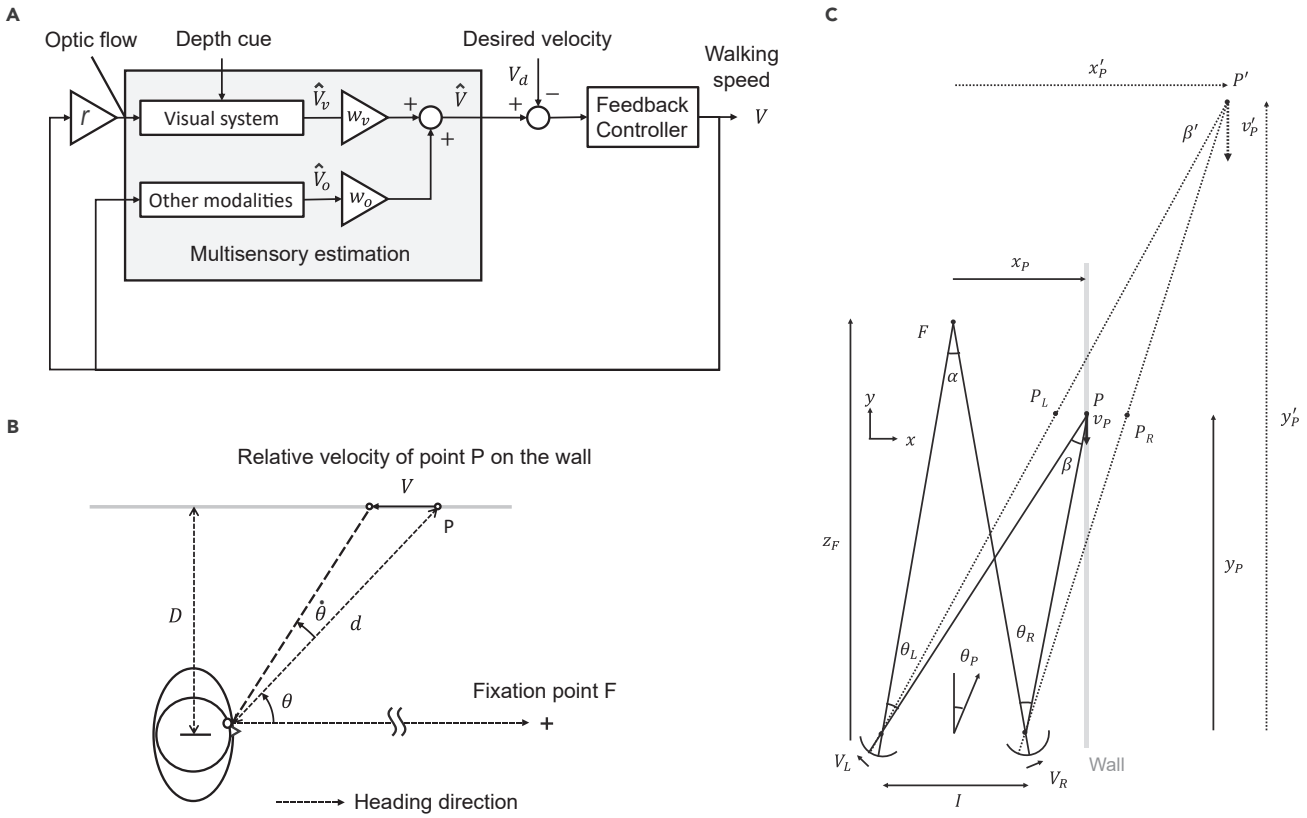


Figure 6. Computational model for the walking speed adjustment

(A) shows the block diagram of our model. The visual system estimates the walking speed based on optic flow and depth cues. Other sensory modalities also provide an estimate of the speed. These estimates are integrated based their reliabilities (most-likelihood estimation). The resulting multimodal estimate is then compared with the desired speed to adjust the walking speed (feedback control).

(B) Shows the geometric relationship between the translational velocity, depth, and optic flow speed.

(C) Impact of IOD manipulation on the geometric relationship between the observer, fixation, and point on the wall. Observer fixates on point *F* and point *P* on the wall moves relative to the observer at velocity $\boldsymbol{v}_P$. When point *P* is fixed to the environment, the relative velocity corresponds to the self-motion velocity ($-\boldsymbol{v}_P$). Dotted lines indicate a situation when IOD is virtually halved ($\boldsymbol{k = 1/2}$) for the wall (P_R and P_L indicate the position of point *P* shown to the right and left eye, respectively). The manipulation doubles the estimated distance of the side walls and the estimated velocity.

of stereo vision could be explained based on depth cue (binocular disparity) or the combination of depth and motion-in-depth cues (Harris and Watamaniuk, 1995) (changing disparity and/or interocular velocity difference) (see method details). Similarly, visual manipulation in Experiment 4 changed several monocular depth cues (Brenner and Smeets, 2018) together; that is, size-cue (retinal height of the sidewalls), perspective cue (angle of the sidewalls), and height-in-the-field cue (declination angle of the bottom of the side walls from the horizon; see Ooi et al., 2001). Only the last cue specifies the absolute depth. In our experiment, the fixation point was always at eye level. This may have facilitated the use of the height-in-the-field cue.

Finally, it remains unclear how the self-motion velocity is computed in the brain. Optic-flow-velocity coding, invariant to variations in spatiotemporal frequencies, is found in the dorsal stream of the primate brain, starting from the complex cells in V1 (Priebe et al., 2006) and continuing to areas MT (Perrone and Thiele, 2001; Priebe et al., 2003) and MSTd (Kawano et al., 1994). Although cells in V1 are mostly limited to coding motion energy, pattern-direction sensitivity initiates in MT (Movshon et al., 1985; Pack and Born, 2001) and increases in MST (Khawaja et al., 2013). During pursuit eye movements, responses of MST neurons are more correlated with the motion of image on screen whereas those of MT neurons are more correlated with retinal motion (Chukoskie and Movshon, 2009; Inaba et al., 2007). Interestingly, MT and MST neurons also encode depth in terms of binocular disparity or surface layout (Uka and DeAngelis, 2006) and area MT (Bradley et al., 1998; Czuba et al., 2014; Sanada and DeAngelis,

2014), as well as human MT+ (Rokers et al., 2009), are also reported to be sensitive to stereo cues on motion-in-depth. Subpopulations of neurons in areas MSTd (Duffy and Wurtz, 1997; Tanaka and Saito, 1989), VIP (Bremmer et al., 2002; Schaafsma and Duysens, 1996), and 7a (Avila et al., 2018; Phinney and Siegel, 2000) of the primate brain are found to be selective to optic flow patterns indicative of 3D self-motion. Among these, activities in MSTd (Duffy and Wurtz, 1997; Orban et al., 1995) and 7a (Avila et al., 2018; Phinney and Siegel, 2000) are found to depend on flow speed. However, the invariance of these speed coding to variation in-depth as well as the neural processes which enable the invariant estimation remains to be clarified. Our method of examining the speed estimation process by measuring the automatic response in adjusting walking speed may play a pivotal role in addressing these neuroscience questions.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Experimental apparatus
  - General task
  - Experiment 1
  - Experiment 2
  - Experiment 3
  - Experiment 4
  - Geometry of optic flow
  - Model
  - Method for normalization
  - Possible contributions of stereo cues
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Trajectory analysis
- STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103390.

ACKNOWLEDGMENTS

This study was supported by KAKEN JP16H06566. We thank Tomoyuki Nagasawa and Koya Yoshino for their help in collecting the data. We also thank Shin'ya Nishida and Patrick Haggard for their helpful comments on the study.

AUTHOR CONTRIBUTIONS

S.T. and H.G. designed the experiment and wrote the paper. S.T. collected and analyzed the data.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 24, 2021
Revised: August 23, 2021
Accepted: October 28, 2021
Published: December 17, 2021

REFERENCES

Altshuler, D.L., and Srinivasan, M.V. (2018). Comparison of visually guided flight in insects and birds. Front. Neurosci. *12*, 157. https://doi.org/10.3389/fnins.2018.00157.

Avila, E., Lakshminarasimhan, K.J., DeAngelis, G.C., and Angelaki, D.E. (2018). Visual and vestibular selectivity for self-motion in macaque posterior parietal area 7a. Cereb. Cortex N. Y. N. *1991*. https://doi.org/10.1093/cercor/bhy272.

Bradley, D.C., Chang, G.C., and Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. Nature *392*, 714–717. https://doi.org/10.1038/33688.

Brandt, T., Dichgans, J., and Koenig, E. (1973). Differential effects of central verses peripheral vision on egocentric and exocentric motion perception. Exp. Brain Res. *16*, 476–491. https://doi.org/10.1007/BF00234474.

Bremmer, F., Duhamel, J.-R., Ben Hamed, S., and Graf, W. (2002). Heading encoding in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. *16*, 1554–1568. https://doi.org/10.1046/j.1460-9568.2002.02207.x.

Brenner, E., and Smeets, J.B.J. (2018). Depth perception. In Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Sensation, Perception, and Attention, J.T. Wixted and J. Serences, eds. (John Wiley & Sons), pp. 385–414.

Bridgemen, B., Kirch, M., and Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. Percept. Psychophys. *29*, 336–342.

Britten, K.H., and van Wezel, R.J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. Nat. Neurosci. *1*, 59–63. https://doi.org/10.1038/259.

Campos, J.L., Butler, J.S., and Bülthoff, H.H. (2012). Multisensory integration in the estimation of walked distances. Exp. Brain Res. *218*, 551–565. https://doi.org/10.1007/s00221-012-3048-1.

Chukoskie, L., and Movshon, J.A. (2009). Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. J. Neurophysiol. *102*, 3225–3233. https://doi.org/10.1152/jn.90692.2008.

Creamer, M.S., Mano, O., and Clark, D.A. (2018). Visual control of walking speed in Drosophila. Neuron. https://doi.org/10.1016/j.neuron.2018.10.028.

Crowell, J.A., Banks, M.S., Shenoy, K.V., and Andersen, R.A. (1998). Visual self-motion perception during head turns. Nat. Neurosci. *1*, 732–737. https://doi.org/10.1038/3732.

Czuba, T.B., Huk, A.C., Cormack, L.K., and Kohn, A. (2014). Area MT encodes three-dimensional motion. J. Neurosci. *34*, 15522–15533. https://doi.org/10.1523/JNEUROSCI.1081-14.2014.

David, C.T. (1982). Compensation for height in the control of groundspeed by Drosophila in a new, 'barber's pole' wind tunnel. J. Comp. Physiol. *147*, 485–493. https://doi.org/10.1007/BF00612014.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. Front. Psychol. *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781.

Dijkstra, T.M., Gielen, C.C.A.M., and Melis, B.J.M. (1992). Postural responses to stationary and moving scenes as a function of distance to the scene. Hum. Mov. Sci. *11*, 195–203. https://doi.org/10.1016/0167-9457(92)90060-O.

Dijkstra, T.M., Schöner, G., and Gielen, C.C. (1994). Temporal stability of the action-perception cycle for postural control in a moving visual environment. Exp. Brain Res. *97*, 477–486.

Distler, H.K., Gegenfurtner, K.R., van Veen, H.A., and Hawken, M.J. (2000). Velocity constancy in a virtual reality environment. Perception *29*, 1423–1435. https://doi.org/10.1068/p3115.

Duffy, C.J., and Wurtz, R.H. (1997). Medial superior temporal area neurons respond to speed patterns in optic flow. J. Neurosci. *17*, 2839–2851.

Duffy, C.J., and Wurtz, R.H. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. J. Neurosci. *15*, 5192–5208.

Durgin, F.H., Gigone, K., and Scott, R. (2005). Perception of visual speed while moving. J. Exp. Psychol. Hum. Percept. Perform. *31*, 339–353. https://doi.org/10.1037/0096-1523.31.2.339.

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

Festl, F., Recktenwald, F., Yuan, C., and Mallot, H.A. (2012). Detection of linear ego-acceleration from optic flow. J. Vis. *12*. https://doi.org/10.1167/12.7.10.

Flach, J.M., Junaid, A.A., and Warren, R. (2004). Judgments of speed of self-motion: modeling the relative effects of speed and altitude change. Proc. Hum. Factors Ergon. Soc. Annu. Meet. *48*, 1923–1927. https://doi.org/10.1177/154193120404801628.

Fraundorfer, F., and Scaramuzza, D. (2012). Visual odometry : Part II: matching, robustness, optimization, and applications. IEEE Robot. Autom. Mag. *19*, 78–90. https://doi.org/10.1109/MRA.2012.2182810.

Frenz, H., Bremmer, F., and Lappe, M. (2003). Discrimination of travel distances from "situated" optic flow. Vis. Res *43*, 2173–2183.

Frenz, H., Lappe, M., Kolesnik, M., and Bührmann, T. (2007). Estimation of travel distance from visual motion in virtual environments. ACM Trans. Appl. Percept. *4*. https://doi.org/10.1145/1227134.1227137.

Fry, S.N., Rohrseitz, N., Straw, A.D., and Dickinson, M.H. (2009). Visual control of flight speed in Drosophila melanogaster. J. Exp. Biol. *212*, 1120–1130. https://doi.org/10.1242/jeb.020768.

Gomi, H., Abekawa, N., and Nishida, S. (2006). Spatiotemporal tuning of rapid interactions between visual-motion analysis and reaching movement. J. Neurosci. *26*, 5301–5308. https://doi.org/10.1523/JNEUROSCI.0340-06.2006.

Goodale, M.A., Pelisson, D., and Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. Nature *320*, 748–750. https://doi.org/10.1038/320748a0.

Greer, D.A., Bonnen, K., Huk, A.C., and Cormack, L.K. (2016). Speed discrimination in the far monocular periphery: a relative advantage for interocular comparisons consistent with self-motion. J. Vis. *16*, 7. https://doi.org/10.1167/16.10.7.

Harris, J.M., and Watamaniuk, S.N. (1995). Speed discrimination of motion-in-depth using binocular cues. Vis. Res *35*, 885–896.

Inaba, N., Shinomoto, S., Yamane, S., Takemura, A., and Kawano, K. (2007). MST neurons code for visual motion in space independent of pursuit eye movements. J. Neurophysiol. *97*, 3473–3483. https://doi.org/10.1152/jn.01054.2006.

Ito, H., and Shibata, I. (2005). Self-motion perception from expanding and contracting optical flows overlapped with binocular disparity. Vis. Res *45*, 397–402. https://doi.org/10.1016/j.visres.2004.11.009.

Kawano, K., Shidara, M., Watanabe, Y., and Yamane, S. (1994). Neural activity in cortical area MST of alert monkey during ocular following responses. J. Neurophysiol. *71*, 2305–2324.

Khawaja, F.A., Liu, L.D., and Pack, C.C. (2013). Responses of MST neurons to plaid stimuli. J. Neurophysiol. *110*, 63–74. https://doi.org/10.1152/jn.00338.2012.

Konczak, J. (1994). Effects of optic flow on the kinematics of human gait: a comparison of young and older adults. J. Mot. Behav. *26*, 225–236. https://doi.org/10.1080/00222895.1994.9941678.

Lakshminarasimhan, K.J., Petsalis, M., Park, H., DeAngelis, G.C., Pitkow, X., and Angelaki, D.E. (2018). A dynamic bayesian observer model reveals origins of bias in visual path integration. Neuron. https://doi.org/10.1016/j.neuron.2018.05.040.

Landy, M.S., Maloney, L.T., Johnston, E.B., and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. Vis. Res *35*, 389–412.

Larish, J.F., and Flach, J.M. (1990). Sources of optical information useful for perception of speed of rectilinear self-motion. J. Exp. Psychol. Hum. Percept. Perform. *16*, 295–302.

Lestienne, F., Soechting, J., and Berthoz, A. (1977). Postural readjustments induced by linear motion of visual scenes. Exp. Brain Res. *28*, 363–384. https://doi.org/10.1007/BF00235717.

Li, R., Wang, S., Long, Z., and Gu, D. (2018). UnDeepVO: monocular visual odometry through unsupervised deep learning. In 2018 IEEE International Conference on Robotics and Automation (ICRA). (IEEE), pp. 7286–7291.

Mather, G., Sharman, R.J., and Parsons, T. (2017). Visual adaptation alters the apparent speed of

real-world actions. Sci. Rep. *7*, 6738. https://doi.org/10.1038/s41598-017-06841-5.

McKee, S.P., and Smallman, H.S. (1998). Size and speed constancy. In Perceptual Constancy: Why Things Look as They Do, V. Walsh and J. Kulikowski, eds. (Cambridge University Press), pp. 373–408.

Mohler, B.J., Thompson, W.B., Creem-Regehr, S.H., Pick, H.L., and Warren, W.H. (2007). Visual flow influences gait transition speed and preferred walking speed. Exp. Brain Res. *181*, 221–228. https://doi.org/10.1007/s00221-007-0917-0.

Moraes, R., Lopes, A.G., and Barela, J.A. (2009). Monocular vision and increased distance reducing the effects of visual manipulation on body sway. Neurosci. Lett. *460*, 209–213. https://doi.org/10.1016/j.neulet.2009.05.078.

Movshon, J., Adelson, E.H., Gizzi, M.S., and Newsome, W.T. (1985). The analysis of moving visual patterns. Pattern Recognit. Mech. 117–151.

Niehorster, D.C., Li, L., and Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. Iperception *8*. 2041669517708205. https://doi.org/10.1177/2041669517708205.

O'Connor, S.M., and Donelan, J.M. (2012). Fast visual prediction and slow optimization of preferred walking speed. J. Neurophysiol. *107*, 2549–2559. https://doi.org/10.1152/jn.00866.2011.

Ohmi, M., and Howard, I.P. (1988). Effect of stationary objects on illusory forward self-motion induced by a looming display. Perception *17*, 5–11. https://doi.org/10.1068/p170005.

Ohmi, M., Howard, I.P., and Landolt, J.P. (1987). Circular vection as a function of foreground-background relationships. Perception *16*, 17–22. https://doi.org/10.1068/p160017.

Ooi, T.L., Wu, B., and He, Z.J. (2001). Distance determined by the angular declination below the horizon. Nature *414*, 197–200. https://doi.org/10.1038/35102562.

Orban, G.A., Lagae, L., Raiguel, S., Xiao, D., and Maes, H. (1995). The speed tuning of medial superior temporal (MST) cell responses to optic-flow components. Perception *24*, 269–285. https://doi.org/10.1068/p240269.

Ott, F., Pohl, L., Halfmann, M., Hardiess, G., and Mallot, H.A. (2016). The perception of ego-motion change in environments with varying depth: interaction of stereo and optic flow. J. Vis. *16*, 4. https://doi.org/10.1167/16.9.4.

Pack, C.C., and Born, R.T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature *409*, 1040–1042. https://doi.org/10.1038/35059085.

Pailhous, J., Ferrandez, A.M., Flückiger, M., and Baumberger, B. (1990). Unintentional modulations of human gait by optical flow. Behav. Brain Res. *38*, 275–281. https://doi.org/10.1016/0166-4328(90)90181-d.

Palmisano, S. (2002). Consistent stereoscopic information increases the perceived speed of vection in depth. Perception *31*, 463–480. https://doi.org/10.1068/p3321.

Palmisano, S. (1996). Perceiving self-motion in depth: the role of stereoscopic motion and changing-size cues. Percept. Psychophys. *58*, 1168–1176.

Palmisano, S., Davies, R.G., and Brooks, K.R. (2019). Vection strength increases with simulated eye-separation. Atten. Percept. Psychophys. *81*, 281–295. https://doi.org/10.3758/s13414-018-1609-5.

Perrone, J.A., and Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat. Neurosci. *4*, 526–532. https://doi.org/10.1038/87480.

Peuskens, H., Sunaert, S., Dupont, P., Van Hecke, P., and Orban, G.A. (2001). Human brain regions involved in heading estimation. J. Neurosci. *21*, 2451–2461.

Phinney, R.E., and Siegel, R.M. (2000). Speed selectivity for optic flow in area 7a of the behaving macaque. Cereb. Cortex N. Y. N. *10*, 413–421.

Priebe, N.J., Cassanello, C.R., and Lisberger, S.G. (2003). The neural representation of speed in macaque area MT/V5. J. Neurosci. *23*, 5650–5661.

Priebe, N.J., Lisberger, S.G., and Movshon, J.A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. J. Neurosci. *26*, 2941–2950. https://doi.org/10.1523/JNEUROSCI.3936-05.2006.

Prokop, T., Schubert, M., and Berger, W. (1997). Visual influence on human locomotion. Modulation to changes in optic flow. Exp. Brain Res. *114*, 63–70.

Redlick, F.P., Jenkin, M., and Harris, L.R. (2001). Humans can use optic flow to estimate distance of travel. Vis. Res. *41*, 213–219.

Regan, D., and Beverley, K.I. (1979). Binocular and monocular stimuli for motion in depth: changing-disparity and changing-size feed the same motion-in-depth stage. Vis. Res *19*, 1331–1342. https://doi.org/10.1016/0042-6989(79)90205-0.

Rokers, B., Cormack, L.K., and Huk, A.C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. Nat. Neurosci. *12*, 1050–1055. https://doi.org/10.1038/nn.2343.

Royden, C.S., Banks, M.S., and Crowell, J.A. (1992). The perception of heading during eye movements. Nature *360*, 583–585. https://doi.org/10.1038/360583a0.

Rudin-Brown, C.M. (2006). The effect of driver eye height on speed choice, lane-keeping, and car-following behavior: results of two driving simulator studies. Traffic Inj. Prev. *7*, 365–372. https://doi.org/10.1080/15389580600851927.

Saleem, A.B., Ayaz, A., Jeffery, K.J., Harris, K.D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. Nat. Neurosci. *16*, 1864–1869. https://doi.org/10.1038/nn.3567.

Sanada, T.M., and DeAngelis, G.C. (2014). Neural representation of motion-in-depth in area MT. J. Neurosci. *34*, 15508–15521. https://doi.org/10.1523/JNEUROSCI.1072-14.2014.

Scaramuzza, D., and Fraundorfer, F. (2011). Visual odometry: Part I - the first 30 Years and fundamentals. IEEE Robot. Autom. Mag. *18*, 80–92. https://doi.org/10.1109/MRA.2011.943233.

Schaafsma, S.J., and Duysens, J. (1996). Neurons in the ventral intraparietal area of awake macaque monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their responses to optic flow patterns. J. Neurophysiol. *76*, 4056–4068. https://doi.org/10.1152/jn.1996.76.6.4056.

Scholtyssek, C., Dacke, M., Kröger, R., and Baird, E. (2014). Control of self-motion in dynamic fluids: fish do it differently from bees. Biol. Lett. *10*, 20140279. https://doi.org/10.1098/rsbl.2014.0279.

Seya, Y., and Shinoda, H. (2018). Relationship between vection and motion perception in depth. Atten. Percept. Psychophys. *80*, 2008–2021. https://doi.org/10.3758/s13414-018-1567-y.

Si, A., Srinivasan, M.V., and Zhang, S. (2003). Honeybee navigation: properties of the visually driven "odometer. J. Exp. Biol. *206*, 1265–1273. https://doi.org/10.1242/jeb.00236.

Srinivasan, M.V. (2011). Honeybees as a model for the study of visually guided flight, navigation, and biologically inspired robotics. Physiol. Rev. *91*, 413–460. https://doi.org/10.1152/physrev.00005.2010.

Stoffregen, T.A., Schmuckler, M.A., and Gibson, E.J. (1987). Use of central and peripheral optical flow in stance and locomotion in young walkers. Perception *16*, 113–119. https://doi.org/10.1068/p160113.

Tanaka, K., Fukada, Y., and Saito, H.A. (1989). Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. J. Neurophysiol. *62*, 642–656. https://doi.org/10.1152/jn.1989.62.3.642.

Tanaka, K., and Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. J. Neurophysiol. *62*, 626–641. https://doi.org/10.1152/jn.1989.62.3.626.

Tozawa, J. (2008). Speed constancy and the perception of distance. Perception *37*, 3–21. https://doi.org/10.1068/p5585.

Uka, T., and DeAngelis, G.C. (2006). Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. J. Neurosci. *26*, 6791–6802. https://doi.org/10.1523/JNEUROSCI.5435-05.2006.

Wallach, H. (1939). On constancy of visual speed. Psychol. Rev. *46*, 541–552.

Wang, S., Clark, R., Wen, H., and Trigoni, N. (2017). DeepVO: towards end-to-end visual odometry with deep recurrent convolutional neural networks. In 2017 IEEE International Conference on Robotics and Automation (ICRA). (IEEE), pp. 2043–2050.

Wardle, S.G., and Alais, D. (2013). Evidence for speed sensitivity to motion in depth from binocular cues. J. Vis. *13*. https://doi.org/10.1167/13.1.17.

Warren, R. (1976). The perception of egomotion. J. Exp. Psychol. Hum. Percept. Perform. *2*, 448–456.

Warren, W. (2004). 84. Optic flow. In The Visual Neurosciences, L.M. Chalupa and J.S. Werner, eds. (MIT Press), pp. 1247–1259.

Warren, W.H., and Hannon, D.J. (1988). Direction of self-motion is perceived from optical flow. Nature *336*, 162–163. https://doi.org/10.1038/336162a0.

Warren, W.H., Kay, B.A., Zosh, W.D., Duchon, A.P., and Sahuc, S. (2001). Optic flow is used to control human walking. Nat. Neurosci. *4*, 213–216. https://doi.org/10.1038/84054.

Wist, E.R., Diener, H.C., Dichgans, J., and Brandt, T. (1975). Perceived distance and the perceived speed of self-motion: linear vs. angular velocity? Percept. Psychophys *17*, 549–554. https://doi.org/10.3758/BF03203967.

Yabe, Y., and Taga, G. (2008). Treadmill locomotion captures visual perception of apparent motion. Exp. Brain Res. *191*, 487–494. https://doi.org/10.1007/s00221-008-1541-3.

Zhan, H., Garg, R., Weerasekera, C.S., Li, K., Agarwal, H., and Reid, I.M. (2018). Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. (IEEE), pp. 340–349.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MATLAB (R2007b)	MathWorks	RRID: SCR_001622
Unity (5.4.1.f1)	Unity Technologies	

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Shinya Takamuku (shinya.takamuku.ka@hco.ntt.co.jp)

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Experimental data is provided as supplemental information.
- For providing the stimuli, we used HTC Vive (Niehorster et al., 2017) and Unity5. For data analysis, we used MATLAB (version R2007b) and the Bayes factor calculator: https://medstats.github.io/bayesfactor.html
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A total of 50 neurologically intact participants (15 males and 35 females; 37.1 ± 8 y.o., mean ± s.d.) were recruited to participate in four experiments. Experimental procedures were approved by the NTT Communication Science Laboratory Research Ethics Committee and were conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants.

METHOD DETAILS

Experimental apparatus

HTC Vive (Niehorster et al., 2017) was used to provide the visual stimuli and record the head position (both at 90 Hz). Distance between the left and right displays was adjusted for each participant by measuring their IOD. The controller was used to answer the wall distance. The corridor was built using a VR software (Unity). The floor and the ceiling were white and had no textures. The sidewalls always had a height of 2.5 m, but had different panels, checkerboard patterns and disparities depending on the wall conditions used in each experiment.

General task

Participants walked seven steps starting from the right foot (Figure 1B) along the virtual corridor while viewing a fixation point always placed 3 m ahead. The rhythm of their locomotion was controlled by giving an instruction to step on the floor in sync with periodic beeps at interval of 700 ms. Their task was to keep a constant walking velocity by concentrating on their bodily sensations while disregarding vision. A questionnaire was given to the participants after completing the experiment, and all of them denied their use of vision on adjusting their walking speed. In each trial, participants answered the distance of the sidewalls before initiating their locomotion. There were no time limits for giving this answer, but eye and head motions were prohibited.

Experiment 1

Ten typical adults (2 males and 8 females; 40.0±7 y.o., mean±s.d.) participated in the experiment. Participants walked along a virtual corridor in which the sidewalls were placed either at a distance of 0.5 (N) or 1.0 (F) m from the center (green and red walls in Figure 2C). Checkerboard pattern with full-height (2.5 m) and a spatial frequency of 2.0 cycle/m was attached to both walls (green and red panels in Figure 2A). The two walls were easily distinguishable (Figure 2B). There were 6 conditions in total: two wall conditions combined with three wall-velocity conditions (relative velocity ratio: r=0.5, 1.0 and 1.5). Conditions with wall motion (r=0.5 and 1.5) and without wall motion (r=1.0) were repeated 20 and 120 times, respectively. The order of the conditions were pseudo-randomized across trials.

Experiment 2

Ten typical adults (3 males and 7 females; 38.1±10 y.o., mean±s.d.) participated in the experiment. The experiment was similar to Experiment 1, but the near wall had a checkerboard pattern with half-height (1.25 m), doubled (4.0 cycle/m) spatial frequency (N: blue panels in Figure 3A). The height of the pattern was adjusted for each participant so that patterns of the far and near walls would appear at the same retinal area (blue wall in Figure 3C). The panel of the wall had a color (gray) which differed from the color of the floor or the ceiling (white). This provided a monocular cue to distinguish the distances of the two walls (red and blue panels in Figure 3B). Again, there were 6 conditions in total: two wall conditions combined with three wall-velocity conditions. Conditions with wall motion and without wall motion were repeated 30 and 60 times, respectively. The order of the conditions were pseudo-randomized across trials.

Experiment 3

Ten adults (3 males and 7 females; 31.3±9 y.o., mean±s.d.) participated in the experiment. Here, interocular distance (IOD) was virtually manipulated by slightly shifting the position of the sidewalls shown on the left and right displays (Figure 4C). The far wall was identical to those used in Experiments 1 and 2, and the near wall was similar to that used in Experiment 2 except the fact that the panel was white. The walls were shown either with normal IOD (CF and CN walls shown in red and cyan panels of Figure 4A) or with manipulated IOD (IF and IN walls shown in orange and purple panels of Figure 4A). The walls were designed to limit the contribution of monocular depth cues (Figure 4B) and to maximize the contribution of the binocular cues. Far wall shown with doubled IOD (IF) was perceived to be near, and near wall shown with halved IOD (IN) was perceived to be far (Figure 4D). There were 12 conditions in total: two cue-consistency conditions, two wall conditions, and three wall-velocity conditions. Conditions with wall motion and without wall motion were repeated 20 and 60 times, respectively. The order of the conditions were pseudo-randomized across trials.

Experiment 4

Twenty typical adults (7 males and 13 females; 38.2±6 y.o., mean±s.d.) participated in the experiment. Number of participants was doubled based on finding of pilot study that sensitivity to wall motion are approximately halved with monocular vision. Here, in order to specifically examine the contribution of monocular depth cues, images were shown only to the right display. The walls used in this experiment are shown in Figure 5A. For near walls with consistent and inconsistent monocular cues, we used the near walls used in Experiment 2 (indicated in blue) and Experiment 3 (indicated in cyan), respectively. For far wall with consistent monocular cues, we used the far wall used in all the previous experiments (indicated in red). Finally, for far walls with inconsistent monocular cues, we used far walls with horizontal gray panels attached to its top and bottom to partly cover the floor and the ceiling (IF, magenta panel in Figure 5A). As shown in Figure 5B, the wall distance appeared erroneously in the conditions with inconsistent monocular cues (IN and IF). There were 12 conditions in total: two monocular-cue-consistency conditions, two wall conditions, and three wall-velocity conditions. Conditions with wall motion and without wall motion were repeated 20 and 60 times, respectively. The order of the conditions were pseudo-randomized across trials.

Geometry of optic flow

Figure 6B illustrates the geometric relationship between the translational self-motion, the depth of the scene, and the optic flow. When participant moves at speed V m/s while observing a fixed point P, the speed can be approximated from three attributes of P, its distance (d m), its angle relative to the heading

direction (θ rad), and its relative angular velocity ($\dot{\theta}$ rad/s; corresponding to the optic flow), based on the following relation.

$$V \cong (d \sin\theta)\dot{\theta} \quad \text{(Equation 1)}$$

This means that, when the rotational components of optic flow can be canceled out, self-motion speed can be estimated by scaling the retinal-motion speed at a certain location on the retina relative to the focus of expansion with the estimated distance. In our setup, the relationship can be further simplified using the distance of the walls (D m) as follows.

$$V \cong D\dot{\theta} \quad \text{(Equation 2)}$$

Model

Figure 6A illustrates the model used to design and analyze our experiments. The estimate of walking speed ($\hat{V}$ m/s) is obtained as a linear summation of estimates based on vision ($\hat{V}_v$ m/s) and other modalities ($\hat{V}_o$ m/s); i.e.,

$$\hat{V} = w_v \hat{V}_v + w_o \hat{V}_o \quad \text{(Equation 3)}$$

where w_v and w_o represent the weight given to each estimate. This is based on an earlier study (Campos et al., 2012) which suggested that such basic linear model based on most-likelihood estimation explains the multimodal estimation of walked distance. The walking speed is then adjusted so that the estimate equals the desired walking speed (V_d m/s).

$$\hat{V} = V_d \quad \text{(Equation 4)}$$

We can assume that the other modalities provide accurate estimate of the walking speed.

$$\hat{V}_o = V \quad \text{(Equation 5)}$$

Solving Equations (3), (4), and (5) gives the following equation which shows that the converged walking speed (V_c) would have a negative correlation with the visual estimate.

$$V_c = -\frac{w_v}{w_o}\hat{V}_v + \frac{V_d}{w_o} \quad \text{(Equation 6)}$$

If vision provides a genuine estimate of the translational velocity, the visual estimate would correspond to the wall-relative head velocity (V_r), the translational velocity of the walker's head relative to the sidewalls, i.e.,

$$\hat{V}_v = V_r = rV. \quad \text{(Equation 7)}$$

r represents the relative velocity ratio. Note that, since head-relative wall velocity is $-V_r$, top left panels in Figures 2D and 3E also indicate the wall-relative head velocity (i.e., the two velocities have the same normalized values). When we fit the converged walking speed with the wall-relative head velocity, the slope would not depend on the distance of the sidewalls (see bottom left panels of Figures 2D and 3E) and would represent the ratio of the integration weights as follows.

$$V_c = -\frac{w_v}{w_o}V_r + \frac{V_d}{w_o} \quad \text{(Equation 8)}$$

On the other hand, if the visual estimate of the walking speed is proportional to the velocities of the motion signals on the retina (retinal velocity hypothesis), it would be proportional to the angular velocity (V_a rad/s) of the sidewalls relative to the walker as follows (remember Equation 2).

$$\hat{V}_v \propto V_a = \frac{V_r}{D} = \frac{r}{D}V \quad \text{(Equation 9)}$$

When we fit the converged walking speed with the wall-relative head velocity, the slope would depend on the wall distance as described by the following equation (see bottom right panel of Figure 2D). K represents a constant.

$$V_c = -\frac{w_v}{w_o}KV_a + \frac{V_d}{w_o} = -\frac{w_v}{w_o}\frac{K}{D}V_r + \frac{V_d}{w_o} \quad \text{(Equation 10)}$$

If the visual estimate depends on the temporal frequencies of motion signals on the retina (temporal frequency hypothesis), it would be proportional to the temporal frequency of the sidewalls (F_t cycle/s) relative to the walker as follows:

$$\widehat{V}_v \propto F_t = \frac{F_s V_r}{D} = \frac{F_s r}{D} V \quad \text{(Equation 11)}$$

where F_s cycle/rad is the spatial frequency of the pattern on the retina. This predicts the relation described by the following equation. K' represents a constant.

$$V_c = -\frac{w_v}{w_o} K' F_t + \frac{V_d}{w_o} = -\frac{w_v}{w_o} \frac{K' F_s}{D} V_r + \frac{V_d}{w_o} \quad \text{(Equation 12)}$$

Since F_s is identical among the two walls in Experiment 2, this predicts the relation illustrated in the bottom right panel of Figure 3E.

If the translational self-motion velocity (walking speed) is estimated by scaling the retinal motion velocities with the estimated distance of the sidewalls ($\widehat{D}$ m) i.e., estimation based on Equations 1 or 2), the visual estimate would be as follows.

$$\widehat{V}_v = \widehat{D} V_a = \frac{\widehat{D}}{D} V_r \quad \text{(Equation 13)}$$

This predicts the following relation.

$$V_c = -\frac{w_v}{w_o} \widehat{V}_v + \frac{V_d}{w_o} = -\frac{w_v}{w_o} \frac{\widehat{D}}{D} V_r + \frac{V_d}{w_o} \quad \text{(Equation 14)}$$

The predictions for Experiments 3 and 4 are illustrated in the bottom panels of Figures 4E and 5D, respectively.

Earlier studies (O'Connor and Donelan, 2012; Prokop et al., 1997) have suggested that optimization based on direct sensors of metabolic cost, such as central and peripheral chemoreceptors and group IV muscle afferents, take place in adjusting walking speed over a longer duration. Our model does not cover this aspect since we only observe the initial phase of the walking-speed adjustment.

Method for normalization

Visual features indicated in the upper graphs of Figures 2D and 3E as well as the model-predicted velocity estimate indicated in the upper graph of Figures 4E and 5D are scaled so that the baseline condition would have a value of one. The baseline condition is the baseline wall condition with the walls static (r=1.0). The baseline wall condition is the F condition in Experiments 1 and 2 and the CF condition in Experiments 3 and 4. When U represents the absolute value, the normalized value u is represented as

$$u \approx \frac{U}{U_b} \quad \text{(Equation 15)}$$

where U_b represents the absolute value for the baseline condition.

The model-predicted differences in walking speed relative to the baseline condition indicated in the bottom graphs of Figures 2D, 3E, 4E, and 5D also have normalized values (note that our model only predicts the relative scales of the wall-motion effects). When V represents the walking speed, the difference of walking speed relative to the baseline condition is represented as

$$\Delta V = V - V_b \quad \text{(Equation 16)}$$

where V_b represents the walking speed in the baseline condition. When the difference in walking speed caused by the wall motions in the baseline wall condition is represented as ΔV_{bw} (note that our model predicts equal effect sizes for forward and backward motions), the normalized values that the model predicts is represented as follows.

$$\widehat{\Delta v} = \frac{\Delta V}{\Delta V_{bw}} \quad \text{(Equation 17)}$$

Possible contributions of stereo cues

In Experiment 3, we found that the IOD manipulation (1) caused the wall distances to be misjudged (near wall appear far and vice versa) and (2) broke the depth-invariance of the walking speed adjustment. Two cues could account for these effects.

First, the depth-invariant speed estimation could be realized by scaling the sensitivity to the optic flow based on the estimated depth. Therefore, stereo *depth* cues could contribute to the estimation.

Figure 6C illustrates the effects of the IOD manipulation. When the observer fixates on point F, the vergence angle of fixation F (α rad) can be approximated based on distance of point F (y_F m) and IOD (I m) as follows.

$$\alpha \approx \frac{I}{y_F} \quad \text{(Equation 18)}$$

Meanwhile, vergence angle of point P (β rad) can be approximated based on location of point P (x_P, y_P m) as follows.

$$\beta \approx \frac{I y_P}{x_P^2 + y_P^2} \quad \text{(Equation 19)}$$

Using the wall distance D (=x_P) m and the orientation of point P relative to the midpoint between the two eyes as θ_P rad, this can be represented as follows.

$$\beta \approx \left(\frac{1}{2} sin2\theta_P\right) \frac{I}{D} = A(\theta_P)\frac{I}{D} \quad \text{(Equation 20)}$$

Orientation θ_P rad can be approximated using the orientation of point P relative to the right (θ_R rad) and left (θ_L rad) eyes as follows.

$$\theta_P \approx (\theta_R + \theta_L)/2 \quad \text{(Equation 21)}$$

Binocular disparity of point P (δ rad), on the other hand, can be given as follows.

$$\delta = \theta_R - \theta_L = \alpha - \beta \quad \text{(Equation 22)}$$

Using these relations, wall distance D can be estimated based on fixation vergence α, orientation θ_P, and binocular disparity δ as follows.

$$\widehat{D} = \frac{A(\theta_P)\widehat{I}}{\alpha - \delta} \quad \text{(Equation 23)}$$

Then, self-motion speed can be estimated using the geometry of optic flow by scaling the optic flow speed with the estimated wall distance.

When the IOD for the side walls is increased with gain k,

$$I' = kI \quad \text{(Equation 24)}$$

vergence angle of point P is increased with gain k.

$$\beta' = \delta' - \alpha' = k(\delta - \alpha) = k\beta \quad \text{(Equation 25)}$$

Meanwhile, orientation θ_P is unchanged. Therefore, the estimated distance increases with gain $1/k$.

$$\widehat{D'} = \frac{1}{k}\widehat{D} \quad \text{(Equation 26)}$$

This predicts that the far wall (D = 1.0 m) with doubled IOD (FF-DS) would be estimated to be near ($\widehat{D}$ = 0.5 m) and near wall (D = 0.5 m) with halved IOD (HN-WP-HS) would be estimated to be far ($\widehat{D}$ = 1.0 m). This explains the impact of the IOD manipulation on the wall-distance judgements (Figure 4D). Furthermore, if the sensitivity to optic flow is scaled according to the estimated depth (see Equation 2), estimated self-motion speed would also increase with gain $1/k$. Therefore, this predicts that the walking speed would increase for the far wall with doubled IOD and decrease for the near wall with halved IOD. Namely, the presumed contribution of depth cues also explains the increased difference of sensitivity between the two conditions with abnormal IOD.

The effect of IOD manipulation on walking speed may also reflect the contribution of stereo *motion-in-depth* cues. When we assume that point P moves in the y axis with velocity v_P m/s and the fixation remains at the same distance, temporal derivative of the disparity can be given as follows.

$$\frac{\partial \delta}{\partial t} \approx -\frac{I\left(x_P^2 - y_P^2\right)}{\left(x_P^2 + y_P^2\right)^2} v_P \quad \text{(Equation 27)}$$

This can be represented using D and θ_P as follows.

$$\frac{\partial \delta}{\partial t} \approx sin^2\theta_P \cos(2\theta_P)\frac{I}{D^2}v_P = B(\theta_P)\frac{I}{D^2}v_P \quad \text{(Equation 28)}$$

$\frac{\partial \delta}{\partial t}$ rad/s is obtained as changing disparity (change in disparity over time) and inter-ocular velocity difference ($\frac{\partial \delta}{\partial t} = V_R - V_L$). When point P is fixed to the environment, self-motion speed (v_S) can be estimated from $\frac{\partial \delta}{\partial t}$, θ_P, and $\widehat{D}$ as follows.

$$\widehat{v}_S = -\widehat{v}_P = \frac{\partial \delta}{\partial t}\frac{\widehat{D}^2}{B(\theta_P)\widehat{I}} \quad \text{(Equation 29)}$$

When the IOD is increased with gain k, the temporal derivative of the disparity (and thus the two motion-in-depth cues) increases with gain k.

$$\left(\frac{\partial \delta}{\partial t}\right)' = B(\theta_P)\frac{(kI)}{D^2}v_P = k\frac{\partial \delta}{\partial t} \quad \text{(Equation 30)}$$

Using Equations (26), (29) and (30), the estimated self-motion speed is expected to increase with gain of $1/k$.

$$\widehat{v'}_S = -\left(k\frac{\partial \delta}{\partial t}\right)\frac{\left(D^2/k^2\right)}{B(\theta_P)\widehat{I}} = \frac{1}{k}\widehat{v}_S \quad \text{(Equation 31)}$$

Therefore, the contribution of motion-in-depth cues can also explain the changes in walking speed.

QUANTIFICATION AND STATISTICAL ANALYSIS

Trajectory analysis

Head position was recorded at 90 Hz and low-pass-filtered at 5 Hz using a fourth-order, no-lag, Butterworth filter. Since the participants were going back-and-forth along the corridor, we inverted the head velocity recorded in trials with even number. Then, the velocity profile was aligned based on the time when the participant's head entered the wall-moving area. Figure 1C shows the procedure for the gait analysis. We detected the time point when the vertical acceleration of the head peaked for each step. Forward head velocity tended to peak around these time points, and the use of vertical acceleration provided robust definition of the peak velocities for each step. Trials in which these peak velocities exceeded their 3 SD ranges were excluded from analysis. When we averaged the peak velocities across all conditions, there were no significant differences among the velocities of the third, fourth, and fifth steps ($p > 0.05$). We therefore calculated the head velocity, averaged across the time window between the time point of the third step and the time point of the fifth step, as the approximation of the converged walking speed for each trial.

STATISTICAL ANALYSIS

In analyzing the walking speed for each experiment, we first carried out two- (Experiments 1 and 2: wall distance and velocity) or three- (Experiments 3 and 4: cue-consistency, wall distance and velocity) factor repeated-measures ANOVA using R. In Experiments 1 and 2, we further compared walking speed between two critical conditions to specify the factor underlying the response (paired t test). To analyze how the sensitivity to wall motion differed among wall-types, we fitted a linear function between walking speed and the wall-relative head velocity for each participant. Then, we compared the slopes between the near and far walls or between the measured and predicted values (paired t test). The former difference was further analyzed by calculating a Bayesian Factor (Dienes, 2014)

$$BF_{01} = \frac{p(data|H0)}{p(data|H1)} \quad \text{(Equation 32)}$$

where p(data|H0) and p(data|H1) represents the likelihoods of H0 (no difference between the wall conditions) and H1 (sensitivity lower for far wall; assumed a uniform distribution between 0 and the average slope of the near condition) hypotheses, respectively. Factor above 3 supported the H0 hypothesis and factor below 1/3 supported the H1 hypothesis.

In analyzing the correct rate of judging the wall distance, measured in Experiments 2–4, we carried out two- (Experiment 2: wall distance and velocity) or three- (Experiments 3 and 4: cue-consistency, wall distance and velocity) factor repeated-measures ANOVA using R.

Further details on the datasets used for these analysis can be found in Data S1 and S2.