



Vection induced by low-level motion extracted from complex animation films

Wataru Suzuki^{1,2} · Takeharu Seno³ · Wakayo Yamashita⁴ · Noritaka Ichinohe¹ · Hiroshige Takeichi^{5,6} · Stephen Palmisano⁷

Received: 3 August 2019 / Accepted: 21 October 2019 / Published online: 11 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This study examined the contributions of low-, mid- and high-level visual motion information to vection. We compared the vection experiences induced by hand-drawn and computer-generated animation clips to those induced by versions of these movies that contained only their pure optic flow. While the original movies were found to induce longer and stronger vection experiences than the pure optic flow, vection onsets were not significantly altered by removing the mid- and high-level information. We conclude that low-level visual motion information appears to be important for vection induction, whereas mid- and higher-level display information appears to be important for sustaining and strengthening this vection after its initial induction.

Keywords Vection · Optic flow · Visual motion · Psychological Experiment

Introduction

Humans use multiple senses to perceive their self-motions through the world (including vision, the vestibular system of the inner ear, kinesthesia, somatosensation and even audition). The important role that vision plays in self-motion perception is clearly demonstrated by the occurrence of vection (i.e., compelling visual illusions of self-motion induced in physically stationary observers¹; e.g., Brandt, Dichgans and

Koenig 1973; Brandt, Wist and Dichgans 1971). This vection has traditionally been assumed to be the result of low-level visual motion processing. However, recent evidence shows that vection also depends on mid-level visual processing (such as the computations involved in perceiving surfaces and material properties—see Kim, Khuu and Palmisano 2016) and other high-level/cognitive factors (e.g., Lepcq et al. 1995; Palmisano and Chan 2004; Riecke et al. 2005, 2006). Below we discuss the various high and low-level contributions to visual self-motion perception in more detail.

Communicated by Melvyn A. Goodale.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00221-019-05674-0>) contains supplementary material, which is available to authorized users.

✉ Takeharu Seno
seno@design.kyushu-u.ac.jp

¹ Department of Ultrastructural Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan

² Physical Intelligence Transfer Technology Team, Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

³ Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, Japan

¹ Please see Palmisano, Allison, Schira and Barry 2015 for other self-motion related uses of the term vection.

⁴ Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, 1-21-40 Korimoto, Kagoshima 890-0065, Japan

⁵ RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

⁶ Department of Developmental Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8553, Japan

⁷ University of Wollongong, Wollongong, NSW, Australia 2500

Low-level motion processing in vection

Global patterns of optic flow are regarded to be the major visual stimulus for self-motion perception and vection (e.g., Gibson 1966; Palmisano 1996; Våljamäe and Seno 2016). The term optic flow refers to the way that the patterns of light at our eyes continually change as we move through the environment (Gibson 1966). Global perspective changes in these patterns over time clearly provide important self-motion information (e.g., Gibson, Olum and Rosenblatt 1955). Indeed, research has shown that compelling vection can be induced using *global* motion perspective information alone (e.g., Andersen and Braunstein 1985; Telford and Frost 1993). However, important information about self-motion in depth can also be conveyed by *local* changes in the image sizes of individual objects (i.e., ‘changing-size cues’—see Palmisano 1996; Seya and Shinoda 2018) and *stereoscopic* motion signals (based on changing disparity over time and interocular velocity differences—see Palmisano 1996, 2002; Palmisano et al. 2016, 2019).

Most past studies investigating the role of low-level motion processing in vection have used (non-stereoscopic) schematic dot motion displays as inducing stimuli. These computer-generated displays have typically provided only luminance-defined (or “first-order”) motion signals to observers (which could be detected via spatiotemporal filtering by the neurons in V1—see Adelson and Bergen 1985; Emerson et al. 1992). However, the visual system also normally receives “second-order” motion signals, which are based on stimulus characteristics other than luminance changes, such as contrast, flicker or texture (Chubb and Sperling 1988, 1989; Cavanagh and Mather 1989). Research suggests that this second-order motion contributes much less to vection than first-order motion (Aaen-Stockdale, Ledgeway and Hess 2007; Gurnsey, Fleet and Potechin 1998; Hanada and Ejima 2000; Seno and Palmisano 2012). For example, Gurnsey and colleagues (1998) found that luminance-defined motion induced much stronger vection than the equivalent contrast-defined motion. Similarly, a subsequent study by Seno and Palmisano (2012) found that vection could only be facilitated by adding additional luminance-, not contrast-, defined motion to the optic flow—even when observers were able to clearly report seeing both types of additional display motion.

Higher-level contributions to vection

Self-motions in the real world generate complex patterns of visual motion stimulation that also depend on the 3D shape, as well as the lighting and material properties of any surrounding surfaces. For example, the precise

patterns of light arriving at a moving observer’s eyes will be determined (in part) by the specular and diffuse reflectance properties of such surfaces. Kim, Khuu and Palmisano (2016) found that vection could be induced by optic flow resulting from specular as well as diffuse reflectance, and that the strength of the vection induced by ‘specular-only’ optic flow depended on the perceived relief height of the simulated environmental surfaces. This latter finding shows that vection is not only based on low-level visual motion processing, but also depends on the computations involved in perceiving surfaces and their materials.

Higher-level, top-down cognitive processes have also been shown to play important roles in vection. For example, knowledge about the possibility of actual self-motion has been shown to affect vection induction (Lepecq et al. 1995; Ogawa and Seno 2014; Palmisano and Chan 2004; Schulte-Pelkum et al. 2003; Wright et al. 2006). Studies have also found that vection increases with the naturalness/realism of the visual inducing stimulus (Bonato and Bubka 2006; Bubka and Bonato 2010; Schulte-Pelkum et al. 2003; Riecke et al. 2005, 2006). For example, Riecke et al. (2006) found that photorealistic self-motion simulations induced stronger vection than scrambled and inverted versions of the same visually moving scene. Other studies have also shown that adding natural visual field features to inducing displays improves vection. For example, Bubka and Bonato (2010) found that vection was stronger when their optic flow simulated self-motion relative to naturally colored (as opposed to grayscale or black-and-white) scenes.

There is also evidence that vection can be altered by differences in stimulus meaning (such as changes to either the perceived figure-ground status of objects/surfaces or their semantic meaning; Seno and Fukuda 2011; Seno et al. 2009; Ogawa and Seno 2014). For example, Seno and Fukuda (2012) found that vection could be altered by changing the meaning associated with their motion displays, while holding low-level visual stimulus factors (such as the overall area and speed of the retinal motion stimulation) relatively constant. When simulating the view from a moving train, they found that vection was more compelling when the motion was seen through the cabin’s windows when its doors were closed (as opposed to open) and when the motion of a second train was provided (as opposed to a grating pattern moving at the same speed). Similarly, Ogawa and Seno (2014) found that vertical vection was reduced when the moving elements in their inducing displays had semantic meaning as falling objects—even though these objects (feathers, petals and leaves) were otherwise identical to the dots in their

Table 1 Details of the six original video clips used in this experiment (none were stereo 3D)

No. in database	Title and release date	Duration	Vection strength	Method of construction	Scene details
18	Mind Game (2014)	20 s	176	Computer generated	A pan of the camera shows a small room in an ordinary Japanese home
20	Crayon Shin-chan Serious Battle! Robot Dad Strikes Back (2014)	87 s	192	Hand drawn	A first-person perspective of a man running through a town
28	Stand by Me Doraemon (2014)	44 s	293	Computer generated	Flying through a town. The first-person viewpoint and camerawork show the main character being chased
34	Dragon Ball Z: Battle of Gods (2013)	19 s	138	Hand drawn and Computer generated	A battle scene involving two men flying very quickly through many buildings
120	Porco Rosso (1992)	34 s	57	Hand drawn	This shows a small red airplane taking off from the sea
Standard stimulus	Control condition	30 s	100	Computer generated	A meaningless expanding circular grating

motion control displays (e.g., in terms of color, luminance and size).

One intriguing study by Seno, Ito and Sunaga (2011) has even reported the occurrence of purely metaphorical vection. Participants in their study viewed displays consisting of two stationary curved lines which simulated the winding edges of a road when driving at night. Even though these road displays did not contain any explicit (only implicit) visual motion, they were found to induce vection which was stronger than that found with their control displays (which were also static, but did not resemble a road). Seno and his colleagues found that adding explicitly moving characters to the road display (which gave an impression of driving past a traffic sign) further enhanced the strength of this metaphorical vection.

The current study

This study was aimed at further investigating the contributions of low-, mid- and high-level visual motion information to vection. The research was informed by Seno et al. (2018). This study identified 150 different video clips² from contemporary Japanese animation films that all induced vection, and produced a database³ that described each of them in terms of: (1) the scenario they depicted, (2) the strength of the vection that they induced, and (3) their method of creation. Five of these 150 video clips were chosen to serve as vection stimuli in the present study (see Table 1). While these five stimuli each contained a mixture of low-, mid- and

high-level motion information, they varied in terms of their naturalness/realism and their method of creation (with two of them being hand drawn, two of them being computer generated, and the remaining video clip being a mix of both techniques). These five video clips were also chosen to produce a range of expected vection strengths (based on the previous findings of the Seno et al. 2018 study).

The current study was primarily interested in how the vection induced by these original video clips might be affected by removing their mid- and higher-level information (including their stimulus meanings, chromatic colors, scene lighting, material properties and any metaphorical effects, etc.). Our plan was to compare the vection induced by the original animation clips to that induced by modified versions containing their pure optic flow (i.e., only their low-level visual motion information). The removal of higher-level information was achieved using specialized algorithms that extracted and separated out the low-level motion information from the original video clips, and then moved random dots in the converted clips based solely on this low-level information (see Suzuki et al. 2017). Because this conversion process completely stripped off all shape, color and texture information from the original movies, it was well suited for the purposes of the present study.⁴

If the vection induced by the original video clips is based primarily on its low-level motion information, then vection strength may be relatively unaffected by removing the movie's higher-level information. However, if the vection induced by the original movies relies heavily on higher-level information, then it is possible that the remaining low-level motion signals may not be sufficient to induce compelling vection. We were

² It should be noted that 30 of these 150 video clips had previously been examined by Tokunaga, Ogawa, Ikehata, Masuda, and Seno (2016).

³ This vection video clip database was provided as supplementary materials for Seno et al. 2018.

⁴ Unfortunately, this extraction process also removed some local low-level motion (i.e., the changing-size cues to object motion in depth), so that only first-order, motion perspective remained after the conversion.

also interested in whether the relative importance of higher-level information for vection depended on how the original animation movie was created. To this end, we examined the vection induced by original and pure optic flow versions of hand-drawn movies, computer-generated movies and movies made using a combination of both techniques (i.e., they had both hand-drawn and computer-generated elements).

Method

Ethics statement

The study was pre-approved by the Ethics Committee of Kyushu University. Written informed consent was obtained from each participant.

Apparatus

Stimuli were generated and controlled by a computer (Alienware-M18x, Dell, Austin, TX) and presented on a large plasma display (3D Viera 65-inch, Panasonic, Japan, with 1920×1080 pixel resolution at a 60-Hz refresh rate). The maximum luminance (R, G, B = 255, 255, 255) was 43.9 cd/m^2 . The experiments were conducted in a dark chamber. Each pixel subtended about 0.05×0.05 degree/cycle.

Participants

Fourteen adult volunteers participated in this experiment. They were either graduate or undergraduate students aged between 20 and 52 years (mean age 25.1 (SD=8.2), 11 males and 3 females). All of these participants were unaware of the purpose of the experiment. They were of sound physical and mental health, had normal color vision and eyesight, and had no history of any of the following conditions: ear pain or headaches when boarding aircraft, vestibular system diseases, cardiorespiratory diseases, moderate balance disorders, dizziness, or altitude sickness (based on oral self-reports). This study had a within-subjects design, so that each of the participants experienced all 12 of the different experimental conditions.

Visual stimuli

Original movies

Five of the original vection clips used in this study were chosen from the following popular Japanese animation films: 1) “Mind Game” [2004; STUDIO4 °C; clip start time: 1:18:08; clip duration: 20 s]; 2) “Crayon Shin-chan Serious Battle! Robot Dad Strikes Back” [2014; SHIN-EI ANIMATION; clip start time: 12:12:03; clip duration: 87 s]; 3) “Stand By Me Doraemon” [2014; Shirogumi, ROBOT, SHIN-EI

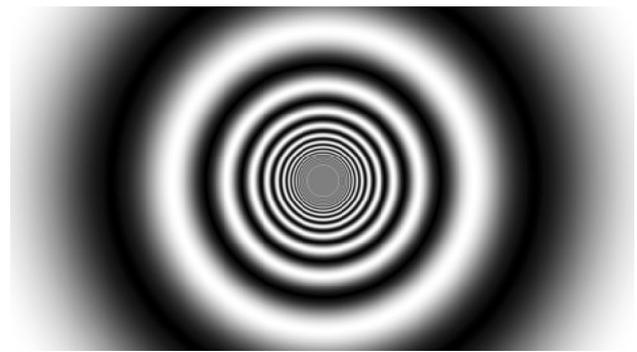


Fig. 1 A single frame of the original expanding grating stimulus

ANIMATION; clip start time: 12:12:43; clip duration: 44 s]; 4) “Dragon Ball Z: Battle of Gods” [2013; TOEI ANIMATION; clip start time: 1:19:29; clip duration: 19 s]; and 5) “Porco Rosso” [1992; STUDIO GHIBLI; clip start time: 12:21:08; clip duration: 34 s]. While these video clips all provided coherent patterns of optic flow, they varied greatly in terms of their actual content (such as their flow types, speeds, densities, luminance, contrast, simulated material properties and surface optics, presence/absence of characters and their actions) and their playing times (19–87 s). In addition to these complex animation-based video clips, we also used a simpler expanding circular grating display in the current study (see Fig. 1). This grayscale display had a 30 s play duration and simulated constant velocity forward self-motion (based on its stimulus, spatial frequency decreasing as eccentricity increased). We had previously obtained vection ratings for all six of these video clips. In the previous Seno et al. (2018) study, participants were told to rate the vection induced by the expanding grating as having a strength of “100”. As can be seen in Table 1, four of the five Japanese animation video clips were found to induce stronger vection than the expanding grating stimulus (with their strength ratings ranging from 138 to 293; only the “Porco Rosso” clip was found to induce weaker vection (57) than the expanding grating stimulus).

Pure optic flow movies

We also created six new “low-level motion only” versions of the original movies described above (i.e., by estimating/extracting the pure optic flow from these five Japanese animation video clips and the expanding grating stimulus). The dynamic vector field of each original movie was initially extracted using an algorithm developed by Suzuki et al. (2017).⁵ The algorithm required that the color videos first be

⁵ The original motivation for developing this algorithm was to separate motion and shape information that interacts in the primate visual association cortices. A preliminary version of this algorithm was found to successfully activate the visual motion area MT (or V5) of a primate (Suzuki et al. 2017).

Table 2 Dot numbers and densities for the converted movies

No. in database	Title and release date	Dot number per frame	Dot density
18	Mind Game (2014)	486	0.02%
20	Crayon Shin-chan Serious Battle! Robot Dad Strikes Back (2014)	436	0.02%
28	Stand by Me Doraemon (2014)	133	0.01%
34	Dragon Ball Z: Battle of Gods (2013)	883	0.04%
120	Porco Rosso (1992)	278	0.01%
Standard stimulus	Control condition	1853	0.09%

converted into grayscale (this conversion was not necessary in the case of the expanding grating stimulus). Then after applying a spatial low-pass filter (circular averaging filter, radius: three pixels), the normalized local normal vectors of the frame images for each pixel were extracted and projected onto the x - y plane using Eqs. 1 and 2. The x and y components of the normalized local normal vector of the frame image projected onto the x - y plane were calculated as:

$$n_x = -\frac{1}{\sqrt{\frac{\partial I(x,y,t)^2}{\partial x^2} + \frac{\partial I(x,y,t)^2}{\partial y^2} + 1}} \frac{\partial I(x,y,t)}{\partial x}, \tag{1}$$

$$n_y = -\frac{1}{\sqrt{\frac{\partial I(x,y,t)^2}{\partial x^2} + \frac{\partial I(x,y,t)^2}{\partial y^2} + 1}} \frac{\partial I(x,y,t)}{\partial y}. \tag{2}$$

where $I(x,y)$ is the pixel value at position (x, y) . Then, the time derivatives of the projected local normal vectors were calculated by subtracting them from those in the subsequent frame. We approximated the normalization factor in Eqs. 1 and 2 by the constant value computed for the last frame. The local normal vectors were also calculated with contrast reversals.

Next, the pure optic flow of the original movie was then estimated based on this calculated vector field. The vector of each pixel in the dynamic vector field was considered to be a visual feature, and its direction and magnitude were used for feature tracking. Although the magnitude of the vector of each pixel was not related to the pixel movement itself, the direction of the vector did indicate the direction of the pixel movement (Suzuki et al. 2017). The algorithm used to estimate the optic flow examined whether there was: a) constant luminance in a local spatial region between three neighboring frames; and b) linear motion in a local time scale between these three frames (i.e., if so, the vector field in the local region between two frames should be the same). Suppose a pixel has a vector with time derivatives of the projected local normal vectors (as described in Eqs. (1) and (2)) that were calculated by subtracting two neighboring t -th and $t + 1$ -th frames. Then the algorithm would search

for a pixel in the next $t + 1$ -th to $t + 2$ -th frame to obtain a local normal vector with the same direction and magnitude. Search areas would be restricted to a line in the direction of the local normal vector of the particular target pixel. If two pixels were found to have the same vector in the next frame, the closest pixel to the target pixel would be the one which was matched.

$$A = \{x_{t+1} | v(x_{t+1}) = v(x_t), x_{t+1} - x_t = av(x_t), a \in R\}, \tag{3}$$

$$x_{t+1}^{track}(x_t) = \arg \min_{x_{t+1}} d(x_{t+1} - x_t), x_{t+1} \in A, \tag{4}$$

where x_t and x_{t+1} indicate the position (x,y) of the pixel at t -th and $t + 1$ -th frame, respectively; $v(x_t)$ and $v(x_{t+1})$ indicate the vector of the pixel x_t and x_{t+1} , respectively; A is a group of pixels at $t + 1$ -th frame whose vector is the same as that of a pixel at the t -th frame and is in line with the direction of $v(x_t)$ of the pixel x_t ; $x_{t+1}^{track}(x_t)$ indicates the tracked pixel of pixel x_t at $t + 1$ -th frame; and $d(x_{t+1} - x_t)$ is the Euclidean distance between the pixels x_t and x_{t+1} . The motion information of the movie should remain the same after contrast reversal (Suzuki et al. 2017), which results in the following tracking criterion.

$$A = \{x_{t+1} | v(x_{t+1}) = \pm v(x_t), x_{t+1} - x_t = av(x_t), a \in R\}. \tag{5}$$

There was a case where no pixel at $t + 1$ -th frame fulfilled this criterion. The pixels that did not move or that moved only with a small magnitude (< 2 pixels) were ignored. Tracked pixels were each visualized by a dot with a radius of 3 pixels presented on a black background. The resulting dot numbers and dot densities for each of the converted movies are shown in Table 2.

Design

Two independent variables were manipulated in this within-subjects experiment: (1) MOVIE. The vection-inducing stimuli used were based on six different video clips: “Grating”, “Dragonball”, “Porco Rosso”, “Mindgame”, and “Crayon” (see Tables 1 and 2) INFORMATION TYPE.

These vection inducing stimuli were either the original video clips (containing a mix of low-, mid- and high-level information) or their extracted pure optic flow (containing only low-level information). Four dependent variables were measured for each of these 12 MOVIE by INFORMATION TYPE conditions, including: (1) the vection onset latency (i.e., the time from the start of the display motion until the participants first felt that they were moving); (2) the %-duration of vection (i.e., the proportion of time that vection was experienced as a function of the total duration of the motion exposure for that trial); (3) the strength of the vection experienced by the participant (rated from 0 to 100); and (4) the perceived scene depth (rated from 0 to 100) for each display.

Procedure

Seated participants viewed each of the above-mentioned motion stimuli from a distance of 57 cm, which resulted in them stimulating a visual area of $100.2^\circ \times 71^\circ$ (note: no fixation point was provided and there was no chin rest). Each trial presented 1 of the 12 different stimuli (the 6 original and 6 low-level motion movies). Participants were asked to press a space key whenever they perceived self-motion (or vection) and keep that key depressed as long as they continued to experience vection. In addition to recording vection onset latency and duration, participants also rated the subjective strength of their experience of vection directly after each trial (using a 101-point rating scale where 0 represented no vection and 100 represented very strong vection, the experimenter entered the rating into the program using the corresponding keys on the keyboard; these procedures have been used in several previous studies, e.g., Seno et al. 2013, 2015, 2017). After obtaining the vection strength rating for the trial, we also obtained a rating of the perceived depth of the display (again using a 101-point rating scale, where 0 indicated that the display was perceived to be flat and 100 indicated that they perceived that the scene extended a large distance behind the screen). Each stimulus condition was repeated four times; thus, there were 48 trials in total. The order of conditions was fully randomized.

Results

Effects of mid- and high-level information on vection and depth

The vection onset latencies, %-durations of vection, vection strength ratings and perceived depth ratings for the six original movies (low-, mid- and high-level information) and six pure optic flow (low-level information only) versions of these movies are shown in Fig. 2a–d. These datasets were each subjected to separate two (INFORMATION TYPE: Original or Pure Optic Flow) by six (MOVIE: “Grating”,

“Dragonball”, “Porco Rosso”, “Mindgame”, or “Crayon”) repeated measures ANOVAs (Greenhouse–Geisser corrections were applied whenever the assumption of sphericity was violated). Post hoc multiple comparisons were performed using the Tukey–Kramer method.

Vection onset latency

The main effect of INFORMATION TYPE was not significant for vection onset latency, $F(1,13) = 0.088$, $p = 0.771$, $\eta_p^2 = 0.007$, suggesting that the original movies did not induce vection more rapidly than the pure optic flow. The main effect of MOVIE was significant, $F(2.234, 29.042) = 16.600$, $p < 0.001$, $\eta_p^2 = 0.561$, indicating that vection onset latency varied significantly depending on which movie was being shown (see Fig. 2a). Of the six movies, “Doraemon” was found to have the shortest ($M = 2.869$ s, $SE = 0.81$ s), and “Crayon” was found to have the longest ($M = 14.854$ s, $SE = 2.609$ s), average vection onset latency. Post hoc comparisons revealed that: (1) “Doraemon” had significantly shorter latencies than all movies except for “Grating” and “Dragonball” ($p < 0.01$); and (2) “Crayon” had significantly longer latencies than all movies except “Porco Rosso” ($p < 0.01$). The interaction between MOVIE and INFORMATION TYPE was not found to be significant for vection onset latency, $F(1.957, 25.447) = 3.298$, $p > 0.05$, $\eta_p^2 = 0.202$.

%-Duration of vection

The main effect of INFORMATION TYPE was significant for %-duration of vection, $F(1,13) = 8.315$, $p = 0.013$, $\eta_p^2 = 0.390$, indicating that the original movies ($M = 49.09\%$, $SE = 4.65\%$) induced larger %-durations of vection on average than the pure optic flow ($M = 38.82\%$, $SE = 4.86\%$). The main effect of MOVIE was significant, $F(2.965, 38.544) = 26.898$, $p < 0.0001$, $\eta_p^2 = 0.674$, indicating that the %-duration of vection varied significantly depending on which movie was shown. Of the six movies, the “Grating” ($M = 60.84\%$, $SE = 6.275\%$) was found to have the largest, and “Porco Rosso” ($M = 17.46\%$, $SE = 4.38\%$) the smallest, average %-duration of vection. The interaction between MOVIE and INFORMATION TYPE was not significant for the %-duration of vection, $F(1.957, 25.447) = 2.640$, $p > 0.05$, $\eta_p^2 = 0.169$.

Vection strength

The main effect of INFORMATION TYPE was significant for vection strength, $F(1,13) = 16.806$, $p = 0.001$, $\eta_p^2 = 0.564$, indicating that the vection induced by the original movies ($M = 58.78$, $SE = 3.49$) was stronger on average than that induced by the pure optic flow ($M = 43.964$,

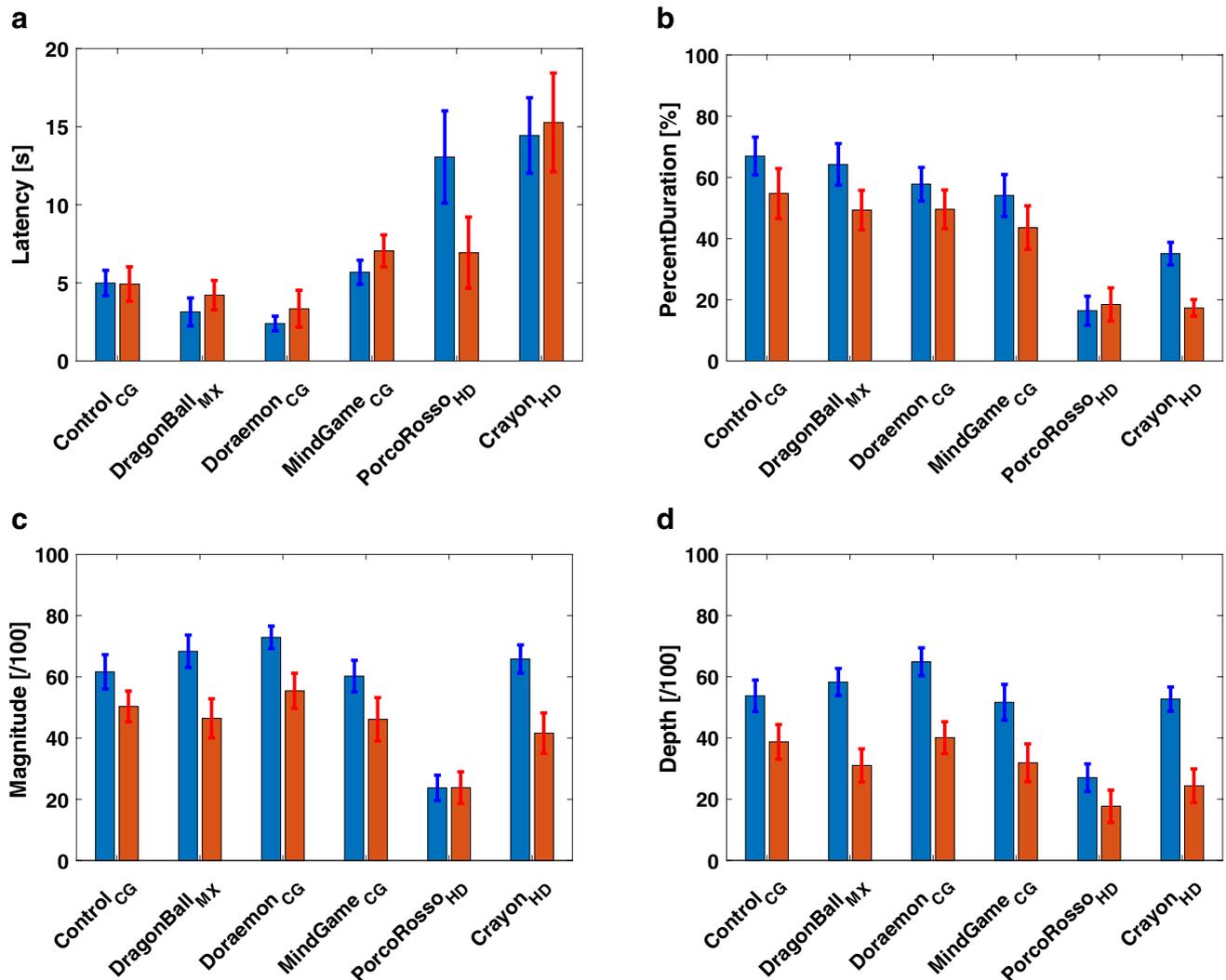


Fig. 2 Mean **a** vection onset latencies, **b** %-durations of vection, **c** vection strength ratings, and **d** perceived depth ratings for each original and pure optical flow movie. The blue and red bars indicate origi-

nal and pure optical flow movies, respectively. The error bars indicate standard error of mean

$SE = 5.06$). The main effect of MOVIE was also significant, $F(2.611, 33.943) = 21.128, p < 0.0001, \eta_p^2 = 0.619$, indicating that vection strength varied significantly depending on which movie was shown. Of the six movies, “Doraemon” was found to have the strongest ($M = 64.17, SE = 3.94$), and “Porco Rosso” the weakest ($M = 23.76, SE = 4.32$), vection strength ratings on average. Post hoc comparisons revealed that: (1) “Doraemon” had significantly stronger vection ratings than all movies except “Grating” and “Dragonball” ($p < 0.01$); and (2) “Porco Rosso” had significantly weaker vection ratings than all of the other movies ($p < 0.01$). The interaction between MOVIE and INFORMATION TYPE was also found to be significant for vection strength ratings, $F(5,65) = 5.187, p < 0.01, \eta_p^2 = 0.285$. As can be seen in Fig. 2c, this was

interpreted as indicating that removing the higher-level visual motion information reduced vection strength ratings for all of the movies except “Porco Rosso”.

Perceived depth

The main effect of INFORMATION TYPE was significant for perceived depth, $F(1,13) = 28.183, p < 0.001, \eta_p^2 = 0.684$, indicating that the scene represented by the original movies ($M = 51.39, SE = 3.01$) was perceived as deeper than that represented by the pure optical flow ($M = 30.64, SE = 4.23$). The main effect of MOVIE was also significant, $F(2.841, 36.937) = 10.396, p < 0.0001, \eta_p^2 = 0.444$, indicating that perceived depth varied significantly depending on which movie was shown. Of the six movies, the scenes in the

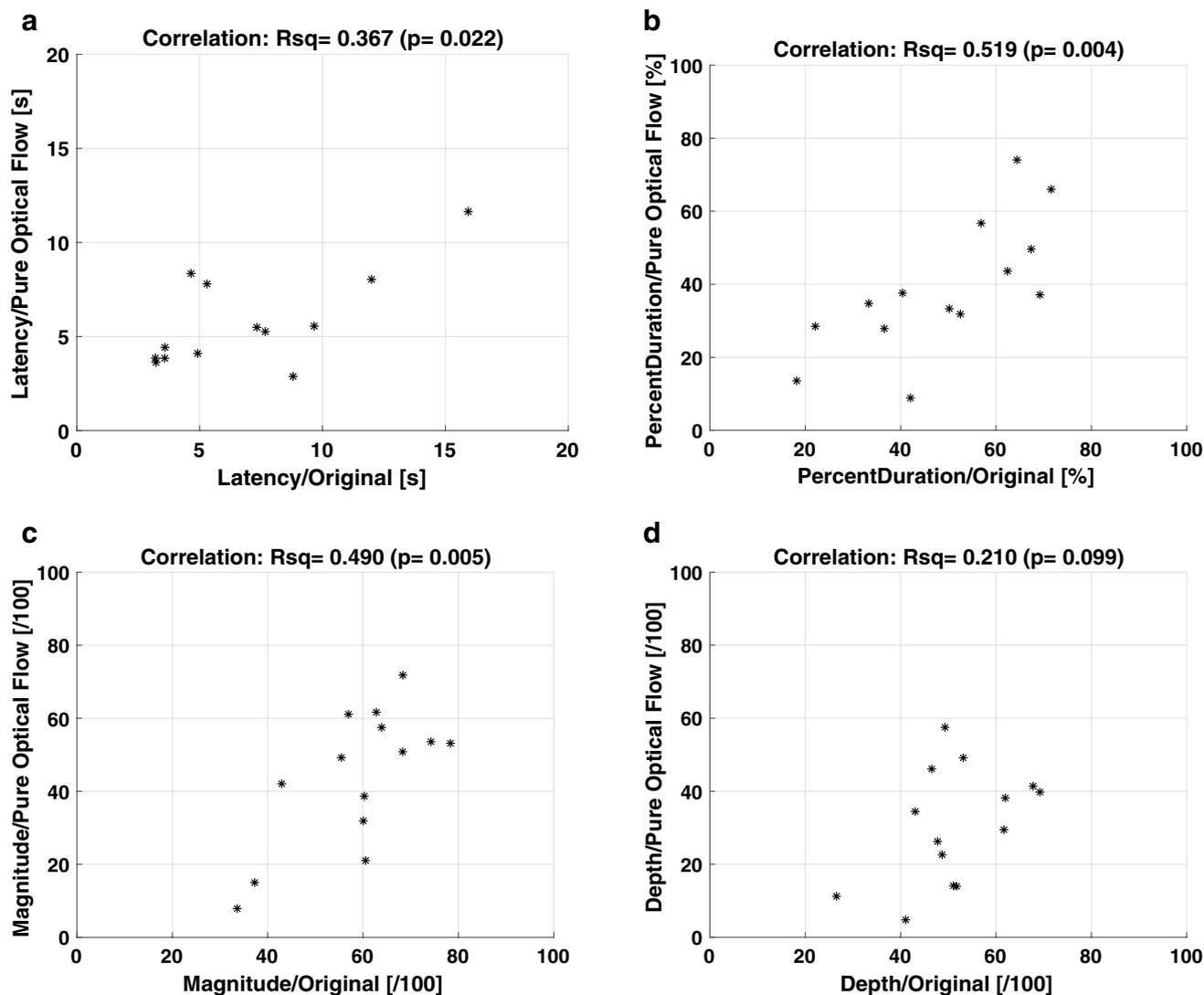


Fig. 3 Scatter plots which show the relationships between the original and the pure optical flow movies in terms of their: **a** vection onset latencies, **b** %-durations of vection, **c** vection strength ratings and **d**

perceived depth ratings. Each data point in these scatter plots represents the average scores for one participant

computer-generated “Doraemon” movie were perceived to be the deepest ($M=52.51$, $SE=3.71$), and those in the hand-drawn “Porco Rosso” movie were perceived to be the flattest ($M=22.35$, $SE=4.24$), on average. The interaction between MOVIE and INFORMATION TYPE was not significant, $F(5,65)=2.364$, $p>0.05$, $\eta_p^2=0.154$.

Correlational analyses

We next examined how consistent the effects of higher- and low-level information were on vection and perceived scene depth across individuals. To this end, we conducted correlation-based analyses which compared the effects of the original and pure optical flow versions of the movies on our participants for each of the dependent measures (i.e., the vection

onset latency, %-duration of vection, vection strength rating and perceived depth rating data). Correlational analyses assume that their data represent independent samples (Lorch and Myers 1990). However, our experiment actually had a repeated measures design. Therefore prior to conducting these analyses, we first calculated each participant’s average vection onset latency (or %-duration, or strength or depth) for the original movies and their average onset latency (or %-duration, or strength or depth) for the pure optic flow versions of these movies. Pearson correlations confirmed that the vection onsets/ %-durations/strength ratings obtained with the pure optical flow (which contained only low-level information) strongly predicted the vection onsets/ %-durations/strength ratings obtained with the original movies (which also contained mid- and high-level information) (Pearson correlations

for vection onset latency: $R=0.61$, $p<0.05$; vection magnitude: $R=0.70$, $p=0.01$; %-duration of vection, $R=0.72$, $p<0.01$ —see Fig. 3a–d). However, the correlation between the perceived depth ratings for the original and pure optic flow movies was not found to be significant ($R=0.46$, $p>0.05$). These findings show that across participants the rank ordering of their vection experiences from least compelling to most compelling (in terms of its strength, onset latency and duration) remained highly consistent irrespective of whether their displays had higher-level or only low-level motion information. However, this was not the case for perceived depth, where there was not a significant relationship between an individual's depth ratings for displays with higher-level motion information and those for displays with only low-level motion information.

Effects of type of animation film (computer generated vs hand drawn)

We were also interested in whether computer-generated and hand-drawn movies were differently affected by the loss of mid- and high-level information. Thus, for each dependent variable (i.e., vection onset latency, %-duration, vection strength ratings and perceived depth), we calculated the average difference between the values obtained during the original movies and the pure optic flow for each participant. We then compared these average differences for the computer-generated (i.e., “Doraemon” and “Mindgame”) and hand-drawn movies (i.e., “Porco Rosso” and “Crayon”) using paired-samples t -tests (the vection and depth data for “Dragonball” was not included in these analyses as this film was created using a mix of hand-drawn and computer-generated techniques). The effects of losing mid- and high-level information were not found to be significantly different for these computer-generated and hand-drawn movies (Vection latency: $t(13)=-1.917$, $p=0.078$ 2-tailed; %-duration of vection $t(13)=0.500$, $p=0.625$ 2-tailed; Vection strength: $t(13)=0.947$, $p=0.361$ 2-tailed; Perceived depth $t(13)=0.628$, $p=0.541$ 2-tailed). However, equivalence tests (Lakens 2017), with the smallest effect sizes of interest being ± 5.77 for Vection latency and ± 14.43 for the other variables,⁶ indicated statistical equivalence for the %-duration of vection ($p_l<0.001$, $p_u<0.001$), Vection strength ($p_l<0.001$, $p_u=0.009$) and Perceived depth ($p_l=0.003$,

$p_u=0.034$), but not for Vection latency ($p_l=0.170$, $p_u<0.001$).

Discussion

In this study, low-level motion information was extracted from six complex (hand-drawn and/or computer-generated) animation movies to create patterns of pure optic flow. We then compared these original and pure optic flow movies in terms of their ability to generate vection and perceived scene depth. As expected, perceived scene depth was significantly reduced by the removal of mid- and high-level information from these movies. However, as can be seen in Fig. 2d, some depth information still remained in the pure optic flow (which must have been based on the remaining motion perspective information).

The main purpose of this study was to investigate how vection is induced and mediated by low-, mid- and high-level motion information. With the exception of the hand-drawn “Porco Rosso” movie, the original movies always induced longer and stronger vection experiences than pure optic flow. While the removal of mid- and high-level information appeared to reduce both of these vection indices,⁷ the vection onset latency was not significantly altered. That is, displays which contained only low-level visual motion information seemed to induce vection as quickly as displays which also contained mid- and high-level visual information. This finding suggests that low-level visual motion information was primarily responsible for the induction of vection in this experiment. However, mid- and higher-level information still appeared to play important roles in enhancing and sustaining these vection experiences once they were triggered. The current results appear to provide a basis for understanding past findings of inconsistencies between these three vection measures (e.g., Seno et al. 2013; also see Seno et al. 2017 for a review). For example, a number of studies have reported that only vection strength ratings, not vection onset latencies, were altered by some display manipulations. Such findings could be explained by these manipulations altering the available mid- and high-level, but not low-level, self-motion information in these displays.

As would be expected based on the above findings, we found that average vection onset latencies for the pure optic flow (which contained only low-level motion information)

⁶ These correspond to one standard deviation of the uniform distribution of the potential range for vection strength, %-duration of vection, and perceived depth. The potential range for latency was estimated to be 40 s. In Seno et al (2018), which examined the same original anime films used in the current study, we obtained only vection magnitude (not latency) data. However, in our other previous studies, when participants were presented with similar self-motion stimuli, they reported vection within 40 s on more than 95% of all trials. This was the reason for that we set 40 s as the upper limit for the vection onset latency in our statistics.

⁷ These two vection advantages for the original movies may have also been due in part to their differences in overall luminance and contrast compared to the pure optic flow. However, it is worth noting that the local contrasts would actually have been higher in the pure optic flow, as they consisted of light white dots against a dark black background (note: there was no way to balance the global contrast and luminance between the pure optic flow and the original movies).

correlated highly with average vection onset latencies for the original movies (which contained extra mid- and high-level information). However, the average %-durations of vection and the average vection strength ratings for the pure optic flow also correlated highly with these same measures for the original movies. These findings indicate that vection responding was highly consistent to both types of movies across our participants. They also suggest that low-level motion was an important contributor to all of three vection indices (although this information was clearly more important for vection onset latency). By contrast, the correlation between the average perceived depth ratings for the original and pure optic flow movies was not significant. This null finding was expected, as the conversion process used to create the pure optic flow was designed to remove important shape, meaning, and 3D layout information. It suggests that mid- and high-level information were much more important for perceived scene depth than they were for vection.

Even though vection and perceived depth did vary significantly depending on which of the original movies was shown,⁸ the effects on vection of removing their higher-level information were quite similar for hand-drawn and computer-generated stimuli. Paired samples *t* tests failed to find any significant differences in these effects across the two types of movies for vection. However, this information manipulation was not confirmed to have equivalent effects for hand-drawn and computer-generated movies in terms of vection onset latency, again pointing to a potentially important difference between this vection measure and the two others. While our information manipulation generally had similar effects on the %-duration and strength of the vection induced by hand-drawn and computer-generated movies, the hand-drawn “Porco Rosso” movie appeared to be an exception. We found a significant interaction between INFORMATION and MOVIE for vection strength. We interpreted this finding as indicating that the removal of mid- and high-level information from the “Porco Rosso” movie had little effect on its vection strength ratings, even though it reduced the vection strength ratings for the other five movies. Since the original version of “Porco Rosso” was always perceived to be flatter, and induced less compelling vection than the other original movies, this suggests that it contained less higher-level information. This would explain why removing the little mid- and higher-level information that it contained had so little effect on vection strength.

⁸ Probably due in part to the different durations of these original movies (which ranged from 19 to 87 s). It is possible that longer movies allowed vection to build up to a greater strength (see Seno et al. 2018).

Relating vection and depth findings to physiology

Neuroimaging research has identified a number of cortical regions that are sensitive to visual motion. Many of these regions lie along the dorsal visual pathway, which starts at the primary visual cortex (V1) in the occipital lobe, and passes through V2, V3 and MT/V5, and then onto the parietal lobe. Suzuki et al. (2017) recently examined the activity of MT/V5 neurons in response to *object* motion movies and their corresponding random dot movies (following the dynamic vector field calculated by the same method used in the current study). The naturalistic movies showed moving animals, human actors and artificial objects. While these original movies contained a mix of low-, mid- and high-level motion information, their vector fields (which were the basis of the pure optic flow displays) did not contain mid- and high-level motion information. Suzuki and colleagues found that their random dot movies generated very similar neural responses in marmoset monkey MT to the original movies, suggesting that this area was primarily activated by low-level visual motion information.

Visual stimuli that are consistent with self- (as opposed to object-) motion have been found to generate activity in V1, V2, V3, MT, MT+, MST, V6 and several parietal, cingulate and insular visual and multisensory association areas (de Jong et al. 1994; Brandt et al. 1998; Morrone et al. 2000; Previc et al. 2000; Wiest et al. 2001; Thilo et al. 2002; Kleinschmidt et al. 2002; Deutschlander et al. 2004; Uesaki and Ashida 2015; Wada et al. 2016; Wall and Smith 2008). Many studies have reported that global optic flow activates area MST in macaques (e.g., Tanaka et al. 1989; Saito et al. 1986; Tanaka and Saito 1989; Duffy and Wurtz 1991; Lagae et al. 1994; see also the review by Duffy and Wurtz 1991) and area MT+ in humans (Cheng et al. 1995; Pitzalis et al. 2010). These findings suggest that much of the visual self-motion processing is done in the dorsal stream. The dorsal stream is thought to be responsible for the preconscious Gibsonian pickup of visual self-motion information (Norman 2002; also see Palmisano, Allison, Schira and Barry, 2015 for a discussion). By contrast, the ventral visual pathway—which also starts at V1 but goes through V2, V4 and onto areas of the inferior temporal lobe—is thought to be responsible for conscious perception (Goodale and Milner 1992). While dorsal system processing is clearly involved, vection is also a conscious experience of self-motion. This conscious aspect of vection implicates the involvement of ventral processing. Consistent with this notion, Deutschlander and colleagues (2004) have reported that vection also activates the fusiform gyrus. The current results suggest that low-level visual motion information is particularly important for vection induction, whereas mid- and higher-level

display information appears to be important for sustaining and strengthening the conscious experience of self-motion after its initial induction (presumably requiring the involvement of both the dorsal and ventral streams).

Conclusions

This paper provides evidence that low- and higher-level information serve different functions in the experience of vection. While low-level motion information is particularly important for vection induction, mid- and high-level display information appears to be important for sustaining and strengthening the vection experience after its initial induction. To our knowledge, the current results are the first clear demonstration of these different functions, although it has been previously proposed that low- and high-level motion information might modulate the vection experience differently.

References

- Aaen-Stockdale C, Ledgeway T, Hess RF (2007) Second-order optic flow processing. *Vis Res* 47:1798–1808
- Adelson EH, Bergen J (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299
- Andersen GJ, Braunstein ML (1985) Induced self-motion in central vision. *J Exp Psychol Hum Percept Perform* 11:122–132
- Bonato F, Bubka A (2006) Chromaticity, spatial complexity, and self-motion perception. *Perception* 35:53–64
- Brandt T, Wist ER, Dichgans J (1971) Optischinduzierte pseudocoriolis-effekte und circularvektion: ein beitrag zur optisch-vestibulären interaktion. *Arch Psychiatr Nervenkr* 214:365–389. <https://doi.org/10.1007/BF00342671>
- Brandt T, Dichgans J, Koenig E (1973) Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Exp Brain Res* 16:476–491. <https://doi.org/10.1007/BF00234474>
- Brandt T, Bartenstein P, Janek A, Dieterich M (1998) Reciprocal inhibitory visual-vestibular interaction. Visual motion stimulation deactivates the parieto-insular vestibular cortex. *Brain* 121:1749–1758
- Bubka A, Bonato F (2010) Natural visual-field features enhance vection. *Perception* 39:627–635. <https://doi.org/10.1068/p6315>
- Cavanagh P, Mather G (1989) Motion: the long and short of it. *Spat Vis* 4:103–129
- Cheng K, Fujita H, Kanno I, Miura S, Tanaka K (1995) Human cortical regions activated by wide-field visual motion: an H2(15)O PET study. *J Neurophysiol* 74(1):413–427
- Chubb C, Sperling G (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J Opt Soc Am A* 5:1986–2007. <https://doi.org/10.1364/josaa.5.001986>
- Chubb C, Sperling G (1989) Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proc Natl Acad Sci* 86:2985–2989. <https://doi.org/10.1073/pnas.86.8.2985>
- de Jong BM, Shipp S, Skidmore B, Frackowiak RS, Zeki S (1994) The cerebral activity related to the visual perception of forward motion in depth. *Brain* 117:1039–1054
- Deutschlander A, Bense S, Stephan T, Schwaiger M, Dieterich M, Brandt T (2004) Roll vection versus linear vection: comparison of brain activation in PET. *Hum Brain Mapp* 21:143–153
- Duffy CJ, Wurtz RH (1991) Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J Neurophysiol* 65:1329–1345
- Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vis Res* 32:203–218
- Gibson JJ (1966) *The senses considered as perceptual systems*. Houghton Mifflin, Boston
- Gibson JJ, Olum P, Rosenblatt F (1955) Parallax and perspective during aircraft landings. *Am J Psychol* 68(3):372–385
- Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci* 15(1):20–25. [https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/10.1016/0166-2236(92)90344-8)
- Gurnsey R, Fleet D, Potechin C (1998) Second-order motions contribute to vection. *Vis Res* 38:2801–2816
- Hanada M, Ejima Y (2000) Heading judgement from second-order motion. *Vis Res* 40:3319–3331
- Kim J, Khuu S, Palmisano S (2016) Vection depends on perceived surface properties. *Atten Percept Psychophys* 78:1163–1173. <https://doi.org/10.3758/s13414-016-1076-9>
- Kleinschmidt A, Thilo KV, Buchel C, Gresty MA, Bronstein AM, Frackowiak RS (2002) Neural correlates of visual-motion perception as object- or self-motion. *Neuroimage* 16:873–882
- Lagae L, Maes H, Raiguel S, Xiao D-K, Orban GA (1994) Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J Neurophysiol* 71:1597–1626
- Lakens D (2017) Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc Psychol Personal Sci* 8(4):355–362
- Lepecq JC, Giannopulu I, Baudonniere PM (1995) Cognitive effects on visually induced body motion in children. *Perception* 24:435. <https://doi.org/10.1068/p240435>
- Lorch RF, Myers JL (1990) Regression analyses of repeated measures data in cognitive research. *J Exp Psychol Learn Mem Cogn* 16(1):149–157
- Morrone MC, Tosetti M, Montanaro D, Fiorentini A, Cioni G, Burr DC (2000) A cortical area that responds specifically to optic flow, revealed by fMRI. *Nat Neurosci* 3(12):1322–1328. <https://doi.org/10.1038/81860>
- Norman J (2002) Two visual systems and two theories of perception. *Behav Brain Sci* 25:73–144. <https://doi.org/10.1017/S0140525X0200002X>
- Ogawa M, Seno T (2014) Vection is modulated by the semantic meaning of stimuli and experimental instructions. *Perception* 63:605–615
- Palmisano S (1996) Perceiving self-motion in depth: the role of stereoscopic motion and changing-size cues. *Percept & Psychophys* 58(8):1168–1176
- Palmisano S (2002) Consistent stereoscopic information increases the perceived speed of vection in depth. *Perception* 31(4):463–480
- Palmisano S, Chan AY (2004) Jitter and size effects on vection are immune to experimental instructions and demands. *Perception* 33:987–1000
- Palmisano S, Allison RS, Schira MM, Barry RJ (2015) Future challenges for vection research: definitions, functional significance, measures and neural bases. *Front Psychol* 6(193):1–15
- Palmisano S, Summersby S, Davies RG, Kim J (2016) Stereoscopic advantages for vection induced by radial circular and spiral optic flow. *J Vis* 16(14):1–19

- Palmisano S, Davies R, Brooks K (2019) The stereoscopic advantage for vection increases with simulated eye-separation. *Atten Percept Psychophys* 81(1):281–295
- Pitzalis S, Sereno MI, Committeri G, Fattori P, Galati G, Patria F, Galletti C (2010) Human v6: the medial motion area. *Cereb Cortex* 20(2):411–424
- Previc FH, Liotti M, Blakemore C, Beer J, Fox P (2000) Functional imaging of brain areas involved in the processing of coherent and incoherent wide field-of-view visual motion. *Exp Brain Res* 131:393–405
- Riecke BE, Västfjäll D, Larsson P, Schulte-Pelkum J (2005) Top down and multi-modal influences on self-motion perception in virtual reality. 11th International conference on Human-Computer Interaction (HCI International 2005). Erlbaum, Mahwah, pp 1–10
- Riecke BE, Schulte-Pelkum J, Avraamides MN, Heyde MVD, Bühlhoff HH (2006) Cognitive factors can influence self-motion perception (vection) in virtual reality. *ACM Trans Appl Percept* 3(3):194–216. <https://doi.org/10.1145/1166087.1166091>
- Saito H, Yukie M, Tanaka K, Hikosaka K, Fukada Y, Iwai E (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J Neurosci* 6:145–157
- Schulte-Pelkum J, Riecke BE, von der Heyde M, Bühlhoff HH (2003) Circular vection is facilitated by a consistent photorealistic scene. 6th Annual Workshop of Presence (Presence 2003). Aalborg, Denmark
- Seno T, Fukuda H (2011) Stimulus meanings alter illusory self-motion (vection)—experimental examination of the train illusion. *Seeing Perceiving* 25(6):631–645
- Seno T, Fukuda H (2012) Stimulus meanings alter illusory self-motion (vection)—experimental examination of the train illusion. *Seeing Perceiving* 25(6):631–645. <https://doi.org/10.1163/18784763-00002394>
- Seno T, Palmisano S (2012) Second-order motion is less efficient at modulating vection strength. *Seeing Perceiving* 25(2):213–221
- Seno T, Ito H, Sunaga S (2009) The object and background hypothesis for vection. *Vis Res* 49:2973–2982
- Seno T, Ito H, Sunaga S (2011) Attentional load inhibits vection. *Atten Percept Psychophys* 73:1467–1476
- Seno T, Abe K, Kiyokawa S (2013) Wearing heavy iron clogs can inhibit vection. *Multisens Res*. 26(6):569–580
- Seno T, Sawai KI, Kanaya H, Wakebe T, Ogawa M, Fujii Y, Palmisano S (2017) The Oscillating Potential Model of Visually Induced Vection. *Iperception* 8(6):2041669517742176
- Seno T, Murata K, Fujii Y, Kanaya H, Ogawa M, Tokunaga K, Palmisano S (2018) Vection increases with optic flow exposure duration. *Iperception* 9(3):1–16
- Seya Y, Shinoda H (2018) Relationship between vection and motion perception in depth. *Atten Percept Psychophys* 80(8):2008–2021
- Suzuki W, Ichinohe N, Tani T, Hayami T, Miyakawa N, Watanabe S, Takeichi H (2017) Novel method of extracting motion from natural movies. *J Neurosci Met* 291:51–60
- Tanaka K, 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
- Tanaka K, Fukada Y, Saito H (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
- Telford L, Frost BJ (1993) Factors affecting the onset and magnitude of linear vection. *Percept Psychophys* 53:682–692
- Thilo KV, Kleinschmidt A, Gresty MA (2002) Perception of self-motion from peripheral optokinetic stimulation suppresses visual evoked responses to central stimuli. *J of Neurophysiol* 90:723–730
- Tokunaga K, Ogawa M, Ikehata S, Masuda T, Seno T (2016) Constitution of a database of vection scenes in Japanese movies and animations and experimental assessments of them. *Transact Virtual Reality Soc Japan* 21:35–47. https://doi.org/10.18974/tvrsj.21.1_35 (in Japanese)
- Uesaki M, Ashida H (2015) Optic-flow selective cortical sensory regions associated with self-reported states of vection. *Front Psychol* 6:775
- Väljamäe A, Seno T (2016) Modulation of recognition memory for emotional images by vertical vection. *Front Psychol* 7:39
- Wall MB, Smith AT (2008) The representation of egomotion in the human brain. *Curr Biol* 18(3):191–194. <https://doi.org/10.1016/j.cub.2007.12.053>
- Wada A, Sakano Y, Ando H (2016) Differential responses to a visual self-motion signal in human medial cortical regions revealed by wide-view stimulation. *Front Psychol* 7:309
- Wiest G, Amorim MA, Mayer D, Schick S, Deecke L, Lang W (2001) Cortical responses to object-motion and visually induced self-motion perception. *Cogn Brain Res* 12:167–170
- Wright WG, DiZio P, Lackner JR (2006) Perceived self-motion in two visual contexts: dissociable mechanisms underlie perception. *J Vestib Res* 16:23–28

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.