Vestibular nuclei and cerebellum put visual gravitational motion in context

William L. Miller¹,²*, Vincenzo Maffei¹, Gianfranco Bosco³,⁴, Marco Iosa¹, Myrka Zago¹, Emiliano Macaluso², and Francesco Lacquaniti¹,³,⁴

¹Department of Neuromotor Physiology and ²Neuroimaging Laboratory, Santa Lucia Foundation, via Ardeatina 306, 00179 Rome (Italy), ³Department of Neuroscience and ⁴Center of Space Biomedicine, University of Rome Tor Vergata, Via Montpellier 1, 00133 Rome, Italy.

Abbreviated title: Visual context effects in brainstem and cerebellum

*Correspondence to:

William L. Miller, Ph.D.
Department of Neuromotor Physiology
Santa Lucia Foundation
via Ardeatina 306
00179 Rome, Italy
ph +39-06-51501483
fax +39-06-51501482
e-mail: w.miller@vegascientific.com
Abstract

Animal survival in the forest, and human success on the sports field, often depend on the ability to seize a target on the fly. All bodies fall at the same rate in the gravitational field, but the corresponding retinal motion varies with apparent viewing distance. How then does the brain predict time-to-collision under gravity? A perspective context from natural or pictorial settings might afford accurate predictions of gravity’s effects via the recovery of an environmental reference from the scene structure. We report that embedding motion in a pictorial scene facilitates interception of gravitational acceleration over unnatural acceleration, whereas a blank scene eliminates such bias. Functional magnetic resonance imaging (fMRI) revealed BOLD correlates of these visual context effects on gravitational motion processing in the vestibular nuclei and posterior cerebellar vermis. Our results suggest an early stage of integration of high-level visual analysis with gravity-related motion information, which may represent the substrate for perceptual constancy of ubiquitous gravitational motion.

Keywords

fMRI, visuomotor control, internal models, insula, temporo-parietal junction, interception, time-to-contact.
Introduction

Perception and interaction with objects accelerated by Earth’s gravity (g) are important skills in daily life. To predict when an observed object in free-fall will reach us or contact ground, we can estimate the remaining time-to-contact, $TTC(t)$. In line of principle, $TTC(t)$ could be estimated directly from retinal and gaze-position signals, that is, without the need to integrate high-order contextual cues about the scale of the visual scene. Here, however, we show that pictorial information does affect $TTC(t)$ estimates, facilitating interception of gravitational acceleration over unnatural acceleration.

A number of studies have addressed the specific roles in $TTC(t)$ estimation of retinal and gaze-position signals such as dilation of the retinal image, ratio of target distance to velocity (so called $\tau$ function), reduction of the retinal gap between target and contact point, change in binocular disparity, or evolution of pursuit eye movements (Land and McLeod 2000; Lee et al. 1983; Michaels et al. 2001). However, there is now strong evidence that human accuracy in interception of falling objects involves taking into account $g$, in addition to those cues (Lacquaniti and Maioli 1989b; Lacquaniti et al. 1993; McIntyre et al. 2001; Senot et al. 2005; Zago et al. 2004; 2005; Zago and Lacquaniti 2005). Although all objects are accelerated downward by Earth’s gravity at the same rate, the corresponding acceleration of the retinal image is not at all constant, being inversely related to the apparent viewing distance of the object (assuming that the target motion plane is orthogonal to the line of sight). Arbitrary optic accelerations are not well discriminated over short time intervals (Brouwer et al 2002; Werkhoven et al 1992), and generally they are not taken into account in timing manual interceptions (Port et al. 1997; Senot et al. 2003). Thus, it was hypothesized that an internal model calculating the specific effects of $g$ on a falling target is stored in the brain and is used to interpret visual motion that appears consistent with gravity (Lacquaniti and Maioli 1989a; 1989b; Lacquaniti et al. 1992; 1993). Evidence for this internal model was provided by the observation that, in the absence of $g$-determined sensory cues, astronauts initially expect the effects of $g$ on a dropped object (moving at constant speed) when they
attempt to catch it in the Spacelab, and they adapt to the new environment only after many days of space flight (McIntyre et al. 2001).

The internal model hypothesis poses the question of how the brain combines target motion, which is mapped topographically on the retina, with an internal representation of $g$ specified in the world coordinates of the visual scene. This combination may require reference to a common spatial frame. Thus, retinal motion information might be scaled by apparent viewing distance to estimate target motion in world coordinates. Eye vergence, accommodation and stereodisparity may contribute to estimating viewing distance of target motion in 3D space, but these cues are ineffective when the target is far off or when it moves on a 2D video display (as in an interactive videogame). Presupposed knowledge of $g$ might cue apparent viewing distance in all cases, but the evidence is mixed in the absence of a patterned background (Hecht et al. 1996; McConnell et al 1998; Saxberg 1987; Stappers and Waller 1993; Watson et al. 1992).

Here we consider a different possibility, namely that pictorial information from the moving target or the background of the visual scene aids recovery of an environmental reference and scale (DeLucia 1991; Distler et al 2000; McKee and Smallman 1998; Wallach 1939). In other words, a spatial array of natural objects might indicate the reference metric necessary to map between retinal and world motion and to calibrate the effects of $g$ on a visual target. For instance, if an observed object fell near a person, the estimated height of that person could be used to scale the motion of the falling object, effectively recovering the apparent distance from the viewer. Processing of visually rich scenes may routinely involve such transformations of retinal input into allocentric output.

We studied whether the availability of such pictorial cues influences interception accuracy for accelerating and decelerating targets. We measured behavioral and brain responses of human subjects while they intercepted virtual targets descending with either natural ($g$) or reversed ($rg$) gravitational laws of motion under two different visual contexts. $g$ targets decelerated while moving up and
accelerated while moving down, while $rg$ targets accelerated while moving up and decelerated while moving down. In the pictorial context, target motion was embedded in a pictorial scene whose spatial scale in a world reference frame could be roughly estimated from several cues (familiar size, linear perspective, shading, texture gradient). In the non-pictorial context, target motion was embedded in a quasi-uniform background, rendering target size and distance ambiguous in a world reference frame. If subjects relied entirely on the retinal coordinates of the target to guide interception, response timing should be independent of the type of visual context, since target motion was identical in the pictorial and non-pictorial conditions. Instead, we found that response timing did depend on visual context: embedding motion in the pictorial scene facilitated interception of $g$ targets as compared with interception of $rg$ targets, whereas the blank scene eliminated such bias. This is consistent with the hypothesis that subjects used the perceived size of naturalistic objects in the scene to map between retinal and world motion.

We employed fMRI to identify BOLD correlates of visual context effects on gravitational motion processing. A previous fMRI study identified correlates of visual representations of $g$ motion in a pictorial context only (Indovina et al. 2005). In that study, it was found that $g$ trials were associated with significantly more activity than $rg$ trials in a network of regions that were independently activated by vestibular caloric stimuli, and therefore were identifiable as belonging to a multi-modal visual-vestibular network. The network included the insular cortex, temporo-parietal junction (TPJ), premotor areas, middle cingulate cortex, thalamus and putamen. The previous study focused on higher stages of visual processing of targets accelerated by gravity. However, associations between features that mediate spatial reference for gravitational motion may occur across different stages of visuo-vestibular analysis, from subcortical to cortical regions (Angelaki and Hess 2005). Here we employed fMRI to identify correlates of visual context effects on gravitational motion processing by comparing pictorial and non-pictorial conditions. The hypothesis that a pictorial context facilitates the recognition of $g$
effects on a moving target would predict that spatial scale information about the scene be integrated with visual motion information at or before the neural stages where a visual representation of $g$ is used for the interception of falling targets. Consistent with this prediction, we found that the contextual effects were mediated by the vestibular nuclei and posterior cerebellar vermis.
Materials and Methods

Subjects

A total of 70 healthy adults with normal uncorrected vision, right handed dominance and no history of neurological disorders participated in the experiments after giving written informed consent in accordance with the procedures established by the ethics committee of the Santa Lucia Foundation. Predominant use of the right hand was indicated both verbally and according to a short version of the Edinburgh handedness inventory (Oldfield 1971). All subjects were naïve to the stimuli, task, and purpose of the experiments. Thirty-three subjects (16 females and 17 males; mean age 26, range 20-39 years) participated in the main fMRI study. Thirty-four different subjects (20 females and 14 males, mean age 24, range 18-32 years) participated in an additional, purely behavioral study, and three additional subjects (all male, mean age 37, range 25-45 years) participated in an additional fMRI experiment.

fMRI study

Visual stimuli and tasks.

Subjects underwent fMRI in the supine position while viewing visual stimuli and responding by means of a button-press. 768x768 pixel, 32-bit color images were generated with in-house software (compiled C++ with external libraries) and projected with a digital projector (NEC LT158, 60 Hz refresh rate) through an inverted telephoto lens onto a semi-opaque plexiglas screen mounted vertically inside the scanner bore behind the subject's head. The back-projected images were viewed via a mirror mounted on the head coil (Fig. 1c). The total eye-to-screen distance was 66.5 cm, and the size of the projected image was 18 cm x 18 cm, corresponding to 15.4° x 15.4° of visual angle. Subject responses were acquired with an MR-compatible fiber-optic-based button response system (fORP, Current
Designs Inc.), sampled at 1 KHz and registered by the experiment control software through a serial port interface.

Subjects were pseudo-randomly assigned into two gender-balanced groups. Subjects in the *pictorial* group (n=17) viewed target stimuli overlaid upon an image of a human figure standing in front of a building (Fig. 1a). The relative sizes of the elements of the visual scene were consistent with an apparent viewing distance of about 25 m. During *interception* trials, a textured sphere (0.4°) moved vertically relative to the scene at a constant absolute acceleration, first ascending from a box (0.4°) in the person’s hand, elastically bouncing (coefficient of restitution=1) on the building cornice, and then descending into the box. The excursion of the target between the box and the cornice was 9.17°, corresponding to 4 m in the scale of the photograph, and target acceleration was 22.5°·s⁻², corresponding to 9.81 m·s⁻². In *g* trials, target acceleration was consistent with natural gravity, that is, the target decelerated while moving up and accelerated while moving down. In *rg* trials, instead, target acceleration was reversed relative to natural gravity, that is, the target accelerated while moving up and decelerated while moving down. The flight duration (FD) of the target sphere was varied by changing the value of target initial velocity. FD could take one of 5 possible values between 1.40 and 1.78 s, with a corresponding range of mean target speeds of 13.1 - 10.3°·s⁻¹. FD and inter-trial intervals were randomized to make motion onset and duration unpredictable from trial to trial (see *Trial sequence* below). Subjects were instructed to maintain fixation on a red dot (0.24°) located on the box throughout an experimental run, and were asked to press the button with the right thumb so as to intercept the descending target at the expected time of arrival at the fixation point. No feedback of response accuracy was provided.

Subjects in the *non-pictorial* group (n=16) were instead presented with a quasi-uniform background image (Fig. 1b), upon which targets were superimposed with dynamics identical to that in the *pictorial* condition. A starting box with superimposed fixation dot and a filled rectangle were located in the same
absolute positions and served the same functions as the box and the cornice bounce point, respectively, in the *pictorial* background image. Target circle, starting box and top rectangle were all non-textured geometrical figures. Color and luminance of the background in the *non-pictorial* stimulus matched their respective mean values in the *pictorial* stimulus within the region of fixation and target trajectory (mean luminance, 34 cd·m\(^{-2}\), as measured within the scanner), ensuring similarity of target contrast in the two visual context conditions. Otherwise, the stimuli and task were the same as in the *pictorial* condition.

A between-groups design was employed to avoid presentation order effects which had been observed in pilot experiments. Indeed, it is known that when observers see familiar-size scenes first, their subsequent judgments of more abstract scenes may be biased, and vice-versa (DeLucia 2005). To check whether group effects attributed to the interception task in a *pictorial vs. non-pictorial* context were instead related to mere group differences, we included a second visuo-motor task that did not require any prediction of time-to-contact (i.e. a simple reactive task). This task consisted of reactive trials in which subjects maintained fixation on the red dot in the *pictorial* or *non-pictorial* image. Moving targets were not presented in these trials; instead, the fixation point radius expanded by 1.7 times for 200 ms after a total un-cued randomized interval equivalent on average to that employed in *interception* trials (inter-trial interval plus FD) minus 200 ms. Subjects were instructed to press the button with the right thumb as soon as possible after the fixation point expanded. As in the *interception* trials, no performance feedback was provided to subjects.

*Trial sequence.*

Each subject underwent 5 fMRI-runs lasting approximately 6.5 min each. During each run, 95 trials were presented, comprising 35 repetitions of each of the *g* and *rg interception* trial types (36.8% of the total trials), and 25 repetitions of the *reactive* trial type (26.4%). Trials in a single fMRI-run were
grouped into blocks of pseudo-randomized order consisting of either \textit{g} interception (5 blocks), \textit{rg} interception (5 blocks), or reactive trials (4 blocks). In order to resolve differences in BOLD response to individual interception trial types (defined by target acceleration and FD) while maintaining sufficient event detection power, inter-trial intervals throughout a run were randomized according to a long-tailed (geometric) distribution (Hagberg et al. 2001). The resulting mean trial duration was $4.1 \pm 1.2$ s (mean $\pm$ s.d., median 3.8 s, minimum 2.8 s, maximum 11.1 s). At the start of each block, a text cue was displayed for 1 s to alert subjects about the task to perform (interception or reactive), whereas no cue was given about the motion law (\textit{g} or \textit{rg}). The type of the first block of the full 5-run sequence was also varied pseudo-randomly among subjects, to control for any potential influences of first-trial type on subsequent trials.

\textit{Pre-scan and in-scanner fixation tests.}

Prolonged fixation ability was verified in all subjects with a challenging test ($\sim$ 7 min) on the same day and prior to performing the fMRI experiment. Subjects were seated in front of a CRT computer monitor and were instructed to maintain fixation on a stationary illuminated dot on a black background (located at the same relative position as the fixation point in the fMRI visual stimuli) throughout the test. A target (0.4° disk) moved rectilinearly across the screen in the periphery at random constant speed and direction. Subjects were instructed to respond with a button-press as soon as the moving target transiently expanded and changed color (for 200 ms) while always maintaining fixation on the stationary dot. Movements of both eyes were recorded at 500 Hz, mean resolution <0.01°, mean accuracy <0.5° (according to factory-provided specifications), with a head-mounted infrared system (Eyelink II, SR Research, Mississauga, Ontario, Canada).

In addition, to verify that fixation could be maintained adequately throughout a full-length experiment inside the MR scanner, eye movements were monitored in three subjects while they
performed the interception task during fMRI. Stimulus, task (see above) and scanning methods (see below) were identical to those in the main fMRI study except that reactive trials were not included, and there were 90 rather than 75 interception trials for each of the 5 runs (45 g trials per run, 450 total trials, 6.5 minutes per run). All three subjects viewed the pictorial background stimulus. Eye movement data were recorded during experiment execution with an ASL 504 eye tracking system (Bedford, MA, USA) at an acquisition rate of 60 Hz. During offline analysis, trials were flagged as invalid due to excessive missing data points or blinks and excursion trials if positive excursion from the fixation point during target motion exceeded 1° of visual angle for 150 ms (9 consecutive points). The remaining trials were considered valid.

*fMRI data acquisition.*

Imaging was performed with a Siemens Magnetom Allegra 3 Tesla head-only scanning system (Siemens Medical Systems, Erlangen, Germany), equipped with a quadrature volume RF head coil. Subjects were provided with noise suppression apparatus (ear plugs and headphones), and lay supine with the head firmly immobilized with foam cushioning. Whole-brain blood-oxygenation dependent (BOLD) echoplanar imaging (EPI) functional data were acquired with a 3T-optimized gradient echo pulse-sequence (TR=2.47 s, TE=30 ms, Flip angle=70°, FOV=192 mm, fat suppression). Blocks of 38 image slices were acquired in ascending order (64x64 voxels, 3x3x2.5 mm, distance factor 50%). For each participant, a total of 830 volumes of functional data were acquired in 5 consecutive runs. At the end of each run, subjects were given a brief rest period.

*Additional behavioral study*

We performed a second series of behavior-only experiments to control for the possibility that visual context effects on interception are dependent on the alignment of the subject’s body and visually-
determined vertical relative to the gravity vector. In this study, we repeated the protocols with two additional groups of 17 subjects each (one group being assigned to the *pictorial* condition and the other group to the *non-pictorial* condition), who performed the interceptions while seated in front of a CRT monitor in the laboratory. The stimuli and task were identical to those of the fMRI study, with the exception that subjects did not perform *reactive* trials.

**Data analysis**

*Behavioral data analysis.*

For *interception* trials, response timing error (RTE) was computed as the signed difference between the recorded button-press time and the actual arrival time of the target at the fixation point. RTE for *reactive* trials was computed as the difference between the button-press time and the onset of fixation point expansion. We discarded approximately 1% of all behavioral trials collected during the experiments, including those with no response, *interception* responses with RTE larger than 900 ms or smaller than -900 ms, and *reactive* responses occurring either earlier than fixation point expansion or later than 900 ms after fixation point expansion. Mean subject RTE for the different laws of motion (*g* or *rg*), flight durations (5 different values) and visual context (*pictorial* or *non-pictorial*) conditions were submitted to individual *t*-tests and between-groups repeated measures analyses of variance (ANOVA), analyses of covariance (ANCOVA) and linear regression using the Statistica analysis software (Statsoft, Oklahoma). We compared the laboratory- and fMRI-based *pictorial* and *non-pictorial* interception results in a group-by-protocol-by-law of motion ANOVA.

**fMRI data analysis.**

Data pre-processing and statistical analysis were performed using the SPM2 software (Wellcome Department of Cognitive Neurology, University College London; implemented in MatLab version 6.5).
The first four volumes of data from each of the 5 runs were discarded to allow for stabilization of longitudinal magnetization. The remaining 810 volumes were submitted to the following pre-processing steps: realignment of all images to the first volume to compensate for head motion; slice timing adjustment to correct for slice acquisition delays; normalization to Montreal Neurological Institute (MNI) standard space using the default SPM2 EPI image template and the mean of the functional volumes to facilitate group analysis; and 3D image smoothing with an 8-mm FWHM isotropic Gaussian kernel to increase signal-to-noise ratio. Voxel dimensions of the final smoothed images were 2 x 2 x 2 mm. Statistical analysis was carried out in two stages (Penny et al. 2003). For each subject, responses to the experimental stimulus types were first estimated with a fixed effects general linear model (GLM) analysis using the smoothed image time-series. Voxel time-series were pre-processed to remove autocorrelations using a first-order autoregressive model, and high-pass filtering (128 s cut-off). One-parameter regressors of interest were formed for each trial type by convolving the trial onsets with the SPM2's canonical hemodynamic response function, resulting in 11 regressors for each of the five runs (interception trials: separate regressors for the 5 FDs for each of \( g \) and \( rg \) trials; reactive trials: 1 combined regressor). Additional regressors for run constants and for head translation and rotation covariates were included in the design matrix. Second-stage random effects analyses of variance (two-group one-way ANOVA, with sphericity correction) were then carried out on the subtraction contrast images representing the estimated differential BOLD response to \( g \) and \( rg \) trials from the pictorial and non-pictorial subjects (which were generated by applying a weight of 1 on the \( g \) regressors and -1 on the \( rg \) regressors, and vice-versa). Final activation images tables were generated from contrast analyses on the second-stage random effects model, and corrected for multiple comparisons using cluster thresholding based on the theory of random Gaussian fields (Friston et al. 1994). Except where otherwise noted, we employed a probability criterion of \( P \)-corrected \( < 0.05 \) at cluster level (cluster size estimated at \( P \)-uncorrected \( < 0.005 \)), considering the whole brain as
the volume of interest.

The main effect of intercepting $g$ targets versus intercepting $rg$ targets across both visual contexts (main effect of law of motion) was computed as \([g \text{ pictorial} - (rg \text{ pictorial}) + [g \text{ non-pictorial} - (rg \text{ non-pictorial})]\), d.f.=31. The complementary main effect of intercepting $rg$ targets vs. $g$ targets was also computed. The interaction effects between law of motion and visual context were computed as \([g \text{ pictorial} - (rg \text{ pictorial}) - [g \text{ non-pictorial} - (rg \text{ non-pictorial})]\) and as \([g \text{ non-pictorial} - (rg \text{ non-pictorial}) - [g \text{ pictorial} - (rg \text{ pictorial})]\), both with d.f.=31. For the purpose of verifying activation of visual motion areas, we also computed the overall main effect of the interception task (relative to the mean signal) as \([g \text{ pictorial} + (rg \text{ pictorial}) + (g \text{ non-pictorial}) + (rg \text{ non-pictorial})]\).

Given the observation in the behavioral data of an increasing difference in $g$ and $rg$ RTEs with increasing FD across visual context conditions (see Results), we tested for a similar modulation of brain responses by target FD. Using the single-subject GLM estimations described above, we modeled an increasing difference in \((g - rg)\) relative activation with increasing FD as a difference in positive linear trends by simultaneously applying weights of \([-2 -1 0 1 2]\) on the five $g$ regressors and \([2 1 0 -1 -2]\) on the five $rg$ regressors (where each set of regressors is ordered in terms of increasing FD). The resulting contrast images were then submitted to a 1-factor group ANOVA. Additional post-hoc analyses tested for positive or negative trends separately or together for $g$ regressors and $rg$ regressors in a similar way. A related alternative analysis directly assessed the relationship between RTE and brain activity. Rather than imposing weights for linear trends, subject-specific mean RTEs for each FD were used as contrast-weights for the fMRI analysis. As expected, given the observed linear relationship between RTE and FD (see Results), this alternative analysis resulted in a pattern of activation no different from the original analysis, and therefore it will not be discussed further.

In order to investigate whether differential \((g - rg)\) brain activation changed over the course of the fMRI experiment, we created single-subject contrast images coded for a positive or a negative linear
trend in \((g - rg)\) across the five consecutive fMRI runs. Specifically, we weighted the \(g\) regressors for each run with 1 and \(rg\) regressors with -1, and then further multiplied the five sets of regressor weights for the five consecutive runs with \([-2, -1, 0, 1, 2]\) (for a positive trend) or \([2, 1, 0, -1, -2]\) (for a negative trend). The resulting contrast images for positive and negative trends across runs were then submitted to separate 1-factor group ANOVAs.

For all types of \((g - rg)\) contrasts, activity profiles were plotted as the difference \(t\)-values for \(g\) and \(rg\) trials for the \textit{pictorial} and \textit{non-pictorial} contexts, respectively, at the significant local maxima.

We compared brain responses to \textit{reactive} trials between the \textit{pictorial} and \textit{non-pictorial} groups of subjects to control for potential group differences in visuomotor activation in a simple non-motion reaction task which should not depend on visual context. Contrast images for \textit{pictorial} and \textit{non-pictorial reactive} trials were thus compared with a 1-factor group ANOVA.

Finally, to assess whether the midline cerebellar activations identified in the main study (see \textit{Results}) might be dependent on eye movements, we analyzed the fMRI data acquired from three subjects during the in-scanner fixation test (see above). As our aim was solely to confirm our main fMRI results, and given the small group size, we used the variance between-scans rather than the variance between-subjects for statistical inference (i.e. in a fixed-effects rather than a random-effects analysis). The design matrix for this analysis thus included the 15 runs from the three subjects, where for each run we created regressors for valid \(g\) trials and valid \(rg\) trials, as well as separate a regressor for the flagged eye movement (\textit{invalid} and \textit{excursion}) trials. Additional regressors for run constants and head motion covariates were included in the design matrix as in the main fMRI analysis. Upon estimation of this model, a contrast image for (valid \(g\) trial – valid \(rg\) trials) was generated and assessed at \(P\)-uncorrected < 0.05.

Anatomical localization of activity peaks identified in the relevant contrasts was guided by several
widely available software tools, including the Automated Anatomical Labeling database (AAL, Tzourio-Mazoyer et al. 2002) available both for SPM2 and as part of the MRicro software (http://www.mricro.com), and the brain partitioning scheme available in Caret (Van Essen et al. 2001; Van Essen 2005; http://brainmap.wustl.edu/caret).

Anatomical localization of the activity peaks in the cerebellum was derived from the AAL segmentation scheme, which is based on the atlas by Schmahmann et al. (1999). Cerebellar nuclear activations were localized with two atlases published by Dimitrova et al. (2002, 2006), using the 61-70% probabilistic volume (the highest reported probability). Brainstem activations were localized with the atlases by Duvernoy (1995) and Schaltenbrandt and Wahren (1977). Volumetric analysis of the activated cerebellar regions (P<0.05, corrected for multiple comparisons) was performed by determining the number of activated voxels falling within each region of interest. This procedure is equivalent to computing a logical AND between each activation map and the segmentation scheme described above. We considered that the extent of the activation falling within each region was relevant when the activated fraction was > 10% of its volume and the peak $t$-value was > 3.375 (P<0.001, d.f.=31).
Results

Behavioral results: fMRI study

Effect of visual context on interception timing errors.

Figure 2 plots the response timing errors (RTEs) averaged across subject and flight duration for pictorial and non-pictorial visual context conditions. RTE represents a motor correlate of the target TTC\(t\) estimated by the subject: an RTE of 0 corresponds to a perfect estimate of \(TTC(t)\) and to ideal interception, while RTE<0 corresponds to underestimation of \(TTC(t)\) and RTE>0 corresponds to overestimation of \(TTC(t)\). RTEs were significantly influenced by both the law of motion (\(g\) or \(rg\)) and the visual context (pictorial or non-pictorial), and these two factors also interacted significantly (\(P<0.05\) for the individual factors, and for their interaction; between-groups repeated measures ANOVA on mean subject RTEs to \(g\) and \(rg\) trials, d.f. = 31).

The crucial finding was that subjects correctly estimated the \(TTC(t)\) of \(g\) targets but grossly underestimated that of \(rg\) targets in the pictorial condition: on average, they intercepted \(g\) targets accurately, but responded too early to \(rg\) targets. In the non-pictorial condition, instead, the \(TTC(t)\) of \(g\) targets was overestimated, whereas the \(TTC(t)\) of \(rg\) targets was underestimated. In the pictorial condition, the mean RTE for \(g\) trials was not significantly different from 0 (3 ± 15 ms, across subjects mean ± s.e.m., one-sample \(t\)-test, \(n=17\), \(P=0.84\)), while the mean RTE for \(rg\) trials was significantly negative by a substantial margin (-95 ± 23 ms, \(P<0.001\)). By contrast, in the non-pictorial condition, the mean RTE for \(g\) trials was significantly positive (36 ± 13 ms, \(n=16\), \(P=0.013\)), while the mean RTE for \(rg\) responses was negative (-32 ± 16 ms, \(P=0.067\)), but by a much smaller margin than that in the pictorial condition.

In each experiment, target flight duration (FD) was varied among 5 different possible values by randomizing initial target velocity (see Methods). In both visual context conditions, mean RTEs for \(g\)
trials were independent of FD (Fig. 3; linear regressions: pictorial, \( P=0.216, n=85 \); non-pictorial, \( P=0.730, n=80 \)), whereas mean RTEs for \( rg \) trials were increasingly negative with increasing FD (\( P<0.001 \) for both visual context conditions). Moreover, for each law of motion, the slopes of individual pictorial and non-pictorial linear regressions were not significantly different (homogeneity of slopes test for interaction between factors group and FD; \( g \) RTEs, \( P=0.49 \); \( rg \) RTEs, \( P=0.75 \)).

**Additional controls for time effects and group differences.**

We investigated the extent to which the group performance differences might be attributable to response trends across the experiment. Only modest changes of RTE with trial repetition were observed within a given condition, consistent with the lack of performance feedback. In both the pictorial and non-pictorial conditions there were small negative trends in the mean RTE for \( g \) trials across the 5 consecutive runs included in each experiment (linear regression of mean RTE with run, pictorial slope -9.4 ms/run, \( P=0.049 \); non-pictorial slope -8.6 ms/run, \( P=0.04 \)), while the RTE for \( rg \) trials did not change significantly across the experiment (\( P>0.44 \) for both visual context conditions).

In addition to the interception task, subjects also performed a reactive task (responding as soon as possible to a brief expansion of the fixation point in the absence of visual motion stimuli) as an independent control on group differences. Reactive trial RTEs for the two groups of subjects were not significantly different (pictorial, 286 ± 8 ms; non-pictorial, 307 ± 9 ms, unpaired \( t \)-test, d.f.=31, \( P=0.08 \)). In addition, reactive responses did not change significantly across the experiment for either group (linear regressions of mean RTE on run, pictorial \( P=0.6 \), non-pictorial \( P=0.95 \)).

**Behavioral results: eye movement tests**

In the 7-min pre-scan fixation tests performed with all subjects of the main fMRI study (see Methods), we found that all subjects maintained fixation well. The mean deviation from the fixation
point was 0 ± 0.7° (mean ± s.d.) on the horizontal and -0.2° ± 1.6° on the vertical, and the mean saccade rate (>1.5°-amplitude, >30 ms duration) was 0.04 ± 0.02 saccades per second.

An additional in-scanner fixation test was run with 3 subjects who each performed a full protocol of pictorial g and rg interceptions during fMRI acquisition (450 trials in 5 runs, 6.5 min per run). In offline analysis, 2.4% of all trials were flagged as invalid due to excessive missing data points or blinks and an additional 4.5% of trials were flagged as excursion trials where positive excursion from the fixation point during target motion exceeded 1° of visual angle for 150 ms. For all 3 subjects, the numbers of invalid trials and excursion trials were evenly split between g and rg trials (16 vs. 16, and 30 vs. 31, respectively). There was no evidence of a systematic increase in flagged trials of either type with progression of the experiment (P>0.5 for either type, linear regression analysis).

Behavioral results: control experiments performed in a seated posture

The presence of naturalistic cues (person and building, see Fig. 1a) in the pictorial scene might have aided the subjects lying supine in the scanner to mentally rotate the target path from the actual horizontal axis to an apparent vertical (visual gravity axis of the observed background image, Fig. 1c). Instead, the absence of such cues in the non-pictorial condition (Fig. 1b) might have led subjects to perceive the moving target as horizontal (i.e., non-gravitational) throughout the experiment. Therefore, we performed an additional series of purely behavioral experiments with two additional groups of subjects who were seated upright in front of a CRT monitor in the laboratory, so that each subject’s head and body axis, target motion axis, and Earth’s gravity vector were all roughly parallel (see Methods).

The pattern of interception responses for g and rg trials across pictorial and non-pictorial conditions for this seated protocol was similar to that for the fMRI (supine) protocol. In the pictorial condition, the mean RTE for g trials was not significantly different from 0 (-5 ± 14 ms, across subjects mean ±
s.e.m., one-sample $t$-test, $n=17$, $P=0.7$), while the mean RTE for $rg$ trials was significantly negative ($-65 \pm 20$ ms, $P<0.005$). By contrast, in the non-pictorial condition, the mean RTE for $g$ trials was significantly positive ($19 \pm 8$ ms, $P<0.05$), while the mean RTE for $rg$ responses was negative but not significantly different from 0 ($-21 \pm 12$ ms, $P=0.1$). The slopes of the linear regressions between RTEs and target flight duration for the seated protocol were not significantly different from the corresponding slopes computed for the supine protocol ($P>0.65$). Finally, the seated and supine protocols were not significantly different ($P>0.6$) in a direct statistical comparison of the data from the four groups of subjects.

Brain imaging results

By design, target motion was identical in the pictorial and the non-pictorial conditions: $g$ targets were affected by a natural gravity and $rg$ targets by reversed gravity. However, only the pictorial condition included the familiar cues sufficient to gauge the scale of the scene and to accurately calibrate the effects of natural gravity on target motion. Accordingly, we examined the BOLD brain responses for the main effect of gravitational motion independent of visual context, followed by the brain responses for the interaction effect between gravitational motion and pictorial context.

Brain responses selective for gravitational motion independent of visual context.

The main effect of gravitational motion identified the brain regions significantly more active in $g$ trials than in $rg$ trials, irrespective of whether the target was presented in the pictorial or non-pictorial context ($P<0.05$, corrected for multiple comparisons, Fig. 4, local maxima listed in Table 1). We found bilateral frontal, insular, temporo-parietal, and occipital regions, in addition to sub-cortical regions (mainly thalamus, putamen and lateral cerebellum). Frontal activations were predominantly on the left hemisphere and included motor, dorsal premotor and supplementary motor areas, and middle cingulate
regions. Bilateral insular activations encompassed both anterior and posterior regions. A set of regions was activated bilaterally along the temporo-parietal junction (TPJ), including the inferior parietal lobule, parietal operculum, temporal operculum, superior and middle temporal gyri. Peri-sylvian regions (insula and TPJ) are thought to represent the core regions of human vestibular cortex because they receive disynaptic input from the vestibular complex via the thalamus (Guldin and Grüsser 1998) and they can be activated independently by direct vestibular stimulation (caloric or galvanic, Bense et al. 2001; Bottini et al. 2001; Indovina et al. 2005). Bilateral occipital activations were located in the lingual gyrus, mainly in correspondence of ventral V1, V2, V3 (VP) and V4, plus V3A. Putamen was extensively activated bilaterally. Cerebellar activations occurred in hemispheric portions of bilateral lobule VI (extending into crus 1) and right lobule X, as determined by volumetric analysis (see Methods). Modulation of activity related to the direction of gravity in the pictorial group was comparable to that in the non-pictorial group. Figure 5 shows bar-graphs of the difference t-values for g and rg trials at selected sites; a similar pattern was observed at all sites of Table 1. The ensemble of areas more active in g trials than in rg trials roughly corresponds to that previously described by Indovina et al. (2005) for a similar contrast in a pictorial context only.

Also in agreement with the previous results of Indovina and colleagues (2005), we found that low-order motion-sensitive regions (area V2v, V3A, hMT/V5+ and regions along the intra-parietal sulcus) were activated to a comparable extent in g trials and rg trials across visual context conditions compared to the overall signal mean (analysis of the main effect of the interception task, i.e. summed contrast for g and rg interception trials across visual context conditions, $P<0.05$, corrected for multiple comparisons; see Table 2).

By contrast, significant activation for the main effect of reversed gravitational motion (rg - g, i.e. activation greater for rg trials than for g trials, or equivalently, relative deactivations for g trials, across visual context conditions) was confined to a single significant cluster in the right middle/inferior
occipital gyrus, with activation peak at \{36, -92 -4\} (P<0.05 corrected for multiple comparisons, Z-score 4.07, cluster size 470 voxels), in concordance with the similar result from our previous fMRI study (Indovina et al. 2005). This site has also previously been identified as a low-order motion-sensitive region (Orban et al. 2003).

**Brain responses selective for gravitational motion in the pictorial context.**

The interaction effects between gravitational motion and visual context revealed the brain regions exhibiting differential responses to \(g\) and \(rg\) trials that depended on visual context. Significantly greater activation for \(g\) than \(rg\) trials specifically for the *pictorial* rather than the *non-pictorial* condition was found in the midline cerebellum and brainstem (\(P<0.05\) corrected for multiple comparisons, Fig. 6, Table 3). Peak cerebellar activations were localized in the posterior vermis (Fig. 6a,c), corresponding to Larsell Lobules VI-X. The cerebellar activation also encompassed the fastigium bilaterally. The significant brainstem activation (white circle, Fig. 6b) was localized to the left superior vestibular nucleus. Note that while only the left vestibular nuclei reached full statistical significance after correction for multiple comparisons, the right vestibular nuclei also showed higher activity for \(g\) than \(rg\) trials specifically in the *pictorial* context (\(P=0.002\) uncorrected, yellow dashed circle, Fig. 6b).

By contrast, we did not identify any significantly greater activation for \(g\) trials relative to \(rg\) trials specifically for the *non-pictorial* rather than the *pictorial* condition (i.e. the complementary interaction effect).

**Brain responses correlated with target flight duration.**

We reported above that for both the *pictorial* and the *non-pictorial* conditions the mean RTEs for \(g\) trials were independent of target flight duration (FD), whereas the mean RTEs for \(rg\) trials were increasingly negative with increasing FD, giving an overall pattern of increase in the difference of \(g\)
and \( rg \) RTEs with increasing FD. We tested for similar FD-dependent changes in brain activity for \( g \) and \( rg \) interception trials across visual context conditions. No significant FD-dependent decrease of brain activation was found for \( g \) trials, and no significant increases were found in brain activation for either \( g \) or \( rg \) trials with increasing FD. In particular, target flight duration did not significantly affect activity within the cerebellar sites that showed a main effect of law of motion (see Table 1), or those for the law of motion by visual context interaction (see Table 3), even using a small volume correction procedure. We instead identified an anatomically distinct cerebellar region comprising the vermician and para-vermician lobules IV-V which showed an increase in differential \((g - rg)\) brain activation with increasing FD irrespective of pictorial or non-pictorial context (\( P<0.05 \) corrected for multiple comparisons, see Fig 7, Table 4). Additional post-hoc contrast analyses indicated that this activation was due to a selective decrease of activity for \( rg \) trials with increasing FD (\( P<0.05 \), both for pictorial and non-pictorial conditions).

**Analysis of time effects and group differences in the fMRI data.**

We did not identify any significant changes in whole-brain \((g - rg)\) differential activation at the group level across fMRI runs (see Methods). In particular, there were no significant positive or negative trends in \((g - rg)\) activation at the reported loci for either the main effect of law of motion across visual contexts or for interaction effects between the law of motion and visual context, even with small volume correction.

Finally, a 1-way group ANOVA did not reveal significant differences in brain activations between pictorial and non-pictorial groups for the fixation expansion (reactive) task. This result, together with the similar lack of significant difference in pictorial vs. non-pictorial RTEs for the reactive task (see above), support our conclusion that visual context, rather than mere group effects, selectively affected the interception task performance and brain activation.
Analysis of eye movement influence on midline cerebellar activations.

We analyzed the fMRI data acquired from three subjects during the in-scanner fixation test (see Methods) as a means of assessing whether the midline cerebellar activations identified in the interaction effect between law of motion and visual context ([pictorial - non-pictorial], (g - rg)) of the main fMRI study (Fig. 6) might be attributable to eye movements. Trials containing eye movements larger than 1° during target motion were excluded from this analysis, so that the (g - rg) contrast image was generated using only trials associated with good fixation. A cluster of voxels was identified in midline cerebellum with peak activation at coordinates {0, -58, -24}, corresponding to Lobules VII/VIII (near locus {2, -60, -26} reported above for the interaction contrast of the main fMRI study; search volume limited to the regions identified in the interaction contrast of the main fMRI study, P<0.05 uncorrected for multiple comparisons). These results support our main finding of a midline cerebellar response to gravitational motion combined with a pictorial context.
Discussion

The main finding of this study was that embedding target motion in a pictorial scene rather than a blank scene resulted in enhanced interception performance with $g$ targets and degraded performance with $rg$ targets, associated with enhanced BOLD response for $g$ relative to $rg$ targets localized to the vestibular nuclei and the posterior cerebellar vermis. Here we argue that the naturalistic context and perspective metric of a pictorial scene may contribute to map between retinal and world motion and to calibrate the effects of gravity on a visual target. We also argue that the spatial reference information extracted from a pictorial scene might be integrated with visual motion information in the vestibular nuclei and posterior cerebellar vermis, yielding a representation of physical motion consistent with internalized expectations of gravity effects.

Behavioral responses

A pictorial background affected the interception responses in opposite directions for $g$ targets and $rg$ targets. On average, $TTC(t)$ of $g$ targets was correctly estimated in the pictorial condition, but it was overestimated in the non-pictorial condition. By contrast, $TTC(t)$ of $rg$ targets was grossly underestimated in the pictorial condition, while it was still underestimated but to a much lesser extent in the non-pictorial condition. However, the response times for $g$ targets were independent of target flight duration whereas those for $rg$ targets were increasingly negative with increasing FD in both the pictorial and the non-pictorial conditions, indicating that the basic interception strategies employed with each set of targets were conserved across the two visual contexts. The pattern of behavioral results was the same whether subjects lay supine (in the fMRI experiments) or were seated (in the laboratory), indicating that the differences in interception performance between the pictorial and non-pictorial conditions were not dependent either on body orientation in space, or on unidentified group differences. Each subject’s head and trunk were roughly parallel to the visual scene on the monitor,
both when subjects lay supine in the scanner and when they sat in the laboratory. Therefore, we cannot discriminate between the role of allocentric cues related to the visual scene and that of idiotropic cues related to the trunk (Mittelstaedt 1986; Zupan and Merfeld 2005).

In line of principle, presupposed knowledge of $g$ may contribute to cue the apparent viewing distance from observed $g$-coherent motion irrespective of visual surround, allowing predictive estimates of $\text{TTC}(t)$ for gravitational targets both in the pictorial and the non-pictorial conditions. However, the finding that interception of $g$ targets was much less accurate in the absence than in the presence of a pictorial scene goes along with previous observations that viewing distance is estimated inaccurately from a gravitational fall in the absence of patterned background (Hecht et al. 1996; Saxberg 1987). Instead, the pictorial condition included a number of cues (the person standing at the interception point, the ball, linear perspective, shading, texture gradient) that could help to gauge the scale of the scene, and thereby aid calibration of predictive estimates of $\text{TTC}(t)$ for gravitational targets. The present findings extend to interception of visual gravitational motion the conclusion previously reached from studies on the perceptual judgment of $\text{TTC}(t)$ for looming stimuli. Pictorial cues such as familiar size, height in the field, and occlusion have been shown to affect $\text{TTC}(t)$ for such stimuli, especially when the spatio-temporal information is sub-threshold or the retinal and gaze-position signals are inaccurate (DeLucia 1991; 2004; 2005). Both the motor interception of $g$ targets and the perceptual judgment of $\text{TTC}(t)$ for looming stimuli may depend on mapping between retinal and world motion based on a reference metric provided by a naturalistic (or quasi-naturalistic) visual setting.

**Brain imaging**

$g$ trials preferentially engaged a cortical-subcortical set of areas comparably in both visual context conditions, and on top of that, a specific subcortical network only in the pictorial condition. The brain regions significantly more active in $g$ trials than in $rg$ trials, irrespective of visual context, included
bilateral frontal, insular, temporo-parietal, and occipital regions, in addition to thalamus, putamen and lateral cerebellum. This ensemble of areas is similar to that previously described by Indovina and colleagues (2005) for the contrast (\(g\) trials - \(rg\) trials) in a pictorial context. The previous study also showed that activation of these areas occurs irrespective of the specific motor task (interception of the descending target or reaction-time response following target descent at temporally uncoupled delays). Since interception performance (response timing errors) differed substantially between the pictorial and the non-pictorial conditions, the present findings confirm that the presence of natural gravity in the visual stimuli rather than motor output \textit{per se} is the major determining factor engaging this ensemble of areas. Peri-sylvian regions (including the insula and TPJ) are considered the core of the human vestibular cortex due to their receipt of disynaptic input from the vestibular complex via the thalamus (Guldin and Grüsser 1998) and the ability to activate these regions independently using caloric or galvanic vestibular stimulation (Bense et al. 2001; Bottini et al. 2001; Indovina et al. 2005). One can hypothesize that the human vestibular cortex encodes an internal model of gravity accessible by the visual system. The present data indicate that this internal model can be engaged irrespective of whether gravitational motion is presented in a pictorial or non-pictorial context.

The novel findings concern the activation of the midline cerebellum and brainstem in the interaction effect of (\(g - rg\)) specific to the pictorial rather than the non-pictorial condition. Cerebellar activations were localized in the posterior vermis (lobules VI-X) and extended bilaterally to the fastigium nuclei. Previous neuroimaging studies have underlined the role of the posterior cerebellar vermis in processing of vestibular (Bottini et al. 2001; Naito et al. 2003; Bense et al 2001; Stephan et al. 2005) as well as optic flow and optokinetic information (Bense et al. 2006; Kleinschmidt et al. 2002), but also in the discrimination of visual patterns (Dupont et al. 1998; Bundesen et al 2002). Accordingly, lesions of midline cerebellum impair visual motion direction discrimination (Nawrot and Rizzo 1995) and mimic symptoms of peripheral vestibulopathies (Lee et al 2003).
The cerebellar activation identified in the interaction effect between law of motion and visual context could not be accounted for either by temporal trends (e.g., due to learning, see Doyon et al. 2002), or by the behavioral pattern of motor timing errors. Instead, we found significant modulation of activity for target flight duration in anatomically distinct cerebellar regions (vermian and para-vermian lobules IV-V). Activity in these regions is known to relate somatotopically to arm and hand movements and appears to be involved in monitoring motor timing or the dynamic state of the limb (Diedrichsen et al. 2005; 2007; Grodd et al. 2001; Imamizu et al. 2000; Ivry and Spencer 2004; Milner et al. 2006, 2007; Penhune et al., 1998; Ramnani and Passingham, 2001; Xu et al. 2006).

We must also consider the potential role of eye movements in either the main effect of gravitational motion or the interaction effect. In this respect, it should be noted that subjects were instructed to maintain fixation during task execution, and we monitored their fixation ability prior to fMRI. We found no evidence for BOLD activation in (g - rg) contrasts in the cortical areas that are typically involved with saccadic or pursuit eye movements (such as FEF, SEF, intraparietal sulcus and transverse occipital sulcus regions, see Corbetta et al. 1998; Petit et al. 1997). On the other hand, the posterior vermis is known to be involved in oculomotor control, although at locations different from those identified in the present interaction effect (Bense et al. 2006; Burke and Barnes 2007; Corbetta et al. 1998; Nitschke et al. 2004). Here, the midline cerebellar activation in Lobules VII-VIII identified in the present interaction effect was also identified at a lower statistical threshold in the additional small group study (pictorial interception fMRI protocol with eye movement monitoring) even when trials flagged for the presence of eye movements were removed from the analysis.

The interaction effect of (g - rg) specific to the pictorial rather than the non-pictorial condition also showed bilateral activation of the vestibular nuclei (cluster-corrected on the left side, and uncorrected for multiple comparisons on the right side). Limitations in fMRI spatial resolution require that we consider alternative anatomical labeling of the activated brainstem loci. According to the Duvernoy
(1995) atlas, structures neighboring the vestibular nuclei include the superior olivary, trigeminal, abducens and facial nuclei, as well as a number of en passant nerves bundles (facial, trigeminal, vestibular, medial and lateral lemniscii; inferior, middle and superior cerebellar peduncles). This list of neighboring structures does not appear to include BOLD-responsive areas of functional relevance for the present experiment involving manual interception of vertically accelerating motion. Thus, we retain the localization to the vestibular nuclei as the most likely for our study.

The observation that the vestibular nuclei and posterior cerebellar vermis were engaged by the high-level visual analysis in the pictorial condition may appear surprising, given that both short-range and long-range context effects are generally documented in cerebral regions (Albright and Stoner 2002; Bar 2004; Goh et al. 2004; Murray et al. 2006). We speculate that initial scene analysis and extraction of a scale factor from pictorial cues was performed elsewhere in the brain, perhaps in the parahippocampal and retrosplenial cortices which are known to be involved in contextual analysis of static scenes (see Bar 2004 for a review). However, activation of these cerebral areas probably adapted in the very first trials following continued presentation of the same picture, and did not emerge in our whole brain analysis. Instead, activation of the vestibular nuclei and posterior cerebellar vermis throughout the experiment may reflect the integration of the spatial reference information extracted from the scene with on-line visual motion information, integration which was presumably necessary on every trial since visual motion information changed randomly (with flight duration) from trial to trial. As a result of such integration, retinal motion would be mapped into world motion, yielding a representation useful for the internal model of gravity. Perceived size of a scene might act on the retinal coordinates of a target via gain modulation, simulating the effect of changes in viewing distance (Angelaki 2004; Dobbins et al 1998; Murray et al. 2006).

Current evidence on the visuo-vestibular interactions occurring in the vestibular nuclei and cerebellum, mainly based on anatomical and electrophysiological data obtained in the monkey,
suggests that these structures may subserve the computations necessary to scale visual motion coherent with gravity (Angelaki and Hess 2005). In monkeys, the caudal vermis (including nodulus/uvula) receives primary and secondary vestibular afferents, as well as visual signals from the accessory optic system and extrastriate cortical areas (including the MT/MST/FST complex) via pontine nuclei (Glickstein et al. 1994; Voogd et al. 1996), and encodes visual motion signals (Krauzlis and Lisberger 1991). The nodulus and uvula in mammals are critical sites for visuo-vestibular interaction (Precht et al. 1976), receiving this information via climbing-fiber pathways (Takeda and Maekawa 1989). Moreover, during body motion, vestibular nuclear neurons vary their firing rates with viewing distance (Chen-Huang and McCrea 1999; see Angelaki 2004), and neural populations in vestibular nuclei and midline cerebellum compute a vestibular estimate of gravity combining signals from otoliths and semicircular canals (Angelaki et al. 2004). Vestibular-only neurons in the vestibular nuclei and cerebellar nodulus/uvula are thought to be involved in transforming head-referenced movement information into a gravity-referenced coordinate frame (Angelaki and Hess 1995; Cullen and Roy 2004). Further transformations into a scene-referenced frame might render the information directly accessible to visual representations.

Conclusions

We showed that humans can employ pictorial cues to improve performance during interception of observed gravitational targets. Our results add to a growing body of evidence that visual context can influence motion processing as well as the perception of static objects (Albright and Stoner 2002; Bar 2004; DeLucia 1991; Distler et al 2000; Gilroy and Blake 2004; McKee and Smallman 1998). Indeed, our present and previous results indicate that contextual cues in the visual, proprioceptive and haptic modalities can be as important as the motion stimuli themselves in determining motor response strategies during interception (McIntyre et al. 2001; Senot et al. 2005; Zago et al. 2004).
It has previously been hypothesized that the internal vestibular estimate of gravity initially derived in head coordinates (Hess and Angelaki 1999; Merfeld et al. 1999; Zupan et al. 2002) is further transformed into an abstract representation of gravity in visual coordinates to be combined with on-line visual estimates of target motion (Zago et al. 2004; Day and Fitzpatrick 2005; Indovina et al. 2005). This visual estimate of gravity would be processed according to internalized Newton’s laws in the vestibular cortex (Indovina et al. 2005). The present results suggest that cortical vestibular analysis of gravitational motion likely occurs independent of visual context. However, accurate scaling of gravitational motion in world coordinates appears to be contingent on the availability of a reference metric as provided by a pictorial scene. Here we used two-dimensional stimuli in both the *pictorial* and *non-pictorial* conditions. It remains to be seen how the results generalize to three-dimensional scenes, as rendered by means of stereoscopic stimuli (Senot et al. 2005). Our data further indicate that integration of a scene-derived spatial reference with visual motion information occurs in the vestibular nuclei and posterior cerebellar vermis, regions which are reciprocally connected with the vestibular cortex (Fukushima 1997; Guldin and Grüsser 1998; Cullen and Roy 2004). Thus visual gravity and the mechanisms for engaging this internal knowledge during interception are presumably represented in a distributed cortical-subcortical network. The described mechanisms may provide a substrate for perceptual constancy of ubiquitous gravitational motion.
Acknowledgments

We thank Dr. I. Indovina for advice during the design of the experiments and data analysis, and Drs. A. d’Avella, R. Frackowiak, Y.P. Ivanenko, G. Rees, L.C. Robertson and the anonymous reviewers for comments and advice on earlier versions of this article. Current address of M. Iosa: Department of Human Movement and Sport Sciences, Istituto Universitario di Scienze Motorie, Piazza Lauro de Bosis, 6, 00194 Rome, Italy.

Grants

This research was supported by grants from the Italian Ministry of Health (RC and RF ISPESL), the Italian Ministry of University and Research (PRIN grant), and the Italian Space Agency (DCMC grant).

Disclosures

None.
References


Table 1. Brain regions significantly more active in $g$ trials than in $rg$ trials across pictorial and non-pictorial conditions.

<table>
<thead>
<tr>
<th>Brain Area</th>
<th>Side</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cg</td>
<td>L</td>
<td>-4</td>
<td>0</td>
<td>48</td>
<td>3.25</td>
</tr>
<tr>
<td>MeFg (SMA)</td>
<td>L</td>
<td>-6</td>
<td>-2</td>
<td>60</td>
<td>3.46</td>
</tr>
<tr>
<td>MeFg (SMA)</td>
<td>R</td>
<td>6</td>
<td>2</td>
<td>60</td>
<td>3.68</td>
</tr>
<tr>
<td>MdFg (PM)</td>
<td>L</td>
<td>-24</td>
<td>-16</td>
<td>64</td>
<td>3.47</td>
</tr>
<tr>
<td>PrCg (PM)</td>
<td>L</td>
<td>-38</td>
<td>-10</td>
<td>66</td>
<td>4.50</td>
</tr>
<tr>
<td>CeS (M1)</td>
<td>L</td>
<td>-32</td>
<td>-26</td>
<td>56</td>
<td>3.49</td>
</tr>
<tr>
<td>PoCg</td>
<td>L</td>
<td>-44</td>
<td>-20</td>
<td>56</td>
<td>3.98</td>
</tr>
<tr>
<td>Insula</td>
<td>L</td>
<td>-34</td>
<td>-4</td>
<td>-12</td>
<td>4.73</td>
</tr>
<tr>
<td>Insula</td>
<td>R</td>
<td>34</td>
<td>10</td>
<td>-16</td>
<td>4.49</td>
</tr>
<tr>
<td>SMg</td>
<td>L</td>
<td>-52</td>
<td>-40</td>
<td>26</td>
<td>3.63</td>
</tr>
<tr>
<td>STg</td>
<td>L</td>
<td>-50</td>
<td>6</td>
<td>-14</td>
<td>3.51</td>
</tr>
<tr>
<td>STg</td>
<td>L</td>
<td>-62</td>
<td>-42</td>
<td>14</td>
<td>3.96</td>
</tr>
<tr>
<td>STg/SMg</td>
<td>R</td>
<td>64</td>
<td>-38</td>
<td>16</td>
<td>3.77</td>
</tr>
<tr>
<td>MTg/STS</td>
<td>L</td>
<td>-52</td>
<td>-52</td>
<td>6</td>
<td>3.26</td>
</tr>
<tr>
<td>MTg</td>
<td>R</td>
<td>52</td>
<td>-44</td>
<td>4</td>
<td>3.27</td>
</tr>
<tr>
<td>STg / MTg</td>
<td>R</td>
<td>48</td>
<td>6</td>
<td>-26</td>
<td>3.43</td>
</tr>
<tr>
<td>MTg</td>
<td>L</td>
<td>-52</td>
<td>-64</td>
<td>6</td>
<td>3.85</td>
</tr>
<tr>
<td>MTg</td>
<td>R</td>
<td>46</td>
<td>-62</td>
<td>8</td>
<td>3.51</td>
</tr>
<tr>
<td>Lg (V1/V2)</td>
<td>L</td>
<td>-2</td>
<td>-92</td>
<td>-4</td>
<td>4.31</td>
</tr>
<tr>
<td>Lg (V1)</td>
<td>R</td>
<td>18</td>
<td>-78</td>
<td>2</td>
<td>4.24</td>
</tr>
<tr>
<td>Lg (V2)</td>
<td>R</td>
<td>4</td>
<td>-80</td>
<td>-4</td>
<td>5.37</td>
</tr>
<tr>
<td>Lg (V2/V3)</td>
<td>L</td>
<td>-10</td>
<td>-82</td>
<td>-8</td>
<td>5.99</td>
</tr>
<tr>
<td>Lg (V3/V4v)</td>
<td>L</td>
<td>-24</td>
<td>-72</td>
<td>-10</td>
<td>5.56</td>
</tr>
<tr>
<td>Lg (V3a)</td>
<td>L</td>
<td>-20</td>
<td>-90</td>
<td>18</td>
<td>3.38</td>
</tr>
<tr>
<td>Lg (V3/V4)</td>
<td>R</td>
<td>18</td>
<td>-76</td>
<td>-14</td>
<td>4.65</td>
</tr>
<tr>
<td>Thalamus</td>
<td>L</td>
<td>-10</td>
<td>-20</td>
<td>0</td>
<td>3.43</td>
</tr>
<tr>
<td>Thalamus</td>
<td>R</td>
<td>12</td>
<td>-22</td>
<td>-2</td>
<td>4.67</td>
</tr>
<tr>
<td>Putamen</td>
<td>L</td>
<td>-26</td>
<td>6</td>
<td>-4</td>
<td>4.85</td>
</tr>
<tr>
<td>Putamen</td>
<td>R</td>
<td>22</td>
<td>8</td>
<td>-6</td>
<td>4.68</td>
</tr>
<tr>
<td>Lat. Cereb. (Lobule VI)</td>
<td>L</td>
<td>-24</td>
<td>-60</td>
<td>-26</td>
<td>3.88</td>
</tr>
<tr>
<td>Lat. Cereb. (Lobule VI)</td>
<td>R</td>
<td>24</td>
<td>-54</td>
<td>-28</td>
<td>4.22</td>
</tr>
<tr>
<td>Lat. Cereb. (Lobule VI)</td>
<td>R</td>
<td>30</td>
<td>-48</td>
<td>-36</td>
<td>4.02</td>
</tr>
<tr>
<td>Lat. Cereb. (Lobule IX-X)</td>
<td>R</td>
<td>14</td>
<td>-52</td>
<td>-34</td>
<td>2.94</td>
</tr>
</tbody>
</table>
Table 1 Legend. Local maxima for the main effect of gravitational motion irrespective of pictorial context. $x,y,z$ coordinates (in mm) refer to locations in the normalized MNI reference frame. Z-scores were significant at $P<0.05$, corrected for multiple comparisons at cluster level. Abbreviations: Cg, Cingulate gyrus; CeS, Central Sulcus; Lat. Cereb., Lateral Cerebellum; Lg, Lingual gyrus; M1, Primary motor area; MdFg, Middle Frontal gyrus; MeFg, Medial Frontal gyrus; MTg, Middle Temporal gyrus; PM, Pre-Motor area; PoCg, Post-Central gyrus; PrCg, Pre-Central gyrus; SMA, Supplementary Motor Area; SMg, Supra-marginal gyrus; STg, Superior Temporal gyrus; STS, Superior Temporal Sulcus.
Table 2. Selection of motion-sensitive cortical areas activated in g and rg trials across pictorial and non-pictorial conditions

<table>
<thead>
<tr>
<th>Brain Area</th>
<th>Side</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MdOcc/MT</td>
<td>L</td>
<td>-48</td>
<td>-72</td>
<td>6</td>
<td>6.99</td>
</tr>
<tr>
<td>hMT+</td>
<td>L</td>
<td>-34</td>
<td>-70</td>
<td>4</td>
<td>5.23</td>
</tr>
<tr>
<td>V3v/V4v</td>
<td>L</td>
<td>-24</td>
<td>-74</td>
<td>-8</td>
<td>6.13</td>
</tr>
<tr>
<td>V3A</td>
<td>L</td>
<td>-22</td>
<td>-86</td>
<td>34</td>
<td>3.39</td>
</tr>
<tr>
<td>DIPSA</td>
<td>L</td>
<td>-36</td>
<td>-50</td>
<td>64</td>
<td>5.86</td>
</tr>
<tr>
<td>DIPSM</td>
<td>L</td>
<td>-30</td>
<td>-60</td>
<td>58</td>
<td>3.67</td>
</tr>
<tr>
<td>POIPS</td>
<td>L</td>
<td>-22</td>
<td>-84</td>
<td>38</td>
<td>2.67</td>
</tr>
<tr>
<td>VIPS</td>
<td>L</td>
<td>-24</td>
<td>-80</td>
<td>24</td>
<td>3.15</td>
</tr>
<tr>
<td>hMT+</td>
<td>R</td>
<td>54</td>
<td>-70</td>
<td>0</td>
<td>6.81</td>
</tr>
<tr>
<td>V2v</td>
<td>R</td>
<td>6</td>
<td>-86</td>
<td>-12</td>
<td>4.05</td>
</tr>
<tr>
<td>V4v</td>
<td>R</td>
<td>22</td>
<td>-68</td>
<td>-10</td>
<td>4.11</td>
</tr>
<tr>
<td>DIPSA</td>
<td>R</td>
<td>36</td>
<td>-46</td>
<td>60</td>
<td>2.67</td>
</tr>
<tr>
<td>DIPSM</td>
<td>R</td>
<td>24</td>
<td>-60</td>
<td>58</td>
<td>2.86</td>
</tr>
<tr>
<td>VIPS</td>
<td>R</td>
<td>26</td>
<td>-80</td>
<td>24</td>
<td>2.73</td>
</tr>
</tbody>
</table>

Table 2. Legend. Local maxima for selected brain areas activated during execution of the interception task across visual context conditions. x,y,z coordinates (in mm) refer to locations in the normalized MNI reference frame. Z-scores were significant at P<0.05, corrected for multiple comparisons at cluster level. Abbreviations: MdOcc, Middle Occipital Lobe; MT, area MT; hMT+, human motion area MT; V2v, V3v, V4v, visual motion areas V2, V3, V4; V3A, visual motion area V3A; DIPSA, dorsal intraparietal sulcus anterior; DIPSM, dorsal intraparietal sulcus medial; POIPS, parieto-occipital intraparietal sulcus; VIPS, ventral intraparietal sulcus. Functional labeling of selected loci in visual areas accomplished using Caret flatmap software and AAL labeling toolbox. Selected loci in the interparietal sulci are within 8mm (1 FWHM smoothing kernel in the present analysis) of corresponding loci listed in Table 2 of Orban et al. 2006.
Table 3. Brain regions significantly more active in $g$ trials than in $rg$ trials specifically for the *pictorial* rather than the *non-pictorial* condition.

<table>
<thead>
<tr>
<th>Brain Area</th>
<th>Side</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereb. (Lobule VI)</td>
<td>L</td>
<td>-14</td>
<td>-42</td>
<td>-28</td>
<td>3.17</td>
</tr>
<tr>
<td>Cereb. Vermis (Lobule VII/VIII)</td>
<td>-</td>
<td>2</td>
<td>-60</td>
<td>-26</td>
<td>3.64</td>
</tr>
<tr>
<td>Cereb. Vermis (Lobule IX/X)</td>
<td>-</td>
<td>-6</td>
<td>-50</td>
<td>-28</td>
<td>3.47</td>
</tr>
<tr>
<td>Pons (vestibular nuclei)</td>
<td>L</td>
<td>-12</td>
<td>-36</td>
<td>-32</td>
<td>3.52</td>
</tr>
</tbody>
</table>

Table 3 Legend. Local maxima for the interaction between gravitational motion and pictorial context. Z-scores were significant at $P<0.05$, corrected for multiple comparisons at cluster level.
Table 4. Brain regions significantly more active in g trials than in rg trials specifically for longer flight durations than for shorter ones.

<table>
<thead>
<tr>
<th>Brain Area</th>
<th>Side</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereb. Vermis (Lobule IV-V)</td>
<td>-</td>
<td>-2</td>
<td>-52</td>
<td>-10</td>
<td>4.04</td>
</tr>
<tr>
<td>Lat. Cereb. (Lobule IV-V)</td>
<td>L</td>
<td>-14</td>
<td>-42</td>
<td>-14</td>
<td>3.85</td>
</tr>
<tr>
<td>Lat. Cereb. (Lobule IV-V)</td>
<td>R</td>
<td>10</td>
<td>-44</td>
<td>-12</td>
<td>3.19</td>
</tr>
</tbody>
</table>

Table 4 Legend. Local maxima for the interaction between gravitational motion and flight duration. Z-scores were significant at $P<0.05$, corrected for multiple comparisons at cluster level.
Figure Legends

Figure 1. Schematic representation of visual stimuli. **a, b.** Pictorial stimulus and non-pictorial stimulus, respectively. During interception trials, targets launched vertically from inside the container held by the human figure in **a,** and the lower box in **b,** rebounded at the trajectory apex (building cornice in **a,** upper box in **b**), and returned to the starting point (motion direction arrows are shown for illustrative purposes only). Subjects were instructed to maintain fixation on the central red dot. During reactive trials, no moving target appeared, but the fixation point expanded after a randomized delay. **c.** In the fMRI experiment, back-projected visual stimuli were viewed through a 45° inclined mirror attached to the MR head coil so that the visual downward direction (Z\text{vis}) was towards the subject’s feet (i.e. aligned with the body axis), and orthogonal to the Earth’s gravity vector (Z\text{g}).

Figure 2. Effect of visual context on interception of accelerated targets. Group mean (± between-subjects s.e.m.) response timing errors (RTEs) averaged across flight durations for pictorial and non-pictorial visual context conditions, g (white) and rg (black) trials. Negative (positive) time values correspond to responses occurring before (after) the target arrival time.

Figure 3. Interception errors across target flight duration. Group mean (± between-subjects s.e.m.) response timing errors for each flight duration (FD) for pictorial (red) and non-pictorial (blue) visual context conditions, g (empty) and rg (filled) trials. Dashed lines are least-squares linear regressions of the data for each experimental condition using individual subject mean values (see text).
Figure 4. Main effect of gravitational motion. Brain areas showing significantly greater BOLD response to g motion across visual context conditions. Color-coded areas of activation surviving correction for multiple comparisons (P<0.05, see Methods; dark-light color range corresponds to t-value range of 2.74-8.32). Data are presented in normalized stereotactic space, overlaid on a high-resolution anatomical MR image (CH2, MNI) with the left side of the brain shown on the left (numerical labels, z-coordinates of axial slices). Representative peak activations are listed in Table 1, a selection of which is indicated with circles and labels on the figure.

Figure 5. Bar-graphs of the mean difference (± between-subjects s.e.m.) t-values for g and rg trials are plotted for the pictorial (white) and non-pictorial (black) contexts for selected activation sites identified in the main effect of gravitational motion (refer to Fig. 4 and see text).

Figure 6. Interaction effect between gravitational motion and visual context. Brain areas showing preferential activation for g motion specifically for the pictorial visual context. a, b, c. Activations in the midline cerebellum (a, axial section; c, medial sagittal section) and vestibular nuclei (b, axial section), overlaid on the same MNI anatomical template as in Fig. 4 at an enlarged scale (scale bars 20 mm, dark-light color range: t-value range 2.74 - 4.10). In the axial sections, the left side of the brain is shown on the left. Roman numerals in c denote Larsell lobules. White circles are centered on maximal statistical activation peaks of the cluster surviving correction for multiple comparisons (P<0.05, see Methods, and Table 3 for Z-scores): a, lobules IX/X (x=-6, y=-50, z=-28 mm); b, left vestibular nuclei (x=-12, y=-36, z=-32 mm); c, lobules VII/VIII (x=2, y=-60, z=-26 mm). Plotted data are uncorrected for multiple comparisons (voxel threshold P<0.005) in order to show bilateral response of vestibular nuclei in b (right vestibular nuclei within the yellow dashed circle, x=16, y=-36, z=-32 mm, Z-score=2.87, voxel level P=0.002, uncorrected for multiple comparisons). d. Bar-graphs of the
difference (± between-subjects s.e.m.) $t$-values for $g$ and $rg$ trials in pictorial (white) and non-pictorial (black) context for the activity peaks circled in a, b and c.

Figure 7. Interaction effect between gravitational motion and target flight duration. Brain areas showing a significantly more reduction in activation with increasing target flight duration during $\sim g$ trials than $g$ trials across both pictorial and non-pictorial visual context conditions. a. Cerebellar activations overlaid on the same MNI anatomical template as in Fig. 4 (left, medial sagittal, middle, coronal section, right, axial section). Activation cluster for the interaction between law of motion and target flight duration surviving correction for multiple comparisons ($P<0.05$) is localized to lobules IV/V (refer to Table 4 for activation peaks coordinates and Z-scores). b. Bar-graphs of difference $t$-values for $g$ and $rg$ trials at the indicated activation peak.
Figure 1
Figure 2

![Bar chart showing RTE (ms) for pictorial and non-pictorial conditions. The chart includes error bars indicating variability. The legend includes symbols for 'g' and 'rg'.]
Figure 3
Figure 4
Figure 5

Bar charts showing T-values for different anatomical regions with and without pictorial or non-pictorial information.

Regions include:
- L MeFg (SMA)
- L SMg
- R STg/SMg
- L Lg (V3/V4v)
- L Insula
- R Insula
- L Lat. Cereb. (Lobule VI)
- R Lat. Cereb. (Lobule VI)
Figure 6
Figure 7

(a) Images showing the location of the Vermis, Lob. IV/V with coordinates x = -2, y = -52, z = -10.

(b) Graph showing the T-value for flight duration (ms) with different conditions:
- Pictorial
- Non-pictorial

The graph displays the mean values and standard errors for flight duration at different values, indicating a comparison between pictorial and non-pictorial conditions.