Exp Brain Res
DOI 10.1007/s00221-006-0792-0

RESEARCH ARTICLE

A Bayesian model of the disambiguation of gravitoinertial force by visual cues

Paul R. MacNeilage · Martin S. Banks · Daniel R. Berger · Heinrich H. Bülthoff

Received: 7 September 2005 / Accepted: 31 October 2006

Abstract The otoliths are stimulated in the same fashion by gravitational and inertial forces, so otolith signals are ambiguous indicators of self-orientation. The ambiguity can be resolved with added visual information indicating orientation and acceleration with respect to the earth. Here we present a Bayesian model of the statistically optimal combination of noisy vestibular and visual signals. Likelihoods associated with sensory measurements are represented in an orientation/acceleration space. The likelihood function associated with the otolith signal illustrates the ambiguity; there is no unique solution for self-orientation or acceleration. Likelihood functions associated with other sensory signals can resolve this ambiguity. In addition, we propose two priors, each acting on a dimension in the orientation/acceleration space: the idiotropic prior and the no-acceleration prior. We conducted experiments using a motion platform and attached visual display to examine the influence of visual signals on the interpretation of the otolith signal. Subjects made pitch and acceleration judgments as the vestibular and visual signals were manipulated independently. Predictions of the model were confirmed: (1) visual signals affected the interpretation of the otolith signal, (2) less variable signals had more influence on perceived orientation and acceleration than more variable ones, and (3) combined estimates were more precise than single-cue estimates. We also show that the model can explain some well-known phenomena including the perception of upright in zero gravity, the Aubert effect, and the somatogravic illusion.

Keywords Self-motion · Body orientation · Bayesian estimation · Gravitoinertial force · Optic flow · Acceleration · Vestibular system

P. R. MacNeilage (✉) · M. S. Banks
Vision Science Program, University of California, Berkeley, CA 94720, USA
e-mail: pogen@berkeley.edu

M. S. Banks
Department of Psychology and Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA

D. R. Berger · H. H. Bülthoff
Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany

Introduction

Knowing the body's orientation in world coordinates is essential for a range of behaviors including guidance of self-motion and vehicle operation. One way that humans and other animals estimate body orientation is from the direction of gravity. Because gravity's direction is constant in world coordinates, knowing the direction in body coordinates is sufficient for knowing body orientation in world coordinates. However, gravitational force and inertial force from self-acceleration cannot be measured independently because of the equivalence principle (Einstein 1907): Only the resultant force—the gravitoinertial force (Fig. 1a, blue vector)—can be measured. Said another way, during self-acceleration other forces act on the body's gravitational sensors, or ''graviceptors'' (the otoliths, and kinesthetic and somatosensory signals), so they cannot signal orientation unambiguously.

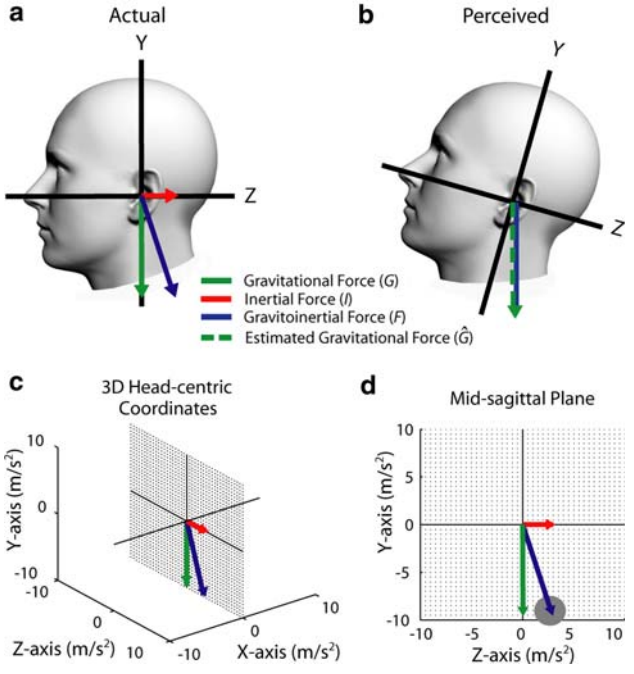

Fig. 1 Somatogravic illusion and gravitoinertial force. **a** The actual situation. The head is upright and accelerating in the forward direction, so the gravitoinertial force (*blue*), which is the sum of gravitational (*green*) and inertial (*red*) forces (Eq. 2), is pitched relative to the body. **b** The perceived situation. The person interprets the rotated gravitoinertial force vector as a pitch of the head, which leads to the perception of a ''nose-up'' attitude. **c** 3d force vectors. The head-centric axes *X*, *Y*, and *Z* represent lateral, vertical, and forward, respectively. The gravitational and inertial vectors define a plane. The orientation of the plane in head-centric coordinates depends on the direction of the gravitational and inertial forces, but the resultant gravitoinertial force vector will always lie in that plane. **d** Force vectors in their common plane. The *gray disk* represents the noise associated with the resultant vector. In the literature, the relevant vectors are sometimes described as accelerations and sometimes as apparent forces. We will adopt the convention of plotting the apparent force vectors in units of m/s^2. For example, the apparent force due to gravity for an upright head will be plotted down on the head and the force due to forward acceleration as back on the head

Despite this ambiguity, we are able to estimate our orientation accurately in most situations using other sources of sensory information and prior experience. Some of the sensory information is visual. For example, visible environmental features with a fixed orientation in world coordinates, such as the horizon, can be used to estimate the direction of gravity directly (Howard 1982). Likewise, visual cues to self-motion—optic flow (Gibson 1950, 1966)—could in principle be used to estimate self-acceleration from which the inertial component and thus the gravitational component of the gravitoinertial force can be estimated. Some of the potentially disambiguating sensory information is non-visual. An estimate of self-rotation from the semi-circular canals could be used to update estimates of gravitational force (Angelaki et al. 2004; Angelaki et al. 1999; Merfeld et al. 1999). Additionally, during active locomotion, knowledge of executed motor movements and the usual consequences can in principle be used to generate an estimate of self-acceleration and thereby the inertial component of gravitoinertial force (Mittelstaedt and Mittelstaedt 2001).

However, misperceptions of self-orientation and self-motion caused by an inability to disambiguate gravitational and inertial forces often occur, especially during vehicle operation. A significant example is the somatogravic illusion (Gillingham and Previc 1993; Glasauer 1995), which we will describe with respect to aviation (Fig. 1a, b). The plane accelerates forward in an earth-horizontal direction producing an inertial force in the opposite direction. The direction of the resultant gravitoinertial force is in-between the directions of the gravitational and inertial forces (Fig. 1a).[1] The pilot often interprets the gravitoinertial force to be largely due to gravitational force (Fig. 1b), so he/she perceives an erroneous ''nose-up'' pitch. If the pilot corrects the plane's attitude based on the erroneous percept, he/she will pitch the nose down, which causes more acceleration and more illusory pitch. Understandably, the somatogravic illusion has caused many fatal crashes (Gillingham and Previc 1993; Cheung et al. 1995; Ercoline 1997). The illusion occurs most frequently when the pilot cannot see the ground: for example, at night or flying in a cloud. This suggests that disambiguating visual information generally overrides the somatogravic illusion.

Here we investigate the principles that govern the disambiguation of gravitoinertial force. We first describe a Bayesian model for using sensory data from the vestibular and visual systems along with knowledge from previous experience. We then describe an experiment in which we measured the influence of visual signals on the interpretation of gravitoinertial force. The experiment reveals clear influences of visual signals. We then show how those effects and others can be understood in the framework of the Bayesian model.

[1] In the vestibular literature, it is conventional for the *X*, *Y*, and *Z* axes to correspond to forward, upward, and lateral, respectively. We have chosen to use axes that are conventional in the vision literature.

Bayesian model of visual-vestibular integration

In order to understand how self-orientation is estimated, we need to consider the available sensory information and the influence of previous experience. Useful sensory information includes signals from the otoliths and semi-circular canals, somatosensory and kinesthetic signals, and signals from the visual system (Zupan et al. 2002). In addition, there is useful prior information such as the fact that the head and body are usually upright in earth coordinates and large inertial accelerations are uncommon. All of the sensory and prior information is to some degree uncertain, so generating estimates of orientation relative to the world requires probabilistic inference. Bayes' Law prescribes the method by which the inference should proceed (Yuille and Bülthoff 1996; Kersten et al. 2004). It states that:

$$p(O|S) \propto p(S|O)p(O) \tag{1}$$

where O is orientation and S is the sensory input. The first term on the right side of the equation is the likelihood function characterizing the probability of observing the sensory input if O is the actual orientation. The second term is the prior distribution, which is the probability of observing O independent of the sensory data. The left side of the equation is the posterior probability distribution and the observer should use it to make his/her perceptual judgment. If there are no immediate consequences to the judgment (i.e., no penalties or rewards), the maximum a posteriori estimate is typically employed. That is, the observer chooses an orientation estimate that is most probable given the vestibular, kinesthetic/somatosensory, and visual signals, and prior expectations. Thus, we choose the value of O that maximizes $p(O|S)$. There is an increasing body of evidence that humans combine signals within and across the senses in a fashion that closely approximates Bayes' Law. For example, some studies measured the variances associated with two sensory signals and then empirically tested predictions for the appearance of a two-cue stimulus as well as two-cue discrimination thresholds (Alais and Burr 2004; Ernst and Banks 2002; Gepshtein and Banks 2003; Knill and Saunders 2003; Landy and Kojima 2001). The two-cue appearance and discrimination thresholds were quite close to those predicted by Bayes' Law. These studies did not reveal an influence of priors because the experiments were designed such that priors would have essentially no influence.

There are several specific sensory signals and prior expectations that are relevant to estimating orientation; each has different properties, so we consider them individually. Figure 2 depicts the model's components.

We are also interested in estimating the self-motion of an earth-bound observer. We focus here on the linear acceleration component of self-motion. For this estimation problem, we need to make a distinction between inertial acceleration, which involves translation relative to earth, and gravitational acceleration, which does not. The self-motion estimation problem is to estimate the inertial acceleration.

Otolith signals

The otoliths signal forces due to linear acceleration. When the head is subject to gravitational and inertial forces, the resulting gravitoinertial force ($\boldsymbol{F}$) is:

$$\boldsymbol{F} = \boldsymbol{G} + \boldsymbol{I} \tag{2}$$

where $\boldsymbol{G}$ and $\boldsymbol{I}$ are the vectors associated with gravitational and inertial force, respectively. These force vectors are schematized in Fig. 1c. Notice that any two gravitational and inertial force vectors and the resulting gravitoinertial force vector specify a plane in 3-dimensional head-centric coordinates. Henceforth we focus on the representation of the vectors in that plane. For the situation in Fig. 1a, the plane is sagittal through the middle of the head.

The blue vector in Fig. 1d shows an instance of the force vector sensed by the otoliths. For the purposes of explanation, we assume that the noise associated with the measurement is Gaussian and isotropic (represented by the gray disk).[2] Naturally, other assumptions could be made concerning the noise distribution depending on the sensitivity of the otoliths to different directions and magnitudes of gravitoinertial force. For the problem investigated in this paper—determining self-orientation and -acceleration relative to the earth—we need to convert the otolith likelihood into the relevant units of pitch and inertial acceleration. Pitch is the direction of gravitational force in head-coordinates; it is measured in degrees and written $\angle\boldsymbol{G}$. Inertial acceleration is the acceleration of the head relative to the earth; it is a vector with a magnitude and direction. In this paper, we are primarily concerned with the magnitude, which is measured in m/s^2 and written $\|\boldsymbol{I}\|$. This is because subjects in our experiment were asked to judge acceleration magnitude. One can show that

[2] The noise would actually be 3d in X–Y–Z space, but here we consider it collapsed onto the Z–Y plane. Note that Eq. 3 is valid for any head-centric plane, but for simplicity we discuss it in terms of the Z–Y plane only.

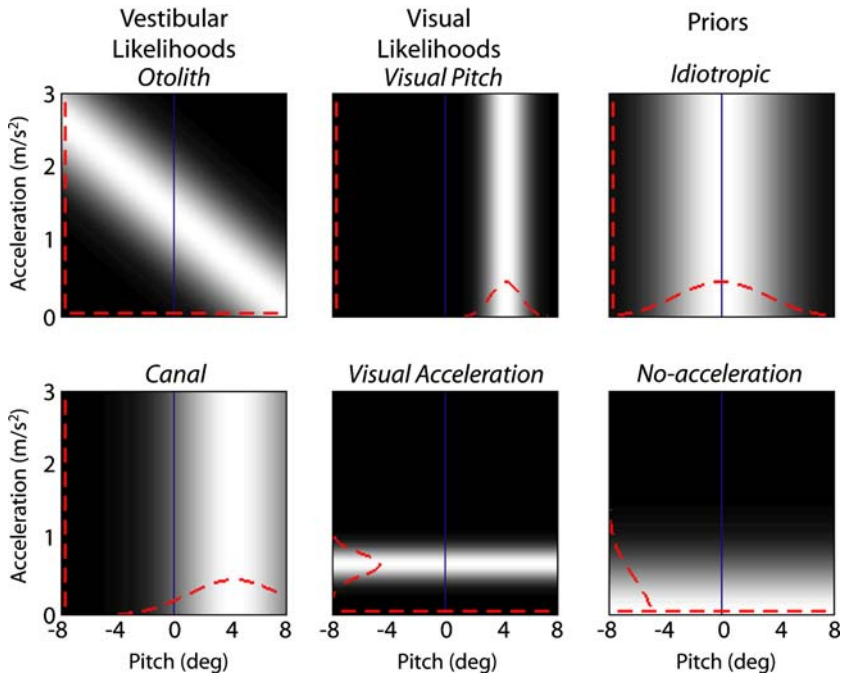

Fig. 2 Bayesian model for pitch and acceleration estimation. *Each panel* shows a probability density as a function of pitch and inertial acceleration. The ordinate has been truncated at the *bottom of each panel* because we are plotting the magnitude of inertial acceleration, which can never be negative. The *red dashed lines* at the margins of the panels are the marginal probability densities; these are either Gaussian or uniform in shape. The *left panels* show likelihood functions associated with vestibular signals. The gravitoinertial force has a direction of 8° and magnitude of 9.81 m/s^2. The otolith likelihood (derived from Eq. 3) is consistent with the direction and magnitude of the gravitoinertial force. The illustrated canal likelihood is associated with a pitch of 4°. The *middle panels* show the likelihood functions associated with the visual signals. The visual pitch signal is centered at 4° and the visual acceleration signal at 0.68 m/s^2. (Note that the visual acceleration likelihood includes the use of a distance estimate to solve the scale ambiguity; Eq. 5). The *right panels* show priors: the idiotropic prior and the no-acceleration prior. Perceptual decisions should be based on the posterior distribution (not shown) which is the product of the likelihoods and priors

$$\|\boldsymbol{I}\| = \sqrt{\|\boldsymbol{G}\|^2 + \|\boldsymbol{F}\|^2 - 2\|\boldsymbol{G}\|\|\boldsymbol{F}\|\cos(\angle\boldsymbol{F} - \angle\boldsymbol{G})} \quad (3)$$

where $\angle\boldsymbol{F}$ and $\|\boldsymbol{F}\|$ are given by the force measured by the otoliths, and $\|\boldsymbol{G}\|$ is constant[3] and equal to 9.81 m/s^2. We can then plot the pairings of acceleration magnitude ($\|\boldsymbol{I}\|$) and pitch ($\angle\boldsymbol{G}$) that are consistent with an observed $\boldsymbol{F}$. Figure 3 shows the relationship between likelihoods in Z–Y space (head-centric coordinates, mid-sagittal plane) and pitch-acceleration space for different situations. The top row represents an upright observer who is stationary in earth coordinates. The measured force is shown in the left panel as a blue vector; it is down on the head at (0 m/s^2, –9.81 m/s^2). The spread of the distribution represents the uncertainty of the estimate. When converted into pitch-acceleration space (right panel), the otolith likelihood has a minimum at a pitch of 0° where $\angle\boldsymbol{G} = \angle\boldsymbol{F}$ and acceleration is 0 m/s^2; the function is symmetrical about this minimum value.[4] Notice that there are infinite combinations of pitch and inertial acceleration that are consistent with the observed otolith signal: hence gravitoinertial force is an ambiguous indicator of self-orientation. The middle row of Fig. 3 shows the otolith likelihoods when the head of a stationary observer is pitched back by 45°. The measured force (6.94 m/s^2, –6.94 m/s^2) is shown in the left panel. The likelihood in pitch-acceleration space is shifted along the pitch axis by 45°. Again the minimum occurs where $\angle\boldsymbol{G} = \angle\boldsymbol{F}$, and there are infinite combinations of pitch and acceleration that are consistent with the observed otolith signal. Thus, the measurement is again an

[3] This conversion requires the assumption that $\|\mathbf{G}\|$ = 9.81 m/s^2. This can be thought of as a prior on $\|\mathbf{G}\|$. This prior also has some variability associated with it, and that will generate additional variability in the otolith likelihood.

[4] Pitch is a circular variable and acceleration is not. So plots of the likelihood functions in pitch-acceleration space should be cylindrical plots. Each point on the cylinder has a position along the circumference that represents pitch and a position perpendicular to the circumference that represents acceleration. For simplicity, we show the unwrapped cylinders in our figures.

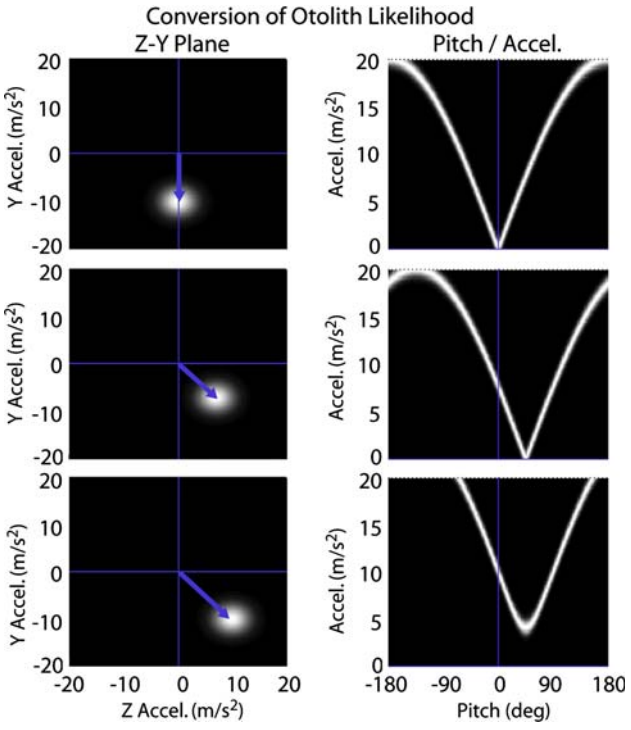

Fig. 3 Otolith likelihoods in *Z–Y* space and pitch-acceleration space. The *left column* shows the otolith likelihood functions when plotted in head-centric coordinates (mid-sagittal, *Z–Y* plane) using units of acceleration. The *right column* shows the same likelihoods when converted into units of pitch and inertial acceleration (Eq. 3). The *top row* shows the likelihoods for an upright observer who is stationary in earth coordinates. The acceleration stimulus is (0 m/s^2, –9.81 m/s^2) and is represented by the blue vector in the *left panel*. The distribution around the *tip* of the vector represents the probability of different observations given the signal. The pitch-acceleration likelihood in the *right panel* is V-shaped with a minimum at (0°, 0 m/s^2). The *middle row* shows the likelihoods for a stationary observer with the head pitched back by 45°. The acceleration stimulus, represented by the blue vector are 6.94 and –6.94 m/s^2. The pitch-acceleration likelihood is again V-shaped but is shifted along the pitch axis by 45°. The *bottom row* shows the likelihoods for an upright observer experiencing a forward inertial acceleration equal to 9.81 m/s^2. The acceleration stimulus are 9.81 and –9.81 m/s^2. The pitch-acceleration likelihood again has a minimum at a pitch of 45°, but the likelihood is now less V-shaped

ambiguous indicator of self-orientation. The bottom row shows the otolith likelihoods for an upright observer accelerating forward at 9.81 m/s^2. The measured force (9.81 m/s^2, –9.81 m/s^2) is shown in the left panel. The direction of the resultant gravitoinertial force is again 45° relative to the *Y*-axis, but the magnitude, 13.87 m/s^2, is greater than gravitational force. The minimum of the likelihood still occurs where $\angle \boldsymbol{G} = \angle \boldsymbol{F}$, but the acceleration is greater than 0 m/s^2; the function also becomes less V-shaped (indeed, one can show that as the magnitude of $\boldsymbol{F}$ approaches 0 or infinity, the likelihood approaches a horizontal ridge). The otolith signal is again an ambiguous indicator of self-orientation.[5]

Semi-circular canals

The canals are driven by angular and not linear acceleration. For brief accelerations, their afferent signals are proportional to angular velocity (Goldberg and Fernandez 1971), so those signals by themselves do not provide an unambiguous indication of absolute head orientation. If we assume that the head is initially upright in earth coordinates (as it was in our experiments), the canal afferent signals integrated over time could in principle be used to infer the pitch angle at the end of a motion sequence. The lower left panel in Fig. 2 shows a canal likelihood function: for a canal signal indicating backward rotation about the *X*-axis, the final specified angle is pitched backward. Because the canal signal is unaffected by linear acceleration, the likelihood is anisotropic: narrow in pitch and infinite in linear acceleration. Such a signal can thus be used directly for estimating pitch, but not acceleration.

Angelaki and colleagues and Merfeld and colleagues have shown that canal and otolith signals are combined in the nervous system to resolve the ambiguity of the otolith signal (Angelaki et al. 1999, 2004; Merfeld et al. 2001). In the Bayesian model, this can be represented by the product of the otolith and canal likelihoods. However, in the conditions of our experiment, the canal signal was made intentionally unreliable by using low angular velocities. Hence the variance of the canal likelihood in Fig. 2 is depicted as large relative to the variance of the other signals.

Visual pitch

The visual orientation of the horizon is a useful cue to orientation. For example, when the observer's pitch changes relative to flat terrain, the elevation of the horizon changes systematically. For a pilot whose head is upright relative to the aircraft, the visual elevation of the horizon lowers as the aircraft's pitch increases. We can represent this signal with the visual-pitch likelihood in Fig. 2. The function is narrow in pitch because visual elevation is a reliable pitch cue. It is infinite in acceleration because visual elevation does not vary with acceleration. When no visual pitch cue is present, the likelihood function is uniform and therefore has no effect on the posterior distribution.

[5] For simplicity, we do not consider separately kinesthetic and somatosensory signals because the forces that affect them are the same as the forces that drive the otoliths.

Visual acceleration

The velocity and acceleration of texture elements in the optic flow at the retina is a useful cue to acceleration relative to a rigid scene. For an observer moving in the forward direction (Z) relative to a rigid scene, there are two equations for the observer's speed:

$$\dot{Z} = \frac{Z\dot{x}}{x_0 - x} \quad \dot{Z} = \frac{Z\dot{y}}{y_0 - y} \tag{4}$$

where X, Y, and Z are the horizontal, vertical, and forward coordinates of a given texture element relative to the eye, x and y are the horizontal and vertical retinal coordinates associated with that element, x_0 and y_0 are the retinal coordinates of the focus of expansion, and $\dot{x}$ and $\dot{y}$ are the velocities of the texture element (Longuet-Higgins and Prazdny 1980; Royden et al. 1994). The acceleration of the observer is found by taking the derivatives of Eq. 4:

$$\ddot{Z} = \frac{Z\ddot{x}}{x_0 - x} \quad \ddot{Z} = \frac{Z\ddot{y}}{y_0 - y} \tag{5}$$

where $\ddot{x}$ and $\ddot{y}$ are the accelerations of the texture element. All of the quantities on the right side of Eqs. 4 and 5 are measurable from the optic flow at the retina except for Z, the distance to the texture element. Thus, an estimate of Z is required to estimate the speed and acceleration of translation. This is known as the scale ambiguity of optic flow (Longuet-Higgins and Prazdny 1980; Royden et al. 1994); one needs to scale optic flow measurements by an estimate of the distance to obtain absolute velocity and acceleration estimates.[6]

Figure 2 shows the likelihood associated with the visual acceleration signal. It is narrow in acceleration because the signal varies with acceleration and is infinite in pitch because the signal does not vary with orientation. When the visual acceleration signal is absent, the visual acceleration likelihood is uniform and has no effect on the posterior.

Upright prior: the idiotropic vector

Mittelstaedt (1983) observed that humans in gravity-free environments with no disambiguating visual information perceive vertical as aligned with the longitudinal body axis. He referred to this bias as the idiotropic vector. In the Bayesian framework, this corresponds to a prior probability distribution; it is illustrated in the upper right panel of Fig. 2 for a head upright with respect to the body. It is centered on a pitch of 0°, indicating the bias toward upright with respect to the head and body. The distribution is infinite in acceleration because the prior concerns pitch only.

No-acceleration prior

The somatogravic illusion demonstrates that humans, faced with interpreting an ambiguous otolith signal in the absence of other information, prefer a solution that minimizes the inertial acceleration. We represent this tendency with the no-acceleration prior (lower right panel, Fig. 2). The prior represents the tendency for an earth-bound observer to not be accelerating relative to earth. The distribution is much narrower in forward acceleration than in pitch because the prior concerns acceleration only. It peaks at 0 m/s^2 indicating the bias toward no inertial acceleration. The no-acceleration prior is somewhat similar to the prior for slow speeds described by Weiss et al. (2002).

Priors generally have large variance relative to likelihoods (e.g., Weiss et al. 2002; Körding and Wolpert 2004; Hillis et al. 2004; Stocker and Simoncelli 2006), so they have little effect on perception when rich sensory information is available. Accordingly, we have depicted the prior distributions as having large variance.

Posterior distribution

The sensory signals under consideration are physically quite different and are signaled by different sensory apparatuses. It is likely, therefore, that their noises are statistically independent (that the signals are conditionally independent). With this assumption, the likelihood can be written as the product of likelihoods associated with each individual signal: otoliths, canals, visual pitch, and visual acceleration. The prior distributions—idiotropic and no-acceleration—are orthogonal so the joint prior distribution can be represented by their product. With the assumptions of conditional independence and separable priors, Bayes' Law for estimating pitch and acceleration can be written as:

$$\begin{aligned} p(O, A|S_o, S_c, S_p, S_a) \propto\ & p(S_o|O, A)p(S_c|O, A) \\ & p(S_p|O, A)p(S_a|O, A)p(O)p(A) \end{aligned} \tag{6}$$

where O is pitch relative to earth horizontal, A is inertial acceleration (that is, acceleration relative to the earth), and S_o, S_c, S_p, and S_a are the signals from otoliths, canals, visual pitch, and visual acceleration,

[6] Note that any error in scaling the scene will lead to an error in the velocity and acceleration estimates.

respectively. The posterior is simply proportional to the product of the likelihoods and priors.

There are two important properties of Bayesian estimation that can be visualized when one represents the posterior distribution as the product of conditionally independent likelihoods and priors. First, the combined estimate is influenced most by the least variable of the likelihoods and priors. In other words, the most reliable signal has the greatest influence on the final estimate. Second, the variance of the posterior is guaranteed to be smaller than the variance of any of the individual likelihoods or the priors: Bayesian estimation yields a less variable estimate than can be derived from any of the individual signals (Ghahramani et al. 1997).

In the experiments reported here, subjects were asked to judge pitch and inertial acceleration. In the framework of the Bayesian model, they would base those judgments on the marginal posterior distributions. These 1d distributions are obtained by integrating the 2d posteriors across the irrelevant dimension. The marginal distributions of the likelihoods, priors, and posterior are shown by the red dashed lines at the margins of each panel in Fig. 2.

Some examples

Figure 4 shows two examples of orientation/acceleration estimation in the framework of the Bayesian model. The examples depict the likelihoods and the posteriors associated with conditions in our experiment. (Priors are not dependent on sensory input so they remain the same as in Fig. 2.) In both examples, the observer is pitched backward by 4°, so the otoliths signal a force vector that is rotated from upright (Fig. 4, top row). The head was pitched by slow passive rotation, so the canal signal has high variance (second row). Note that the vestibular stimuli are the same in both examples; only the visual stimuli differ.

In the ''pitch only'' example (left column), the visual pitch cue (third row) signals that the head has been pitched by 4°. The visual acceleration (fourth row) is 0 m/s^2. The posterior (bottom row) is the product of the likelihoods and priors (Eq. 6). The model's pitch and acceleration estimates (the maximum of the posterior) are approximately 4° and 0 m/s^2, respectively, in this condition: the model's estimates are very similar to the actual state of affairs.

In the ''acceleration only'' example (right column), the visual pitch cue signals 0° and the visual acceleration is $\|G\|tan(4^\circ) = 0.68$ m/s^2, a value consistent with the hypothesis that the rotation of the gravitoinertial force vector is due to linear acceleration. The posterior is shown in the bottom row. The model's pitch and acceleration estimates are now approximately 0° and 0.68 m/s^2, respectively. In this condition, the model's estimates are consistent with forward acceleration and no pitch, which is quite different from the actual state of affairs.

These examples illustrate in the framework of the Bayesian model how visual signals affect the interpretation of the otolith signal. The examples are also consistent with the observation that the somatogravic illusion only occurs when useful visual information is not available (Gillingham and Previc 1993).

In the experiments reported here, we examined whether the perception of orientation and acceleration can be characterized as a Bayesian estimation process. To do so, we looked for influences of visual signals on the interpretation of otolith signals and for improvements in discrimination performance when visual and vestibular signals were both present compared to when only vestibular signals were present.

Methods

Overview

The experiments were designed to measure each subject's ability to discriminate pitch and the magnitude of inertial acceleration on the basis of visual cues only, vestibular cues only, and visual and vestibular cues together. Depending on the experimental condition, we varied three signals in these experiments: visual pitch, visual acceleration, and body pitch (thereby affecting the otolith signal). We used a 2-interval, forced-choice procedure in all conditions; on a given trial, subjects were presented two stimuli in succession and they indicated which stimulus had more pitch and/or more acceleration.

There were three two-modality, visual-vestibular conditions: the pitch condition in which subjects reported perceived pitch, the acceleration condition in which they reported perceived acceleration, and the pitch-plus-acceleration condition in which they reported both. In each of those conditions, two signals were varied systematically and one was held constant. In the pitch condition, visual pitch and body pitch were varied and visual acceleration was held constant. Thus, this condition was designed to measure the influences of visual pitch and body pitch on perceived pitch. In the acceleration condition, visual acceleration and body pitch were varied and visual pitch was held

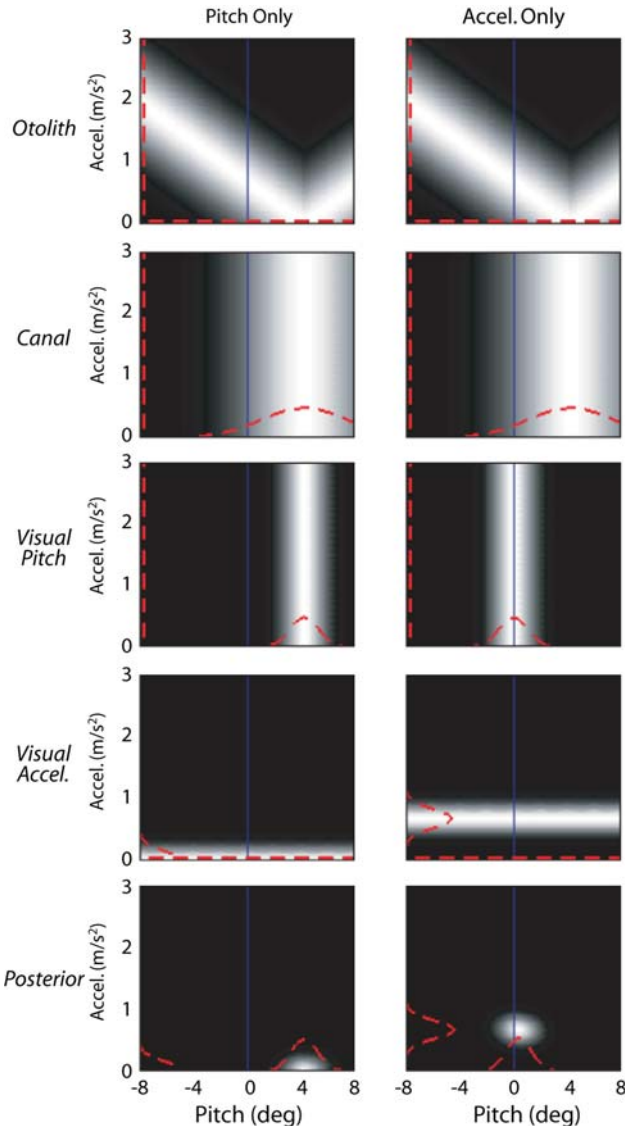

Fig. 4 Two examples of Bayesian estimation of pitch and acceleration. The *upper four rows* are the likelihoods associated with the otoliths, canals, visual pitch, and visual acceleration. The *bottom row* is the posteriors for the two situations. The marginal distributions are represented by the *red dashed curves*. The *left and right columns* are two different situations. In both situations, the otolith signals are consistent with a gravitoinertial vector rotated 4° with respect to upright. In the situation on the *left*, the visual signals specify 4° pitch and forward acceleration of 0 m/s^2. In the situation on the *right*, they specify 0° pitch and acceleration of 0.68 m/s^2. Note that the posterior distributions are quite different in the two situations because of the disambiguating visual information

constant. Thus, this condition measured the influences of visual acceleration and body pitch on perceived acceleration. In the pitch-plus-acceleration condition, visual pitch and visual acceleration were varied and body pitch was held constant. This condition was therefore designed to measure the influences of visual signals on perceived pitch and acceleration.

Subjects

Eight naïve subjects (four males and four females) participated. They were 20–30 years old and had normal or corrected-to-normal vision. They were registered in the subject database at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, so most had prior experience with psychophysical experimentation. Before beginning the experiment, each subject participated in a 2-h, cues-consistent training session in which they made pitch and acceleration judgments. No feedback was given during the training session. All experiments were conducted in accordance with the requirements of the Helsinki Declaration.

Apparatus

The experiment was conducted in the Motion Lab at the Max Planck Institute. The main piece of equipment is the Maxcue 600 platform manufactured by Motion-Base PLC (Fig. 5). The motion platform is connected to an earth-fixed base with six variable-length legs. The platform moves via changes in the lengths of the legs. The platform can move with six degrees of freedom: three translations (*X*, leftward-rightward; *Y*, upward-downward; and *Z*, forward-backward) and three rotations (about *X*, *Y*, and *Z*; pitch, yaw, and roll, respectively). Subjects were seated on the platform and

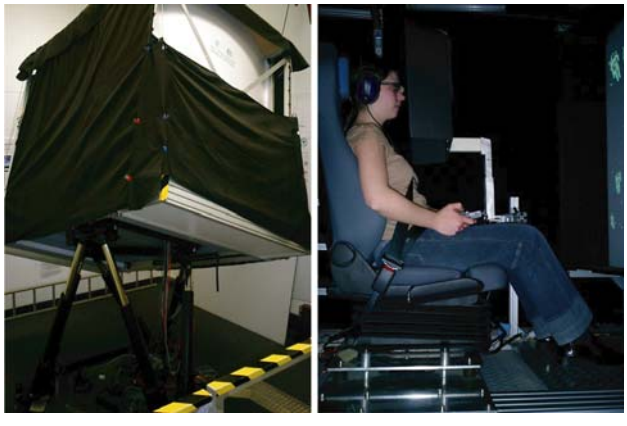

Fig. 5 The motion platform. *Left* : Outside view of the platform. The platform translates in three dimensions and rotates about three axes through the action of six variable-length legs. *Right* : Inside view the subject is seated in front of a projection screen and views the screen monocularly through an aperture such that the edges of the screen cannot be seen. Responses are made with the hand-held response box

were instructed to keep their heads still during the experiment. A shaped foam pad behind the head aided the head stabilization. Platform movement generated vestibular and somatosensory stimulation.

Visual stimuli were presented on a projection screen in front of the subject. The projector and projection screen were attached to the platform, so their positions were fixed relative to the subject. Viewing was monocular. An aperture 15 cm in front of the subject's face restricted the field of view to 50 × 50° so that the edges of the projection screen were not visible.

Verbal instructions were given through headphones. Noise played in the headphones masked the sounds associated with platform movement. Vibrations associated with platform movement were masked by vibrating the seat mount and footrest independently from platform movement. Subjects responded with key presses.

Stimuli

We created the vestibular stimulus by pitching and translating the motion platform. Movement profiles of the motion platform during one presentation are shown in the left column of Fig. 6. Platform movements change the direction and magnitude of the gravitoinertial force experienced by the subject; the right column of Fig. 6 shows gravitoinertial force profiles generated by the platform movements shown on the left.

A stimulus presentation lasted 9 s. Platform rotation about the *X*-axis began after 1 s, continued at a constant rate for 3 s to a pre-specified angle, remained for 2 s at the target pitch angle, and then rotated back to upright at a constant rate for 3 s (Fig. 6a). This movement modified the direction of gravitational force in head-centric coordinates (Fig. 6b). Target pitch angle was the only parameter of platform motion that was varied systematically from presentation to presentation. The platform rotations (pitch up followed by pitch down) had angular velocities of 0.67–2.67°/s depending on the target pitch angle.[7] We wanted to isolate the otolith signal as well as possible, so we chose those slow angular velocities because they are below or just above the canal threshold (Benson et al. 1989). In Bayesian terms, we aimed to generate a canal likelihood with high variance relative to the otolith likelihood. In the Discussion, we present evidence that we succeeded, and that the canal signal had little if any influence on perceptual judgments.

The platform also translated during each presentation in order to increase the realism of the simulation (Groen et al. 2001; Reymond and Kemeny 2000). Random vertical oscillations (±5 cm) simulated bumps that are often experienced during vehicular movement (Fig. 6c). This earth-vertical movement generated an inertial force in the *Y* direction; the resulting gravitoinertial force profile is shown in Fig. 6d. The platform also translated forward 50 cm and then backward in the *Z* direction during each presentation (Fig. 6e), generating an earth-horizontal inertial force that also modified the gravitoinertial force in head-centric coordinates (Fig. 6f). During the experiment the rotation and translations occurred simultaneously; Fig. 6h shows an example of the resulting gravitoinertial force profile. In pilot testing, we confirmed that subjects experienced pitch and sustained acceleration during these platform movements.

The visual stimuli depicted self-motion through a rigid environment. Depending on the condition, the type of self-motion was (a) pitch rotation during constant-speed forward translation (a "wheelie"), b) accelerating forward translation (no pitch), or c) pitch rotation during accelerating forward translation. The pitch and/or acceleration specified by the visual stimulus were synchronized with the platform rotation. The visual stimulus faded on during the first second of the presentation. Visual pitch and/or acceleration increased at a constant rate for the next 3 s, remained constant for 2 s, and decreased at a constant rate for the last 3 s. The target pitch and/or acceleration were the only parameters of the visual stimulus that were varied from presentation to presentation. To increase simulation realism, the viewpoint on the scene oscillated up and down in synchrony with the vertical translations of the platform (Fig. 6c).

There were two types of visual scene: one simulating translation above a ground plane and one simulating translation through a cloud of dots (Fig. 7). The texture on the ground plane was composed of an overlay of random patterns of varying spatial frequency (Berger 2003). Most of the elements in the dot cloud were frontoparallel disks distributed uniformly in three dimensions. As we pointed out in the modeling section, one cannot estimate self-acceleration from optic flow without an estimate of the distance to elements in the scene that generated the flow (Eq. 5). To provide distance information, we placed some life-size human figures on the plane and in the dot cloud.

[7] The maximum angular acceleration, which occurred very briefly at the beginning of platform motion, was ~12.5°/s^2. We had to place the rotation axis in the floor of the platform to maximize the motion range. The axis was ~100 cm beneath and ~26 cm in front of the center of the subject's head. Thus, the angular acceleration generated small tangential and centripetal acceleration at the head. These accelerations were quite small and had virtually no influence on the total gravitoinertial force profiles.

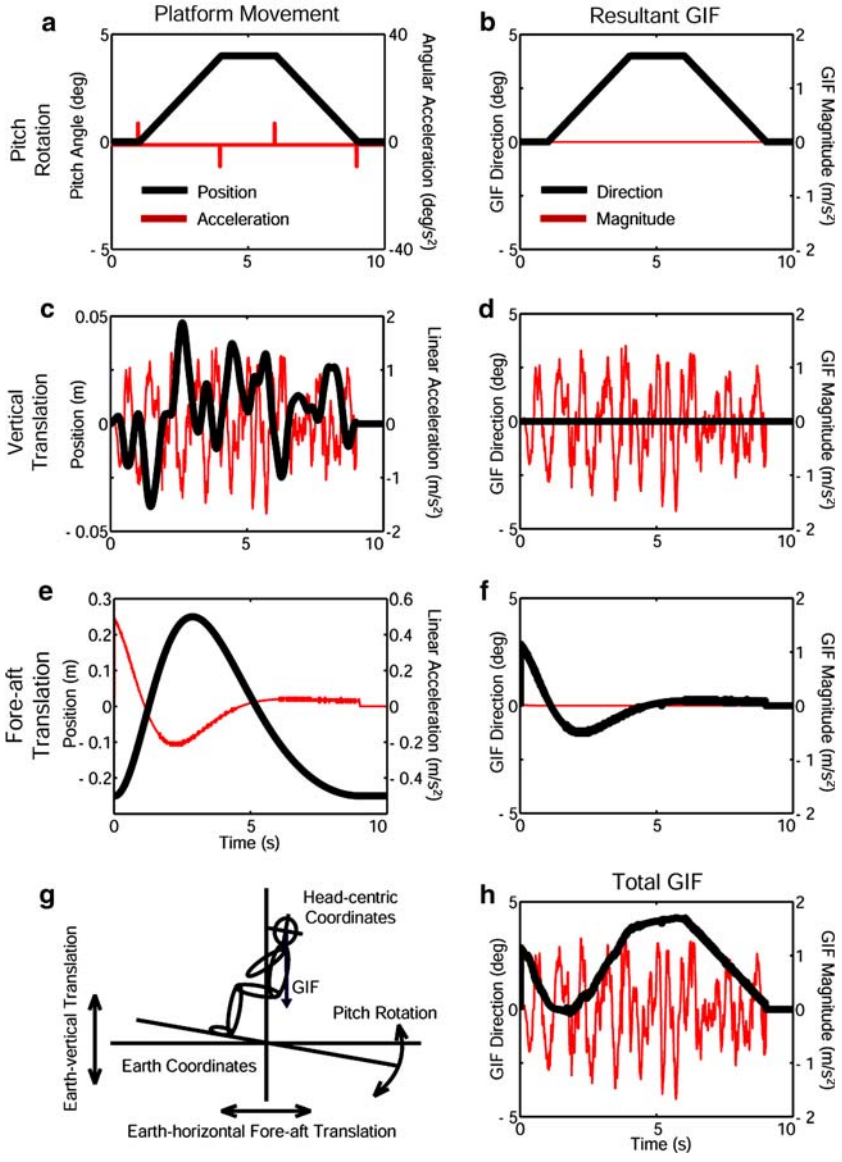

Fig. 6 Platform movements and resultant gravitoinertial forces (*GIF*). The *left column* shows platform movements in earth coordinates. On a single presentation, the platform simultaneously rotated to a target pitch (**a**), oscillated slightly vertically (**c**), and translated fore and aft (**e**). Those movement directions are illustrated in (**g**). The *right column* shows the resulting GIF direction and magnitude (in head coordinates). Pitch (**b**) changes the direction, but not the magnitude of GIF. Vertical oscillations (**d**) affect GIF magnitude but not direction. Fore-aft translation (**f**) generates an inertial force perpendicular to gravitational force. This primarily affects GIF direction; magnitude is minimally affected. The resulting GIF generated during a stimulus presentation is shown in (**h**)

The two scenes convey pitch and acceleration information in different ways. The horizon is earth fixed, so it provides absolute information about self-orientation in earth coordinates (Gibson 1950). Thus, the elevation of the visible horizon was an excellent pitch cue. It was available in the ground-plane but not the dot-cloud scene. The elevation of the focus of expansion varied with simulated pitch in the two scenes because the simulated pitch was independent from the simulated translation (which was always earth horizontal); thus, change in the elevation of the focus of expansion was also a pitch cue, but not a salient one. From these considerations, we expect the variance of the visual-pitch likelihood to be smaller with the ground plane than with the dot cloud. Rotational optic flow was present in both scenes and provided information about the angular velocity of pitch. An angular velocity estimate can be converted into a pitch estimate by assuming a known starting position (upright in our experiment). Acceleration information in both scenes was conveyed by the optic flow of the scene elements. The scenes differed in the spatial distribution of scene elements (along the ground for the ground plane, uniformly 3D for the dot cloud) and this could influence the precision of self-acceleration estimates.

Procedure

We used a 2-interval, force-choice procedure. A trial consisted of two successive 9-s presentations separated by 2 s. One presentation was the standard and the other was the comparison; order was randomized.

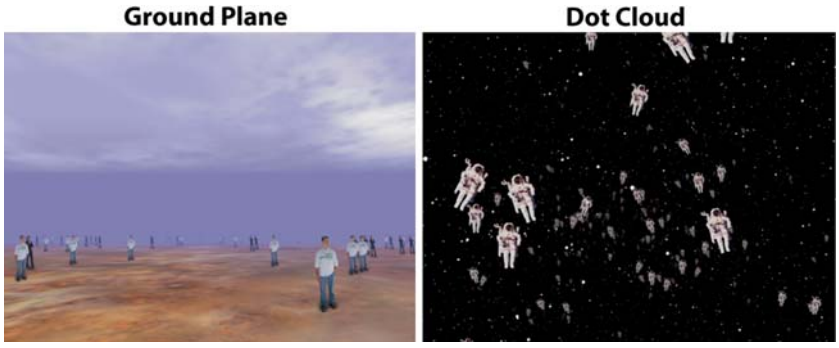

Fig. 7 Visual stimuli. *Left* ground-plane stimulus. The ground plane was textured with random patterns of varying spatial frequency. Human figures (1.9 m high) were placed on the plane to help solve the scale ambiguity. Forward motion across the ground plane was simulated by motion of the texture and figures. Pitch was simulated by the change in elevation of the focus of expansion and by the elevation of the horizon. *Right* Dot-cloud stimulus. The dot cloud was a uniform distribution of fronto-parallel disks (meaning that they are parallel to the image plane). The disks were 19 cm in diameter. The human figures (1.9 m high) were presented to help solve the scale ambiguity. Forward motion through the cloud was simulated by motion of the disks and figures. Pitch was simulated by a change in the elevation of the focus of expansion

After each trial, the subject indicated in which presentation he/she felt more pitched and/or more accelerated. They were asked to base their judgments on the 2-s steady-state portion of the presentations (e.g., the plateau in Fig. 6a) when pitch and/or acceleration were greatest. No feedback was given.

Each standard stimulus value was paired with a range of comparison values, and each pairing was presented 12 times. For each comparison value, we calculated the percentage of times that value was judged as more pitched or accelerated than the standard. A cumulative Gaussian was fit to this psychometric data for each standard stimulus using a maximum-likelihood criterion (Wichmann and Hill 2001). The 50% point on the best-fitting function is the comparison stimulus value that was on average perceptually equivalent to the standard stimulus; this is the point of subjective equality, or PSE. The slope of the best-fitting function is a measure of the just-discriminable difference, or JND. It is expressed by the standard deviation of the best-fitting cumulative Gaussian. For 2-alternative, forced-choice procedures, the standard deviation of the fitted function is $\sqrt{2}$ times the standard deviation of the underlying estimator (Ernst and Banks 2002).

We measured discrimination thresholds in vision-only and non-vision-only experiments to determine the variances associated with those single-modality estimates. These four conditions were non-visual pitch, non-visual acceleration, visual pitch, and visual acceleration.

Subjects wore a blindfold in the non-visual conditions allowing us to isolate the vestibular signals. For the non-visual pitch condition, subjects indicated the presentation in which they felt more pitched; thus, they were instructed to interpret changes in the vestibular signal as caused by changes in pitch and not in inertial acceleration. The standard was 4° of platform pitch and the comparison ranged from 2 to 6° in 0.5° increments. For the non-visual acceleration condition, platform movements were exactly the same, but subjects were instructed to interpret the ambiguous otolith signal as due to acceleration and not pitch.[8] They indicated the presentation in which they felt more accelerated. The forward acceleration necessary to produce a given direction of gravitoinertial force in an upright observer exposed to gravity is $\|\boldsymbol{I_e}\| = \|\boldsymbol{G}\| \tan(\angle \boldsymbol{F})$: this is the equivalent acceleration. (The equivalent pitch for a given acceleration is $\angle \boldsymbol{G_e} = \tan^{-1}(\|\boldsymbol{I}\|/\|\boldsymbol{G}\|)$.) The equivalent acceleration of the 4° pitch standard is 0.68 m/s^2. The equivalent accelerations for the comparison stimulus ranged from 0.34–1.02 m/s^2 in 0.085-m/s^2 increments.

In the vision-only conditions, the platform was always upright and stationary. It was not technically possible to isolate the visual signals with this procedure because gravitational force indicating stationary, upright orientation was always present, but we assume

[8] In pilot testing, we found that people experienced both forward translation and pitch when presented the non-visual stimuli. Thus, it was reasonable to ask subjects which of two intervals yielded greater perceived acceleration and which of two intervals yields greater perceived pitch. Nonetheless, the mere instruction to interpret the otolith signal one way (e.g., as pitch, not acceleration) does not mean subjects are capable of completely succeeding, which can lead both to bias (when they attribute some of the signal to the acceleration instead) and increased variability (when the amount they attribute to acceleration varies from trial to trial).

that subjects ignored this irrelevant and constant signal. In the visual-pitch condition, subjects indicated the presentation in which they felt more pitched. The standard was 4° and the comparison ranged from 2–6° in 0.5° increments. For the visual-acceleration condition, subjects indicated the presentation in which they felt more accelerated. The standard was 0.68 m/s^2 and the comparison ranged from 0.34–1.02 m/s^2 in 0.085-m/s^2 increments.

We also ran three multi-cue conditions to determine the influence of visual signals on the interpretation of the otolith signal. The procedure was very similar to the one in the single-cue conditions. Stimulus values were varied as illustrated in Fig. 8. On each trial, two stimuli were presented: a cues-consistent (comparison) and a cues-inconsistent (standard) stimulus. Each pairing was presented 12 times and presentation order was randomized.

In the pitch condition (Fig. 8a), the platform pitch and visual pitch of the cues-consistent stimulus were both 4 + Δ° (Δ ranged from –2 to 2°). The platform pitch and visual pitch of the cues-inconsistent stimulus were 4–δ° and 4 + δ°, respectively (δ was –1, 0, or 1°). Visual acceleration was always 0 m/s^2 in this condition. Visual speed was 5 m/s. Subjects indicated after each trial the presentation in which they felt more pitched.

In the acceleration condition (Fig. 8b), the platform pitch of the cues-consistent stimulus was 4 + Δ° (Δ ranged from –2 to 2°), and the visual acceleration was the equivalent acceleration for those pitch values: 0.68 + Λ m/s^2 (Λ ranged from –0.34 to 0.34 m/s^2). In the cues-inconsistent stimulus, platform pitch was 4-δ° and visual acceleration was 0.68 + λ m/s^2. δ was –1, 0, or 1°, and λ was –0.17, 0, or 0.17 m/s^2. The visual pitch was always 0° in this condition. Subjects indicated the interval in which they felt more accelerated.

We also presented a pitch-plus-acceleration condition. This condition was run to check for an influence of visual pitch on acceleration judgments and visual acceleration on pitch judgments. In this condition (Fig. 8c), the platform pitch was always 8°, and visual pitch and visual acceleration were varied. For the cues-consistent stimulus, they were varied such that the visual pitch value and the equivalent pitch for the visual acceleration value summed to 8°. The visual pitch was 4-Δ° and the visual acceleration was 0.68 + Λ m/s^2 (equivalent pitch of 4 + Δ°). Δ ranged from –2 to 2°, and Λ from –0.34 to 0.34 m/s^2 (equivalent pitch of –2 to 2°). In the cues-inconsistent stimuli, the equivalent pitch specified by the visual stimulus did not sum to 8° (unless $\delta = \sigma = \lambda = 0$°). The visual pitch was 4 + δ° and visual acceleration was 0.68 + λ m/s^2. δ was –1, 0, or 1°, and λ was –0.17, 0, or 0.17 m/s^2, such that the total equivalent pitch was 6, 8, or 10°, respectively. After each trial, subjects made two responses: the interval in which they felt more pitched and the one in which they felt more accelerated. They gave those responses in different orders from one session to another.

The data in each multi-cue condition were again fit with a cumulative Gaussian. The mean of the Gaussian was the estimate of the PSE, the value of Δ or Λ for which the cues-consistent and cues-inconsistent stimuli were on average perceptually equivalent (equally pitched or accelerated). The standard deviation of the

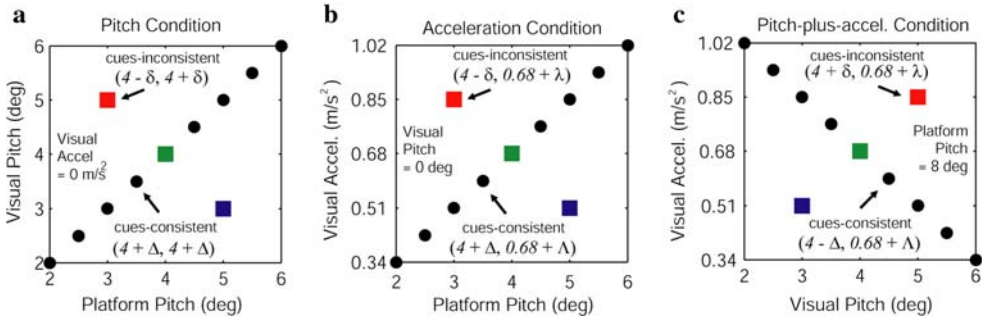

Fig. 8 Stimulus values in the multi-cue conditions. **a** Stimulus values for the pitch condition. Platform pitch and visual pitch were varied. On each trial, a cues-consistent and a cues-inconsistent stimulus were presented. The pitches of the cues-consistent stimulus are represented by the *black circles*; each had a value of 4 + Δ°. The pitches of the cues-inconsistent stimulus are represented by the *red*, *green*, and *blue squares*; platform pitch was 4-δ ° and visual pitch was 4 + δ°. Visual acceleration was always 0 m/s^2. **b** Stimulus values for the acceleration condition. Platform pitch and visual acceleration were varied. Platform pitch and visual acceleration of the cues-consistent stimulus were respectively 4 + Δ° and 0.68 + Λ m/s^2 (an equivalent pitch of 4 + Δ°). Platform pitch and visual acceleration of the cues-inconsistent stimulus were 4-δ° and 0.68 + λ m/s^2. Visual pitch was always 0°. **c** Stimulus values for the pitch-plus-acceleration condition. Visual pitch and visual acceleration were varied. Pitch and acceleration in the cues-consistent stimulus were respectively 4-Δ° and 0.68 + Λ m/s^2; they summed to a specified pitch of 8°. Pitch and acceleration in the cues-inconsistent stimulus were respectively 4 + δ° and 0.68 + λ m/s^2; they did not necessarily sum to a specified pitch of 8°. Platform pitch was always 8°

best-fitting Gaussian was the estimate of the JND. A boot-strap procedure (1999 repetitions) was used to calculate confidence intervals for all PSE and JND estimates (Wichmann and Hill 2001).

Four subjects were presented the ground-plane stimulus and four were presented the dot cloud. The single-cue conditions of visual pitch and visual acceleration were not included in the original experimental design, so those measurements were done after the other conditions had been completed. Unfortunately, we were not able to collect the vision-only data for three of the eight subjects.

There were seven conditions total which took ~12 h of testing for each subject to complete.

Results

Single-cue data

Figure 9 shows examples of the psychometric data from the single-cue conditions. The upper and lower rows show data from the pitch and acceleration judgments, respectively. The left and right columns show data from subjects who saw the ground-plane and dot-cloud stimuli, respectively. Each panel plots the percentage of times the subject indicated that the comparison stimulus was more pitched or accelerated than the standard as a function of the value of the comparison stimulus.

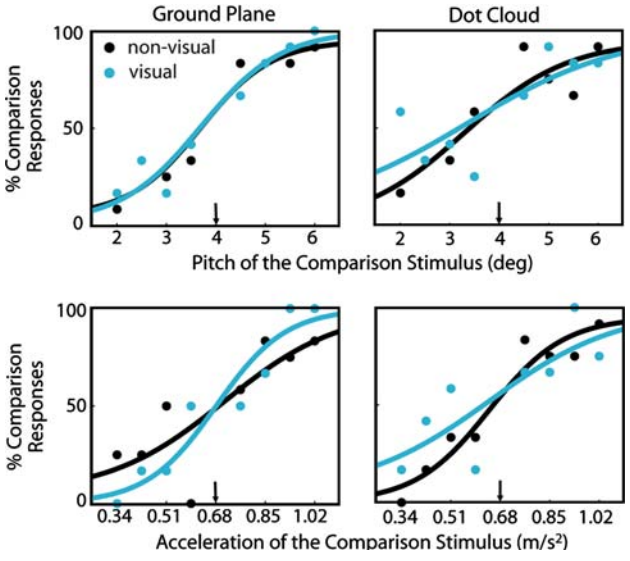

Fig. 9 Example psychometric functions from the single-cue conditions. The *left column* shows data from a subject who saw the ground-plane stimulus and the *right column* shows data from a subject who saw the dot-cloud stimulus. The non-visual data in the *two columns* are from the same condition, but from different subjects. The *upper and lower rows* are from the pitch and acceleration conditions, respectively. The abscissa in each panel is the value of the comparison stimulus; *arrows* indicate the value of the standard stimulus. The ordinates plot the percentage of times the subject indicated that the comparison stimulus was more pitched or accelerated than the standard. The *cyan points and curves* are the data and best-fitting cumulative Gaussians for the visual condition and the *black points and curves* for the non-visual condition

Figure 10 shows the JNDs for all single-cue (and multi-cue, zero-conflict) conditions and subjects. The upper and lower rows show the pitch and acceleration JNDs, respectively. The left and right columns show the JNDs for the ground-plane and dot-cloud stimuli, respectively. The JNDs averaged across subjects are shown on the right of each panel: one set of averages is for all of the subjects and one set is with subjects 4, 6, and 8 removed because they did not contribute data in the visual condition. (The figure does not show data from the pitch-plus-acceleration condition.) We will discuss these findings further in the section on JND analysis.

Multi-cue data

Figure 11 displays examples of the psychometric data in the multi-cue conditions. The upper and lower rows show data from the pitch and acceleration conditions, respectively. The left and right columns show data from subjects who saw the ground-plane and dot-cloud stimuli, respectively. The percentage of times the subject indicated that the cues-consistent (comparison) stimulus was more pitched or accelerated than the cues-inconsistent (standard) stimulus is plotted as a function of the value of the cues-consistent stimulus. The changes from one curve to the next show the effect of changing the conflict value in the cues-inconsistent stimulus.

The main question we addressed is how visual signals affect the interpretation of otolith stimulation. We sought answers in two ways: (1) by measuring perceived pitch and acceleration when otolith and visual cues were varied independently, and (2) by measuring the improvement in pitch and acceleration discrimination when otolith and visual cues were both present compared with when only one was present. The former was based on an analysis of PSEs, and the latter on JNDs.

PSE analysis

For each cues-inconsistent stimulus with conflict δ, there is a cues-consistent stimulus with increment Δ that on average yields the same percept; this is the

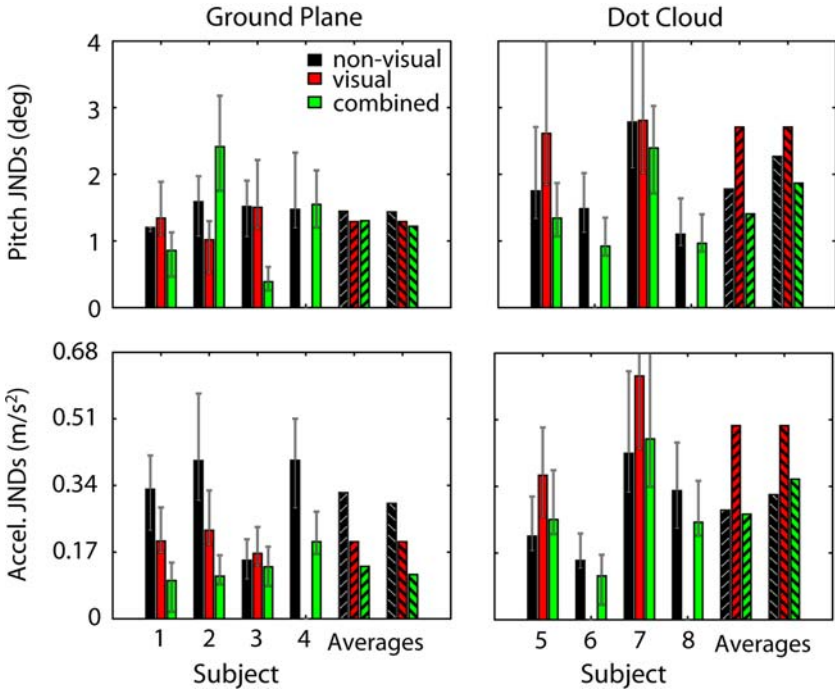

Fig. 10 Just-noticeable differences (*JNDs*) for single-cue and zero-conflict multi-cue conditions ($\delta = \lambda = 0$). JNDs are the standard deviations of the cumulative Gaussians that best fit the psychometric data. The *upper and lower rows* show the pitch and acceleration JNDs, respectively. The *left and right columns* show the JNDs for the ground-plane and dot-cloud stimuli. The non-visual JNDs are represented by the *black bars*, the visual JNDs by the *solid red bars*, and the combined JNDs by the *solid green bars*. The *horizontal axis* represents different subjects. The JNDs averaged across subjects are on the *right* in each panel. Subjects 4, 6, and 8 did not contribute a visual JND, so we show two sets of averages: those with the three subjects included in the averages (for the conditions they provided data; *left diagonal hatching*) and those without the three subjects included in the averages (*right diagonal hatching*). *Error bars* represent 95% confidence intervals obtained by bootstrap (Wichmann and Hill 2001)

PSE. The values of δ and Δ (or equivalently λ and Λ) can be used to plot PSEs in cue-conflict space (Figs. 12, 13). Figure 12 shows the PSEs in the pitch and acceleration conditions for all subjects. The large panels show average data and the small panels show individual subject data. The two columns of panels on the left show data for the pitch condition and the two columns on the right show data from the acceleration condition. The abscissa in each panel represents δ or λ, the value of the conflict in the cues-inconsistent stimulus, while the ordinate represents Δ or Λ, the increment in the cues-consistent stimulus that on average had the same perceived pitch or acceleration as the cues-inconsistent stimulus.

To understand how the use of different cues affects PSEs, consider the pitch condition (left half of Fig. 12). Subjects indicated in which of two stimulus intervals—cues-consistent or cues-inconsistent—they felt more pitched. In principle, this judgment could be made from the visual pitch cue alone; we will refer to this as the direct method. A judgment could not be made from the visual acceleration signal alone or from the otolith signal alone because those signals are individually ambiguous with respect to pitch. But the visual acceleration and otolith signals could be combined to provide an informative pitch estimate; we will refer to this as the indirect method. For the plots on the left side of Fig. 12, the PSEs would lie on lines with slope 1 if subjects based their pitch judgments on only the direct method ($\Delta/\delta = 1$). If they based their judgments only on the indirect method, the PSEs would lie on lines with slope –1 ($\Delta/\delta = -1$). If subjects based their judgments on both methods, the PSEs would lie on lines with slopes in-between –1 and 1. Although there were clear individual differences, some general trends are evident in the pitch PSE data. First, in nearly all cases, the data slopes were between –1 and 1, showing that pitch judgments were affected by both methods. To check the statistical reliability of this result, we performed regression analyses on the PSEs. The slopes in Fig. 12 were significantly greater than –1 ($P < 0.001$, one-tailed for ground plane; $P < 0.01$, one-tailed for dot cloud) and significantly less than +1 ($P < 0.001$, one-tailed in both cases). This means that all three

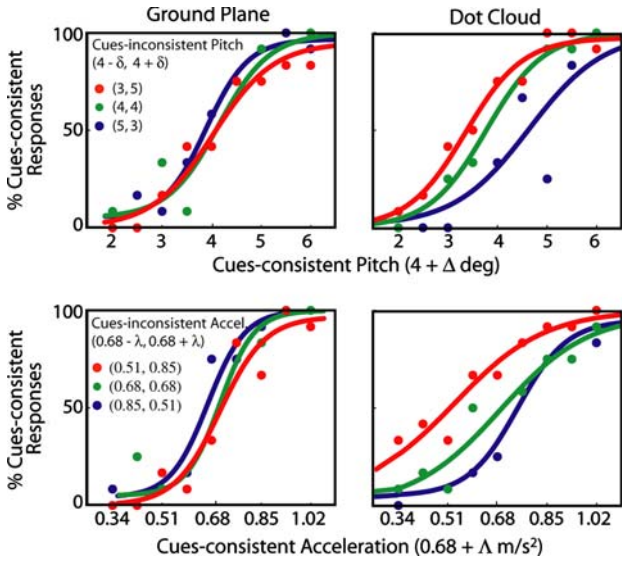

Fig. 11 Psychometric functions from the multi-cue pitch and acceleration conditions. The *left column* shows data from a subject who saw the ground-plane stimulus and the *right column* shows data from a subject who saw the dot-cloud stimulus. *Upper and lower rows* show data from the pitch and acceleration conditions, respectively. *Each panel* plots the percentage of times the subject indicated that the cues-consistent (comparison) stimulus was more pitched or accelerated than the cues-inconsistent (standard) stimulus as a function of the value of the cues-consistent stimulus. The *curves* are best-fitting cumulative Gaussians. *Each symbol color and the corresponding curve* represent data with a different cues-inconsistent stimulus (see Fig. 8)

sensory signals—visual pitch, visual acceleration, and otoliths—influenced the judgments. Second, the slopes in the pitch data were generally greater for the ground plane than for the dot cloud, which means that visual pitch information was more effective when the stimulus was a ground plane. The difference between the slopes with the ground plane and dot cloud was marginally significant statistically ($F(1,6) = 3.74$, $P = 0.1$).

Now consider the acceleration condition in which subjects indicated in which interval they felt more accelerated (right half of Fig. 12). The judgment could be made from the visual acceleration signal alone, which we will refer to as the direct method. An acceleration judgment could not be made from the visual pitch signal alone or otolith signal alone because those signals are individually ambiguous with respect to acceleration. But the visual pitch and otolith signals could be combined to provide an acceleration estimate; this is the indirect method. If subjects based their acceleration judgments on the direct method only, the PSEs would lie on lines with slope 1 ($\Lambda/\lambda = 1$). If they based judgments on the indirect method only, the PSEs would lie on lines with slope –1 ($\Lambda/\lambda = -1$). There were clear individual differences in the acceleration data, but some trends emerged. First, in the clear majority of cases, the data slopes were between –1 and 1, showing that acceleration judgments were affected by both methods of determining acceleration. We conducted regression analysis on the PSEs: the slopes in Fig. 12 were significantly greater than –1 ($P < 0.001$, one-tailed for the ground plane and dot cloud) and significantly less than +1 ($P < 0.001$, one-tailed for the ground plane and dot cloud). This means that all three sensory signals—visual acceleration, visual pitch, and otoliths—influenced the acceleration judgments. Second, although the slopes in the acceleration data were slightly greater for the ground plane than for the dot cloud, the difference was not statistically significant ($F(1,6) = 0.88$, $P = 0.38$).

Figure 13 shows the PSEs for all subjects in the pitch-plus-acceleration condition. Recall that subjects in this condition made two judgments after each trial: the interval in which they felt more pitched and the one in which they felt more accelerated. The stimulus values associated with the cues-consistent and cues-inconsistent stimuli are depicted in Fig. 8c. The left and right halves of the figure show data from the pitch and acceleration judgments, respectively. If subjects based their judgments on only the direct methods (using visual pitch for pitch judgments and visual acceleration for acceleration judgments), the PSEs would lie on lines with slope 1. If they based their judgments only on the indirect methods (using otolith plus visual acceleration for pitch judgments and otolith plus visual pitch for acceleration judgments), the PSEs would lie on lines with slope –1. The results were similar to those in the pitch and acceleration conditions (Fig. 12). For example, the slopes of the pitch PSEs and the acceleration PSEs were generally between –1 and 1 (although they tended to be steeper in the pitch-plus-acceleration condition than they were in the pitch and acceleration conditions). We conducted regression analyses on the PSEs. For the pitch judgments, the slopes in Fig. 13 were significantly greater than –1 ($P < 0.001$, one-tailed for the ground plane and dot cloud) and significantly less than +1 ($P < 0.001$, one-tailed for the ground plane and $P < 0.05$, one-tailed for the dot cloud). For the acceleration judgments, slopes were significantly greater than –1 ($P < 0.001$, one-tailed for the ground plane and dot cloud) and significantly less than +1 ($P < 0.001$, one-tailed for the ground plane and dot cloud). The fact that slopes were less than +1 means that subjects used the indirect method to make their combined estimates. This must be due to an influence of visual pitch on acceleration judgments and visual acceleration on pitch judgments

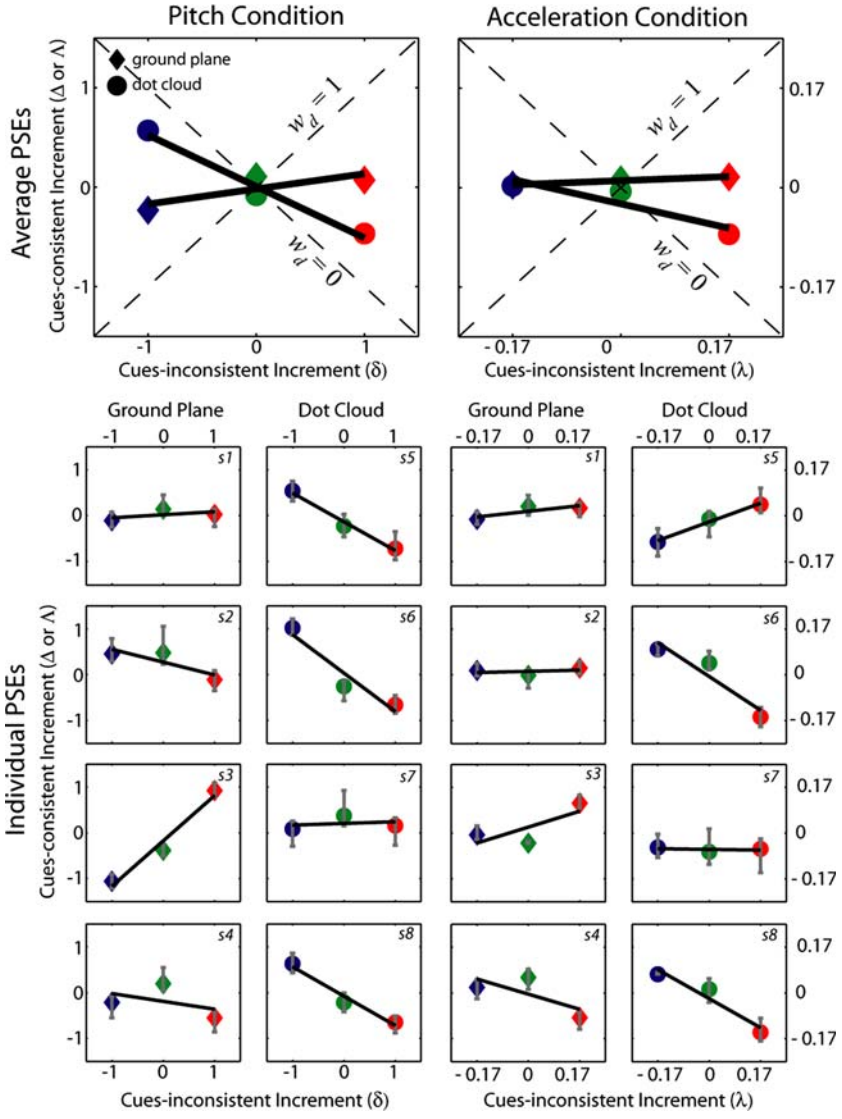

Fig. 12 The PSEs from the multi-cue pitch condition and acceleration condition. The *upper panels* show PSEs averaged across subjects and the *lower panels* show individual PSEs, a separate panel for each subject. The *left panels* show data from the pitch condition and the *right panels* data from the acceleration condition. The *diamonds* represent data with the ground-plane stimulus. The *circles* represent data with the dot cloud. The *slope of the data* indicates the degree to which PSEs were determined by the direct as opposed to the indirect method. A slope of 1 means that subject responses were determined by the direct method only: visual pitch in the pitch condition and visual acceleration in the acceleration condition. A slope of –1 means that responses were determined by the indirect method only. The *lines* are regression fits to the data. *Error bars* are 95% confidence intervals

because the otolith signal (the other component of the indirect method) was always the same (platform pitch = 8°) for the standard and comparison interval. Comparing results from the two visual scenes reveals that slopes for the pitch judgments were very similar for the two scenes ($F(1,6) = 0.01$, $P = 0.93$), while slopes for the acceleration judgments tended to be steeper for the ground plane ($F(1,6) = 2.93$, $P = 0.14$).

It is also interesting to analyze the trial-by-trial correlation between pitch and acceleration responses in this condition. The platform pitch was the same for all presentations in the pitch-plus-acceleration condition; only visual pitch and visual acceleration varied. Therefore, if subjects perceived less pitch in a given presentation, they should also have perceived more acceleration. And if they perceived more pitch, they should have perceived less acceleration. Such an effect occurred in a clear majority of trials. Specifically, pitch and acceleration responses were the same (''more pitch'' and ''more acceleration'', or ''less pitch'' and ''less acceleration'') on 32.5% of the trials and opposite on 67.5 %. This suggests that pitch and acceleration responses were generally inversely related as one would predict from our theoretical analysis.

Calculation of weights based on PSEs

We can conceptualize these relative influences of the direct and indirect methods on the PSEs in the framework of the Bayesian model. We first consider the pitch judgments in the pitch condition. A Bayesian observer would make pitch judgments based on the

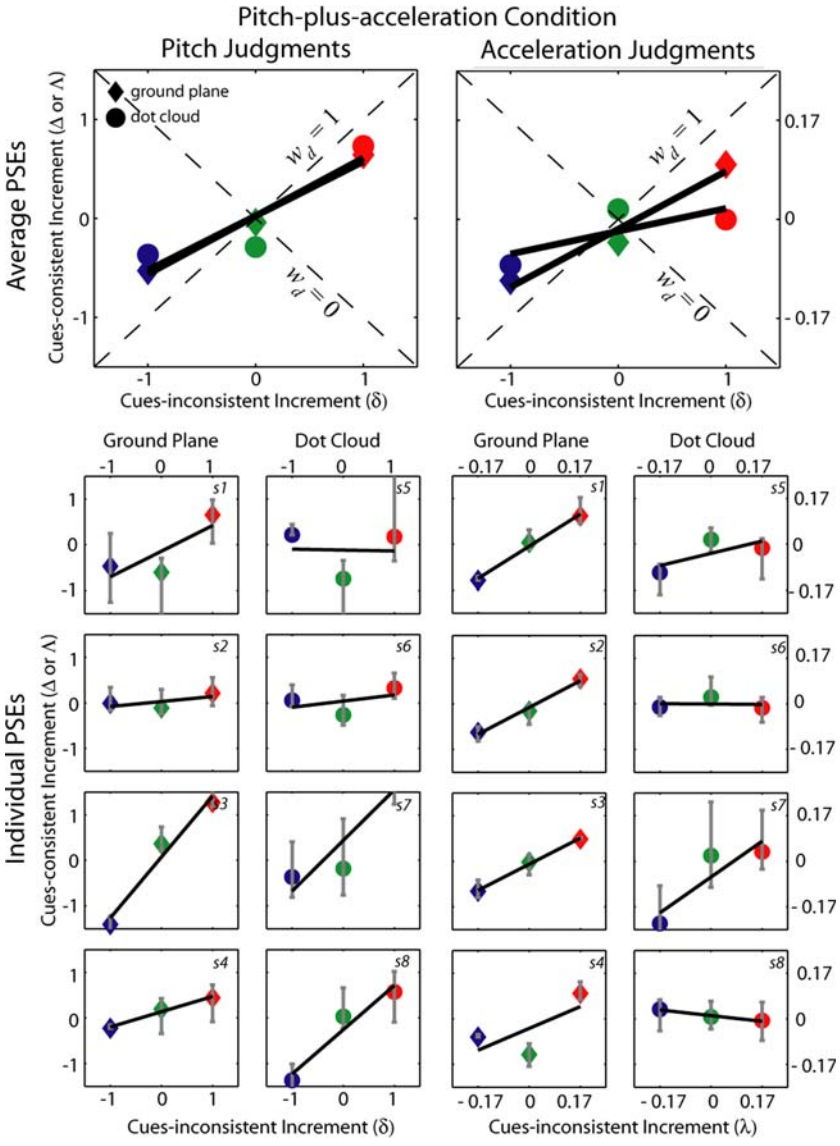

Fig. 13 The PSEs from the multi-cue, pitch-plus-acceleration condition. The *upper panels* show PSEs averaged across subjects and the *lower panels* show individual PSEs. The *left panels and right panels* show data for the pitch and acceleration judgments, respectively. The *diamonds and circles* represent data with the ground-plane and dot-cloud stimulus, respectively. The slope of the data indicates the degree to which PSEs were determined by the direct as opposed to the indirect method. A slope of 1 means that subject responses were determined by the direct method only: visual pitch for the pitch judgments and visual acceleration for the acceleration judgments. A slope of –1 means that responses were determined by the indirect method only. The *lines* are regression fits to the data. *Error bars* are 95% confidence intervals

marginal distributions, which are calculated from the 2d distributions by integrating across the irrelevant acceleration dimension (dashed curves at the bottom of the panels in Fig. 2). Under the assumptions that the likelihoods are conditionally independent and Gaussian distributed, and that the above-mentioned direct and indirect methods are used, the model for pitch judgments can be written as a weighted sum (Ghahramani et al. 1997):

$$\hat{O} = w_{\mathrm{d}} O_{\mathrm{d}} + w_{\mathrm{i}} O_{\mathrm{i}} + w_{\mathrm{p}} O_{\mathrm{p}} \tag{7}$$

where $\hat{O}$ is the estimated pitch, O_{d}, O_{i}, and O_{p} are estimates based on the direct method, indirect method, and idiotropic prior (the no-acceleration prior has a marginal that is uniform in pitch). The weights given to the various estimates are:

$$w_j = \frac{1/\sigma_j^2}{\sum_k 1/\sigma_k^2} \tag{8}$$

In the pitch condition, the cues-consistent stimulus had the same pitch values specified by the direct and indirect methods: they were $O_{\mathrm{d}} = 4{+}\Delta$, $O_i = 4{+}\Delta$. The cues-inconsistent stimulus had different values specified by the two methods: $O_{\mathrm{d}}^* = 4{+}\delta$, $O_{\mathrm{i}}^* = 4{-}\delta$ (the asterisk representing the cues-inconsistent stimulus). PSEs are the cues-consistent stimulus that has a perceived pitch equal to that of the cues-inconsistent stimulus, so $\hat{O} = \hat{O}^*$, and $w_{\mathrm{d}} O_{\mathrm{d}} + w_{\mathrm{i}} O_{\mathrm{i}} + w_{\mathrm{p}} O_{\mathrm{p}} = w_{\mathrm{d}} O_{\mathrm{d}}^* + w_{\mathrm{i}} O_{\mathrm{i}}^* + w_{\mathrm{p}} O_{\mathrm{p}}$. O_{p} is a prior and is therefore the same in the two stimuli. Substituting and rearranging, $\frac{w_{\mathrm{i}}}{w_{\mathrm{d}}} = \frac{\delta - \Delta}{\delta + \Delta}$.

If we make the likely assumption that σ_p^2 is much larger than σ_d^2 and σ_i^2, then $w_i \approx 1 - w_d$ and

$$w_d \approx \frac{1 + \Delta/\delta}{2}. \quad (9)$$

Δ/δ is the equation for the slopes of the regression lines fit to the pitch PSEs in Fig. 12. Thus, we can use these slopes to determine the relative weights given to the direct and indirect methods. Under the assumption that combination is optimal, those relative weights are an index of the relative reliabilities (inverse variances) of the direct and indirect estimates (Ghahramani et al. 1997; Ernst and Banks 2002).

We can apply the same reasoning to the analysis of the PSEs for acceleration judgments in the acceleration condition. The direct method in that case involves the usage of visual acceleration only and the indirect method the combination of the otolith and visual pitch signals. Then,

$$w_d \approx \frac{1 + \Lambda/\lambda}{2}. \quad (10)$$

We can use this equation and the slopes fit to the acceleration PSEs in Fig. 12 to calculate the relative weights given to direct and indirect methods.

Figures 14 and 15 plot the weights derived from the PSE data (Eqs. 9, 10). Note that these weights are calculated directly from the slopes plotted in Figs. 12 and 13; this is essentially a different way of plotting the same data. Therefore, the statistical analyses to test whether the slopes are significantly different from +1 and –1 are equivalent to testing whether the weights are significantly different from 0 or 1. In all cases, average weights are significantly different from 0 and 1. The upper and lower rows of Fig. 14 summarize the findings from the pitch and acceleration conditions, respectively. The direct-method weight w_d is plotted on the left ordinate and the indirect-method weight w_i on the right. If the direct method were the sole determinant of pitch responses, w_d would equal 1; if the direct method had no effect, w_d would equal 0. The average values for w_d in the pitch conditions were 0.58 and 0.24 for the ground plane and dot cloud, respectively. This means that both direct and indirect methods contributed and all three sensory signals were used. The weight for the ground plane was greater than for the dot cloud, showing that the visual pitch cue had greater influence on pitch judgments when the stimulus was a ground plane. The average values for w_d in the acceleration condition were 0.52 and 0.38 for the ground plane and dot cloud, respectively. Again, this means that all three sensory signals contributed to acceleration judgments. The average weights for the ground-plane and dot-cloud stimuli were not significantly different.

Figure 15 plots the weights from the pitch-plus-acceleration condition. The average weights for the direct method w_d were 0.79 (ground plane) and 0.77 (dot cloud) for pitch judgments (top row) and 0.79 (ground plane) and 0.61 (dot cloud) for acceleration judgments (bottom row). These values are somewhat higher than the ones observed in the pitch and acceleration conditions, but some individual tendencies were in general preserved. For example, subjects who gave a signal greater weight in the pitch condition or acceleration condition tended to give it greater weight in the pitch-plus-acceleration condition as well. We speculate that w_d was generally higher in the pitch-plus-acceleration condition because the relative reliability of the direct visual pitch or visual acceleration estimate was greater in this condition. The change in relative reliability is most likely due to decreased reliability of the indirect estimate, rather than an increased reliability of the direct estimate. The direct cue was virtually identical between conditions; visual pitch ranged from 2 to 6° and visual acceleration ranged from 0.17 to 0.68 m/s^2. However, the indirect cue was quite different. In the pitch-plus-acceleration condition, the platform pitch was always 8° while in the pitch and acceleration conditions it varied from 2 to 6°. It is possible that the otolith signal is less reliable at 8° than at 2–6°. Also, the indirect visual cue varied in the pitch-plus-acceleration condition, while it was held constant in the pitch condition (zero visual pitch) and the acceleration condition (zero visual acceleration). It is likely that visual estimates were more reliable when they were constant at zero.

JND analysis

The PSEs show that subjects generally used both the direct and indirect methods to judge pitch and acceleration and this means that they used otolith, visual pitch, and visual acceleration signals. However, this does not prove that subjects combined information from the otolith and visual signals optimally as predicted by the Bayesian model. The PSE shifts could result from cue switching—using one set of signals on some trials and another set on other trials—rather than from using a combined estimate on each trial. Although cue switching can yield PSE data like we observed, it cannot yield an improvement in multi-cue JNDs relative to single-cue JNDs (Ernst and Banks 2002).

Figure 10 plots observed JNDs for the single-cue and the multi-cue conditions in which the conflict was

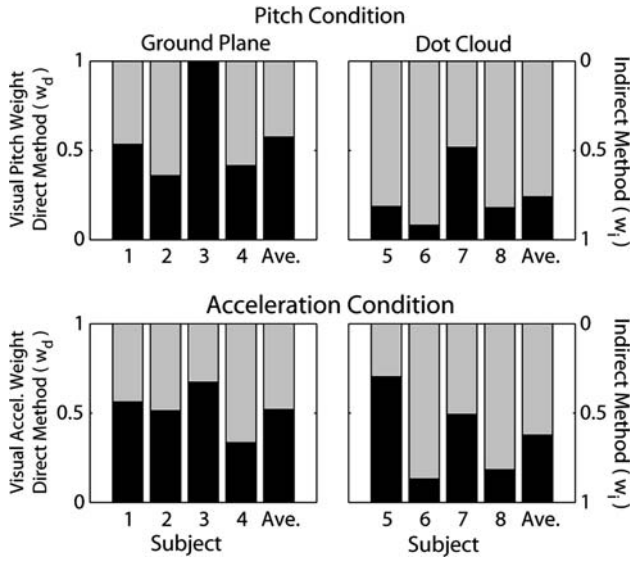

Fig. 14 Weights for the pitch and acceleration conditions. The *upper and lower rows* show the weights for the pitch condition and the acceleration condition, respectively. The *left and right columns* show the weights for the ground-plane and dot-cloud stimuli, respectively. The *horizontal axis* represents different subjects; the average across subjects is on the *right*. The *black bars* represent the weights given to the direct methods (visual pitch for the pitch condition and visual acceleration for the acceleration condition). The *gray bars* represent the weights given to the indirect methods

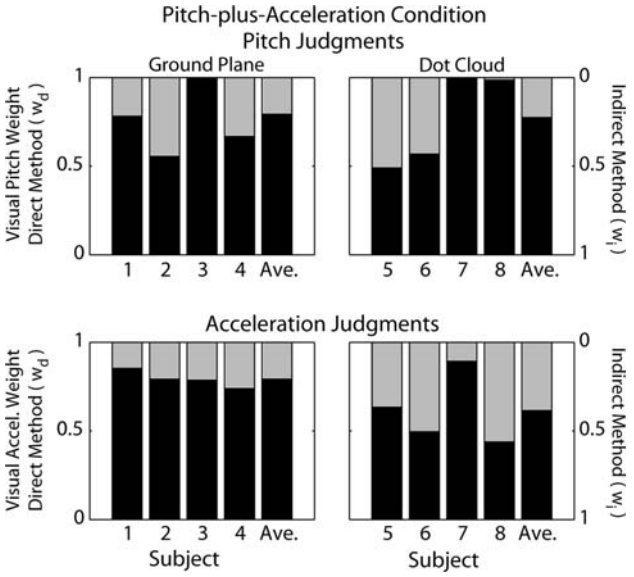

Fig. 15 Weights for the pitch-plus-acceleration condition. The *upper and lower rows* show the weights for the pitch judgments and the acceleration judgments, respectively. The *left and right columns* show the weights for the ground-plane and dot-cloud stimuli, respectively. The *horizontal axis* represents different subjects; the average is on the *right*. The *black bars* represent the weights given to the direct methods (visual pitch for the pitch judgment and visual acceleration for the acceleration judgment). The *gray bars* represent the weights given to the indirect methods

zero ($\delta = \lambda = 0$). We want to know whether the combined JNDs (when vestibular and visual stimuli were both present and informative) were lower than the single-cue JNDs. The multi-cue JNDs were indeed generally lower than single-cue JNDs ($t(15) = 1.85$, $P = 0.042$, one-tailed), which suggests that the cues were being usefully combined and that the PSE effects were not due to cue switching. There are some complexities, however, in comparing the single- and the multi-cue JNDs. Consider first the pitch condition (upper row). The visual JNDs represent discrimination thresholds when the only informative cue was visual pitch, so we assume that they are a measure of the standard deviation of the visual pitch signal. As we said earlier, vestibular signals were present in this condition even though they were constant at zero and irrelevant to the task. Thus, it remains possible, though unlikely, that vestibular signals affected performance in the visual condition and that the measured JND is larger than it would have been if we had truly isolated the visual pitch signal. The non-visual JNDs represent discrimination thresholds when the otoliths provided the only useful information. In this condition, subjects were forced to interpret the otolith signals in the two stimulus intervals in terms of the specified pitch. Thus, the non-visual JNDs should reflect the standard deviation of the otolith signal once converted to a pitch signal under the assumption that no acceleration was present. Any variability in the assumed acceleration would increase the measured non-visual JND. Similar arguments apply to the interpretation of the visual and non-visual acceleration JNDs.

The JND analysis is thus consistent with the idea that vestibular and visual information were being combined in a fashion that improved perceptual precision. But caution is warranted because we cannot be certain that we isolated the underlying signals. We revisit these issues in the Discussion and suggest experiments that could better isolate the signals of interest.

Discussion

Review and evaluation of main findings

It is well recognized that otolith signals are ambiguous with respect to orientation and linear acceleration (Howard, 1982). Many researchers have proposed that disambiguation can be accomplished by combining otolith signals with signals from other sources, such as the semi-circular canals (Merfeld et al. 1999, Angelaki

et al. 1999) and vision (Zupan et al. 2002; Reymond et al. 2002). Our experiment was designed to measure the influence of visual signals on disambiguating otolith signals and we demonstrated several effects.

1. We showed that visual pitch signals affect the perception of pitch. For a given gravitoinertial force on the head, changes in the visual pitch stimulus affected perceived pitch in the direction specified by the visual stimulus ($w_d > 0$ for the pitch condition, Fig. 14, and for pitch judgments in the pitch-plus-acceleration condition, Fig. 15). This result adds to the large literature showing that visual orientation signals affect orientation perception (Howard 1982).
2. We showed that visual pitch signals affect the perception of linear acceleration. For a given gravitoinertial force, changes in the visual pitch stimulus affected perceived acceleration in the expected direction ($w_i > 0$ for acceleration judgments in the pitch-plus-acceleration condition, Fig. 15). We know of only one previous study demonstrating such an effect. Zupan and Merfeld (2003) presented visual roll stimuli while exposing subjects to different gravitoinertial forces. When there was a discrepancy between the measured gravitoinertial force and the visual roll stimulus, subjects' vestibulo-ocular reflex (VOR) eye movements were consistent with lateral acceleration. One cannot conclude from this, however, that subjects perceived lateral acceleration because the orientation-acceleration representation that drives the VOR differs from the representation that drives perception (Merfeld et al. 2005; Zupan and Merfeld 2005). Thus our study provides the first direct evidence that visual pitch signals coupled with ambiguous otolith signals affect perceived acceleration.
3. We showed that visual acceleration signals affect perceived acceleration. Specifically, for a given gravitoinertial force, changes in the visual acceleration stimulus affected perceived linear acceleration in the direction specified by the visual stimulus ($w_d > 0$ for the acceleration condition, Fig. 14, and for acceleration judgments in the pitch-plus-acceleration condition, Fig. 15). This result adds to the large literature showing that optic flow stimuli create linear vection, the sensation of translating through space (Howard 1982). Harris et al. (2000a, 2000b) examined the relative influences of visual acceleration and otolith signals on the perception of distance traveled. They found that the visual stimulus had very little effect and therefore that the vestibular stimulus dominated the percept. We found a substantially larger effect of the visual stimulus on perceived acceleration. It would be interesting to determine the cause of the difference between our results and those of Harris and colleagues.
4. We showed that visual acceleration signals affect perceived pitch. Specifically, for a given gravitoinertial force, changes in the visual acceleration stimulus affected pitch percepts in the predicted direction ($w_i > 0$ for pitch judgments in the pitch-plus-acceleration condition, Fig. 15). To our knowledge, this is the first experimental demonstration of such an effect.
5. We showed that pitch discrimination and acceleration discrimination improved when visual and vestibular information was available as opposed to when only one was available. Specifically, discrimination thresholds were lower in the multi-cue conditions than in the non-visual-only or the visual-only conditions (Fig. 10). We believe that this is the first empirical observation of an improvement in perceptual accuracy when vestibular and visual information are both present.

Previous gravitoinertial-force resolution models

Other researchers have suggested that disambiguation is accomplished by combining otolith signals with information from the semi-circular canals (Merfeld et al. 1999, Angelaki et al. 1999) and vision (Zupan et al. 2002; Reymond et al. 2002); these are known as gravitoinertial-force (GIF) resolution models. Such models propose that internal models are used to represent the relationships between external physical variables like gravitational, inertial, and gravitoinertial force. Here we discuss two previous GIF resolution models and compare them to our approach.

Zupan et al. (2002): sensory weighting model

The model has three stages. The first provides sensory estimates from the raw sensor outputs associated with the otoliths, canals, and visual system. All of these estimates are dynamic. The next stage converts those estimates into a common representation using internal models of body dynamics and physics. The third stage does a weighted average of the converted sensory estimates to generate an estimate based on all of the available sensory signals. The weights are in theory determined by the relative variances of the different sensory estimates expressed in the common

representation. The model has nine free parameters: seven sensory weights and two time constants.

The model's behavior is nicely consistent with human behavior in a variety of situations, so it provides a very useful characterization of the processes that might be involved in visual-vestibular integration. However, the model has a couple of shortcomings in its current form. First, the authors claim that weighted summation provides the most accurate estimate. The minimum-variance solution is indeed a weighted sum when the underlying distributions are Gaussian, but it is not necessarily a weighted sum when they are not (Oruc et al. 2003, Appendix B). Second, Zupan et al. determine the model's nine parameters by fits to the data rather than by determining the parameters independently. (In a later section, we describe how model parameters could be determined independently.)

Reymond et al. (2002): dynamic optimization model

This model works by minimizing four cost functions related to sensory estimates of angular velocity, linear velocity, linear acceleration, and gravity. Sensory estimates are converted into a common representation using internal models. The authors assume that the estimates are unbiased and that they are weighted appropriately to minimize the cost functions. The variability of sensory estimates is not expressly represented, but probably affects the weights assigned to different estimates. The model has 10 free parameters. The best values for fitting the data are found via iterative optimization. The idiotropic vector is not incorporated in the current model, but the authors note that it could be.

The behavior of the Reymond et al. model is quite consistent with human behavior in a variety of situations, but it too has some shortcomings. First, it is unclear whether it offers an optimal solution using multi-sensory information to reduce the uncertainty associated with orientation and acceleration estimates. Second, some aspects of previous experience are not incorporated. Third, the model parameters are determined by fitting the data rather than by independent assessment.

Properties of the Bayesian model of visual-vestibular integration

Percepts can be thought of as the best estimate about what is in the world given the sensory data and previous experience (Helmholtz 1866; Kersten et al. 2004). To make this quantitative, one must specify what is being maximized (or minimized) in obtaining a best estimate, and the way in which the data and previous experience should influence the estimate. Bayes' Law provides a simple framework for optimal solutions to the estimation problem, and we have adopted the framework for the problem of estimating self-orientation and -acceleration. The relationship between observed sensory data and orientation/acceleration is specified by the likelihood functions. Information from previous experience is specified by the priors. Orientation/acceleration estimates based on the posterior distribution yield minimum-variance estimates. Here we discuss some properties of the model's components for the particular problem of estimating pitch and acceleration from vestibular and visual signals. Then we discuss how the model's parameters could be quantified, how dynamics could in principle be added, and how the model accounts for some well-known effects.

Properties of the likelihood functions

Each likelihood function in Fig. 2 represents the probability of the observed sensory signal given each possible pitch and inertial acceleration. As in Zupan et al. (2002) and Reymond et al. (2002), one has to convert the signals from the system in which their measurements are made into a common representation.

Otolith afferents are driven by gravitoinertial force. To convert otolith signals into pitch-acceleration space, we assume that the nervous system has the appropriate mapping; in other words, we assume it has an internal model of the gravitoinertial force equation. There are two degrees of freedom in the measurement of a force vector and they can be described as a direction and a magnitude. Equation 3 shows how the otolith signals expressed as a direction and magnitude can be converted into signals indicating orientation and inertial acceleration. In Fig. 3, we showed some examples of how different otolith signals affect the likelihood in pitch-acceleration space.

For brief angular accelerations, afferent responses from the semi-circular canals signal angular velocity (Goldberg and Fernandez 1971). Conversion into a pitch estimate requires integration of the velocity estimate and an estimate of absolute orientation at some moment in time. Thus, variability in the canal likelihood is due to variability in the sensory signal and in the processes involved in the signal conversion. Obviously, the canal likelihood depends significantly on the situation. For example, when the head is rotated slowly, canal afferents respond minimally and variably so the variance of the canal likelihood would increase with decreasing angular velocity. At the slow rotation rates used in our experiment, the canal signals were

presumably minimally informative. The fact that subjects, when pitched, perceived forward acceleration is consistent with this assumption: if the canals had provided useful pitch signals, subjects would have experienced little or no acceleration and there would have been little effect of manipulating visual information.

Both static and dynamic visual orientation information is relevant to the visual pitch likelihood (Howard and Childerson 1994; Dichgans et al. 1972). Static orientation cues are features of known orientation in world coordinates, such as the horizon on flat terrain. For the horizon to be useful in judging head orientation, the visual system must estimate its orientation on the retina and use eye-position information to convert it into head coordinates. Dynamic orientation cues come from the optic flow measured at the retina. Assuming the retinal-image motion is caused by self-motion relative to a rigid scene, rotational flow provides an estimate of angular velocity in retinal coordinates that can be converted to head coordinates by using eye-position signals and then integrating over time to yield an estimate of head orientation. In our experiment, static and dynamic visual orientation cues were always consistent with one another, so the visual-pitch likelihood function represents both types of information. As with the vestibular likelihoods, the variance of the visual-pitch likelihood depends significantly on the type of stimulus presented.

The visual-acceleration likelihood is derived from sensory estimates of optic flow. Equation 5 show how retinal-image motion can be converted into an estimate of head acceleration. As we said before, the conversion requires an estimate of distance in order to overcome the scale ambiguity of optic flow, so an error in the distance estimate will cause error in the velocity and acceleration estimates. Thus, variability in measuring the retinal motion or in scaling will cause variability in the acceleration likelihood. Again the variance of the visual-acceleration likelihood will depend significantly on the type of visual stimulus presented.

Properties of the priors

A recent Bayesian model developed by Weiss et al. (2002) shows how knowledge from previous experience, expressed as priors, can affect perception. They proposed a prior for slow speed and showed how it can help explain the fact that stimuli appear to move slower at low than at high contrast (Stone and Thompson 1992) and that a line seen through an aperture appears to move in the direction orthogonal to the line's orientation (Adelson and Movshon 1982).

There are two priors in our model: the idiotropic and no-acceleration priors. The former, which is closely related to Mittelstaedt's idiotropic vector, represents the fact that the head and body are usually upright with respect to gravity. It is a bias to perceive gravity as aligned with the body's longitudinal axis. Mittelstaedt hypothesized that perceived orientation is the weighted sum of the idiotropic vector and a gravitational vector derived from otolith and visual information. In our model, the idiotropic vector is a prior centered at a pitch of zero. It is simply another piece of information that has more or less effect on perceived orientation depending on the reliability of other available information.

To our knowledge, the no-acceleration prior has not been previously suggested in the visual-vestibular literature. It represents the fact that zero inertial acceleration is the most common situation for an earth-bound observer. The no-acceleration prior is similar in nature to the slow-speed prior described by Weiss et al. (2002) and the stationarity principle described by Wexler and colleagues (Wexler et al. 2001; van Boxtel et al. 2003).

The model's predictions for zero gravity, the Aubert effect, and the somatogravic illusion

In the Bayesian framework, perceptual illusions are not viewed as sensory errors but rather as byproducts of computations that are generally useful (Weiss et al. 2002). In Fig. 16 we show how the model can account for two illusions—the Aubert effect in the perception of subjective vertical (Howard 1982) and the somatogravic illusion (Glasauer 1995). We also consider how the model explains the perception of vertical in zero-gravity environments (Glasauer and Mittelstaedt 1998).

The left column shows the analysis for the perception of self-orientation in a zero-gravity environment with no visual stimulation. The otoliths signal no acceleration. For an observer on earth, the absence of a force on the otoliths must be due to an inertial acceleration that is equal and opposite to gravity, such as during free fall. Hence the likelihood in orientation-acceleration space is a horizontal ridge at an acceleration of 9.81 m/s^2 (top panel). Because there is no visual stimulus, the visual likelihoods (not shown) are uniform and do not affect the posterior. The observer is not rotating so the canal likelihood (not shown) is centered at zero pitch and presumably has high variance and very little effect on the posterior. The idiotropic and no-acceleration priors are represented by

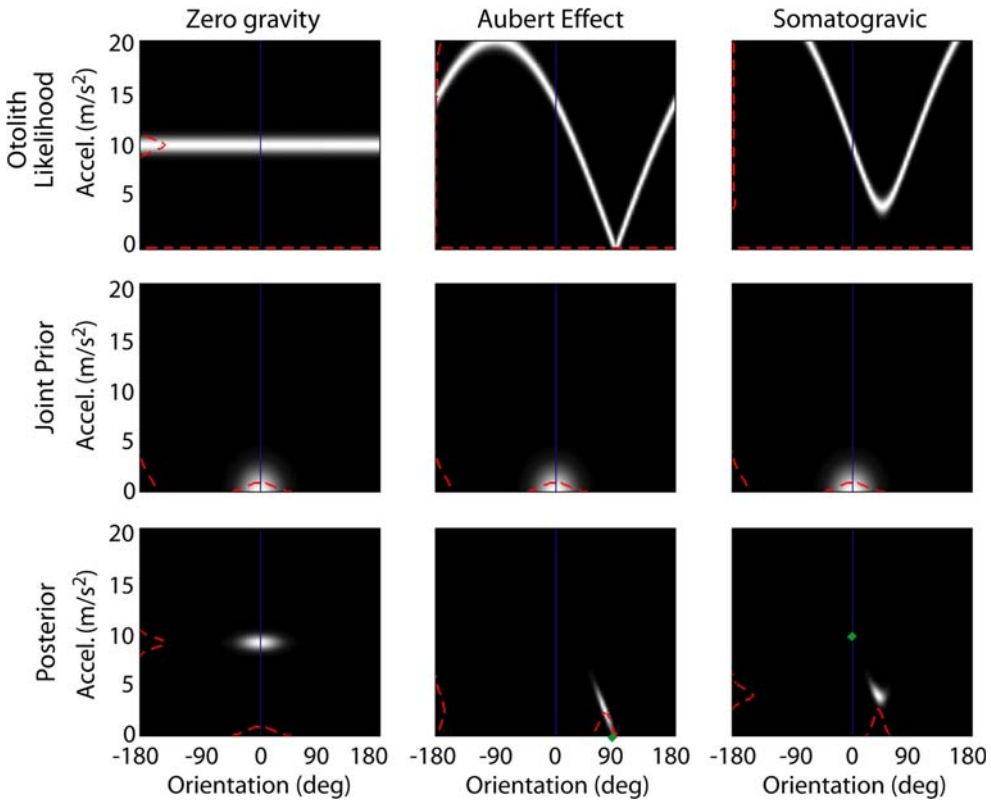

Fig. 16 Analysis of three effects in the perception of self-orientation. The *left, middle, and right columns* show the analyses for perceiving self-orientation in zero gravity, in the conditions that create the Aubert effect, and in the conditions that create the somatogravic illusion. The *top row* represents the otolith likelihoods. The *middle row* represents the product of the idiotropic and no-acceleration priors, under the assumption that they have equal variances. The *bottom row* represents the posteriors; here they are the product of the otolith likelihood and the joint prior because there is no visual stimulus and the canal likelihood is assumed to be very broad and therefore would have negligible effect on the posterior. The marginal probability distributions are represented by *red dashed curves*. The actual orientation (and acceleration) of the observer is represented by a *small green diamond* in the *middle and right posterior panels*. There is no veridical orientation in zero-gravity, so there is no green diamond in the *right posterior panel*

the product of the two, which is isotropic if they have equal variance (middle panel). The posterior is shown in the bottom panel. It is centered at a positive acceleration and zero pitch. Hence the observer should report a pitch of 0°, which is consistent with most reports (Glasauer and Mittelstaedt 1998).

The middle column shows the analysis for the perception of vertical when an earth-bound observer's head and body are rolled. We know that observers tend to perceive subjective vertical (the direction of perceived gravity) as less rolled in head coordinates than is actually the case; this is the Aubert effect (Howard, 1982). The otolith likelihood in this case is like the one shown in the top panel. For a roll of 90°, the minimum in the otolith likelihood is at 90°. The visual and canal likelihoods are not shown because we are modeling the case in which there is no useful visual information and no on-going rotation of the head. Thus, the posterior is the product of the otolith likelihood and the joint prior. As you can see, the maximum of the posterior occurs at a roll less than the observer's actual roll, as is observed in the Aubert effect.[9]

The right column shows the analysis for the somatogravic effect (Gillingham and Previc 1993; Glasauer 1995). The observer is upright with respect to earth and forward accelerated. Therefore, the otolith likelihood is shifted along the pitch axis and has an acceleration minimum at a non-zero pitch value. We are modeling the situation in which visual stimuli are uninformative and no head rotation occurred, so the visual and canal likelihoods have no influence on the posterior. The maximum of the posterior occurs at a pitch value greater than zero, as is observed with the somatogravic

[9] Note that the posterior is centered on non-zero inertial acceleration, which means that the model predicts a perceived inertial acceleration that is slightly greater than zero. To our knowledge, no one has examined whether the conditions that produce the Aubert effect also cause a perceived acceleration.

effect. It is important to note that one would not predict a shift from non-zero pitch without the no-acceleration prior; if the no-acceleration prior is omitted from the analysis, the posterior becomes the product of the otolith likelihood (V-shaped) and the idiotropic vector (vertical ridge), so the posterior's peak would always be centered on a pitch of 0°. Thus the fact that the somatogravic illusion occurs is evidence for the no-acceleration prior.

These three examples show that the Bayesian model can explain some interesting phenomena in self-orientation perception. But the examples are somewhat unsatisfying because the ability to predict the effects depends on the assumed variances of the likelihoods and priors and we have no independent means for knowing those values. In the next section, we consider how one might measure them independently.

Estimating the model's parameters

If we used the Bayesian model in its current form to predict the data quantitatively, we would have to treat the likelihood and prior parameters as free parameters, as Zupan et al. (2002) and Reymond et al. (2002) did. It is in principle possible, however, to estimate the parameters independently and thereby avoid the need to use free parameters. Ernst and Banks (2002) did this for visual and haptic signals for estimating the size of an object. They used single-modality experiments—vision-only and haptic-only—to estimate the parameters of the visual and haptic likelihood functions. If variation in the sensory estimates is the only cause of variation in observer's responses, the slope of the psychometric function is proportional to the standard deviation of the underlying sensory estimator. For 2-alternative data well fit by a cumulative Gaussian, $\sigma_c = \mathrm{JND}/\sqrt{2}$ where σ_c is the standard deviation of the underlying estimator, and JND is the stimulus increment associated with 84% correct. In using the slopes of the psychometric functions to determine the standard deviations of the underlying estimators, Ernst and Banks had to make three assumptions: (1) the prior distribution for size is essentially uniform across the range of sizes presented in the experiment (which was reasonable in their case), (2) high-level decision noise did not affect the subjects' responses (evidence supporting that assumption for 2-alternative, forced-choice tasks with well-trained subjects is provided by Knill and Saunders, 2003; Hillis et al. 2004), and (3) the sensory estimators are unbiased, meaning that their signal values by themselves are interpreted correctly (this assumption is also reasonable because sensory systems become calibrated during the lifespan, Malcolm and Melvill-Jones 1970). With those assumptions, Ernst and Banks made PSE and JND predictions for two-modality experiments. Because the model parameters had been estimated from the single-modality experiments, the predictions had no degrees of freedom. They then compared the predicted and empirical PSEs and the predicted and empirical JNDs and found that they were quite similar. From this they concluded that visual-haptic integration is essentially statistically optimal and well described by Bayes' Law. More recently, Alais and Burr (2004) and Roach et al. (2006) used similar procedures and obtained similar results for visual-auditory integration. Interestingly, Roach and co-workers documented a robustness effect that occurs when the visual and auditory signals are quite different from one another (with large conflicts, responses were driven by one signal rather than a combination of the two). They showed that the result is consistent with Bayesian estimation if the likelihood functions have longer tails than Gaussians.

As a side note, we point out that the experimental techniques used in previous vestibular-visual experiments do not readily allow one to estimate internal model parameters. Most of the work has used eye movements or estimates of perceptual magnitudes and both of these approaches are problematic. For example, many of the eye-movement studies used the VOR to investigate the underlying representation of self-orientation and -acceleration. The problem is that the VOR is determined by both the internal representation and the production of the eye movement itself. An observed effect, therefore, might reflect the representation, the movement production, or both. Indeed, Merfeld et al. (2005) and Zupan and Merfeld (2005) have shown that the VOR and perceived orientation/acceleration can differ substantially. Other studies have assessed perception by using estimation procedures. For example, Zupan and Merfeld (2003) had subjects indicate perceived roll by setting a hand-held rod. The problem with this approach is that the settings do not necessarily reflect the sensory representation because there is an unknown mapping function between that representation and the subject's settings. We argue that 2-alternative, forced-choice psychophysical procedures, as employed by Ernst and Banks (2002), are most able to yield the required information. In the remainder of this section, we show how such procedures could in principle be used to specify the model parameters.

For vestibular-visual interactions like those considered in the current paper, there are three hurdles to measuring the likelihood distributions for the signals of interest. Here we describe the hurdles and how one might overcome them.

1. The likelihoods (and priors) must be expressed in the same units; said another way, the sensory signals must be promoted to a common representation (Zupan et al. 2002; Landy et al. 1995). But the units in which the behavioral measurements should be made differ from one signal to another. The visual-pitch and visual-acceleration likelihood distributions can be estimated from measurements of pitch and acceleration discrimination, but the canal and otolith likelihood distributions cannot be estimated in this way because their signals (considered separately) do not unambiguously specify pitch and acceleration.

To circumvent this problem for the otoliths, one would measure the likelihood in units of forward and vertical acceleration (i.e., in the plane defined by the gravity and inertial vectors as displayed in Fig. 1c). The idea is depicted in Fig. 17a: The Z and Y axes represent forward and vertical acceleration relative to the head. There is a base stimulus at (Z_{base},Y_{base}). On each trial, two stimuli would be presented successively in the dark. The stimuli would either be the base stimulus presented twice, or the base stimulus once and the base plus an increment once. The increment would be added to the base stimulus in a given direction in Z–Y space. The magnitude of the increment would be varied to find the just-discriminable value. The threshold measurements would be made for each of a number of directions with respect to the base stimulus (arrows in Fig. 17). Subjects would indicate after each trial whether the two stimuli were the same or different. From the set of thresholds, one would estimate the contour around the base value that is just discriminable. Figure 17b shows an example of such a threshold contour. To the degree that the experiment isolates the otolith signals (see below how one can separate the effects of sensory signals and priors), the experimental findings would allow an estimate of the otolith likelihood for a given base value. One would, naturally, want to conduct the experiment for different base values. To represent the otolith likelihood in the same units as the other likelihoods, one would then transform the threshold contour into pitch-acceleration space as shown in Fig. 3. This procedure would yield an estimate of the otolith likelihood function in units of pitch and acceleration.

A similar procedure could be used to generate the canal likelihood. The canal threshold could be measured like the otolith threshold, except the stimuli and thresholds would be in units of angular velocity. Two rotations would be presented on each trial and subjects would indicate whether the two stimuli were the same or different. Rotations would have to be around the earth-vertical axis in order to isolate the canal signals, but the subject's head could be oriented differently with respect to gravity to measure the thresholds for different head-centric axes. Such an experiment would provide a good estimate of the variability associated with the canal signal. To represent the canal likelihood in the same units as the other likelihoods, one would have to convert the canal signal into an orientation signal: It would have to be integrated over time and the resulting change in orientation assessed relative to an initial orientation estimate. The variability of the canal likelihood would thus also be affected by the variability of initial estimate, and variability in the integration. It is not clear how one would go about assessing these additional sources of variability.

2. The visual-pitch and visual-acceleration likelihoods might be somewhat difficult to isolate because signals from the vestibular system cannot be simply turned off. Consider, for example, a visual-pitch discrimination experiment in which an upright subject made discriminations based on changes in visual pitch. The visual signals might not be isolated with this procedure because gravitational force indicating stationary, upright orientation would also be present. If subjects could completely ignore the irrelevant, constant vestibular signals, the psychometric function would be a valid indicator of the variability of the visual signals. But if subjects failed to ignore the vestibular signals, the variability associated with those signals would inflate discrimination threshold. One could examine this by comparing the performance of

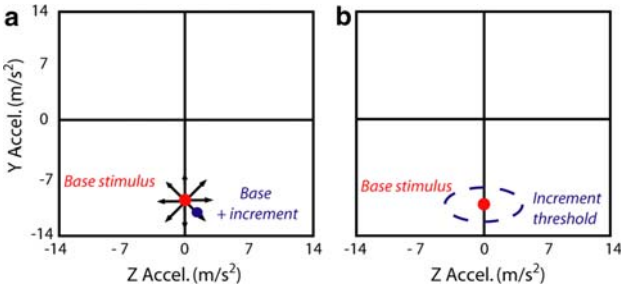

Fig. 17 Method for determining the otolith likelihood empirically. **a** Acceleration stimuli plotted in head-centric coordinates (mid-sagittal, Z–Y plane). A base stimulus at (0 m/s^2, –9.81 m/s^2) represents an upright observer who is stationary in earth coordinates. Another base + increment stimulus contains a different acceleration than the base stimulus. The subject is asked to discriminate the base and base + increment stimuli. The *arrows* represent different directions in which the increment could be added to the base. The magnitude of the increment would be varied in each direction to find the discrimination threshold. **b** The threshold value in each direction would be plotted relative to the base stimulus to define a threshold contour

normal subjects to the performance of subjects with significant vestibular deficits. If visual discrimination performance was on average the same in the two types of subjects, it would suggest that a visual-pitch discrimination experiment can in fact isolate the visual signals. If the visual discrimination performance of normals was poorer than that of other subjects, it would suggest that the normals were unable to ignore the irrelevant signals and therefore that we were unable to isolate the visual signals. One could also compare performance in normal observers on earth and in zero gravity.

3. Non-uniform priors like the ones we propose will influence perceptual judgments. Thus, the JNDs measured in single-modality measurements are affected by both the likelihood of interest and the relevant prior. To measure the prior distributions in the current case, one needs a way to separate their effects from those of the likelihoods. Stocker and Simoncelli (2006) developed a procedure to do just that. Under the reasonable assumption that the prior has higher variance than the likelihood, they take advantage of the fact that PSEs are then more affected by the prior than are JNDs. Their method allows one to estimate the variance of the likelihood and the shape of the prior distribution. The method could be adopted for the measurement of the vestibular and visual likelihoods and the idiotropic and no-acceleration priors.

Frequency dependence of resolution of gravity and acceleration

There are several significant time-dependent effects in visual-vestibular integration. For example, slow changes in otolith stimulation are typically interpreted as changes in orientation while fast changes in stimulation are interpreted as linear acceleration (Glasauer, 1995; Seidman et al. 1998). We have not considered dynamics in this paper, but the model could in principle be extended in that way. The likelihood functions would be described over time to represent time-dependent changes in vestibular and visual signals. The accumulation of information could be represented by the incorporation of a Kalman filter (Maybeck 1979).

The somatogravic illusion and aviation

In aviation, the somatogravic illusion generally occurs only when visibility is poor, such as flying at night or in a cloud (Gillingham and Previc 1993; Cheung et al. 1995). This suggests that disambiguating visual information overrides the illusion, an effect that is well explained by the model presented here.

Several researchers have investigated the role of visual signals in experimental settings. Some have reported that visual signals specifying forward acceleration with the head upright override the illusion (Tokumaru et al. 1998; Lessard et al. 2000), while others have reported no effect of such visual signals (Previc et al. 1992; Otakeno et al. 2002). It is instructive to consider the studies showing no effect of visual stimulation. Previc et al. (1992) varied the direction of gravitoinertial force by using a centrifuge. Visual cues represented earth horizontal. Subjects indicated the perceived direction of earth vertical when the gravitoinertial force (when misperceived as gravitational force) indicated that they were pitched ''nose up'' and visual cues indicated that they were not. All subjects pointed in the direction indicated by the gravitoinertial force; i.e., they experienced the somatogravic illusion. Their responses were the same in the dark and with visual cues present, so there was no effect of visual information. We believe that the properties of the visual display probably made the visual information in the Previc et al. study an uncertain indicator of earth horizontal. The display was head-mounted, so as the subject's head moved relative to the centrifuge, the visual stimulus did not move relative to the subject as it would have if the stimulus were really the earth. Consequently, subjects could probably ascertain that the display was attached to the head and was not representative of earth horizontal and so it was probably ignored as an orientation cue. In a very similar experiment, Lessard and colleagues (2000) reported a clear effect of visual information (in the more experienced subjects). Significantly, the visual display was mounted to the centrifuge rather than the head so its movement relative to the head would have been more appropriate for an earth-referenced stimulus.

A better understanding of the kinds of visual stimulation that affect the interpretation of vestibular signals could lead to the creation of more useful visual indicators of aircraft attitude in the cockpit.

Conclusion

Signals from the otoliths are ambiguous indicators of self-orientation and acceleration, so other sensory signals and previous experience are needed to resolve the ambiguity. We presented a Bayesian model of how to combine noisy vestibular and visual signals and prior information. Likelihoods associated with sensory measurements were represented in an orientation/acceleration space. Two priors—the idiotropic prior and the no-acceleration prior—were also represented

in this space. The model shows how in the Bayesian framework visual cues to self-orientation and self-motion can disambiguate the otolith signal. We also conducted an experiment using a motion platform and attached visual display to examine the influence of visual signals on the interpretation of the otolith signal. Subjects made pitch and acceleration judgments as vestibular and visual signals were manipulated independently. Predictions of the model were confirmed: (1) visual signals affect the interpretation of the otolith signal, (2) less variable signals have more influence on perceived orientation and acceleration than more variable ones, and (3) combined estimates are more precise than single-cue estimates. We also showed how the model can predict a variety of well-known vestibular-visual effects.

Acknowledgments This research was supported by NIH training grant (EY14194) to the Berkeley Vision Science program and AFOSR research grant (F49620) to Martin Banks. Thanks to MarcErnst for helpful discussion and the MPI workshop for techincal assistance. Special thanks to three reviewers who provided thoughtful and thorough comments on the manuscript.

References

Adelson EH, Movshon JA (1982) Phenomenal coherence of moving visual patterns. Nature 300:523–525

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262

Angelaki DE, McHenry MQ, Dickman JD, Newlands SD, Hess BJ (1999) Computation of inertial motion: neural strategies to resolve ambiguous otolith information. J Neurosci 19:316–327

Angelaki DE, Shiakh AG, Green AM, Dickman JD (2004) Neurons compute internal models of the physical laws of motion. Nature 430:560–564

Berger DR (2003) Spectral texturing for real-time applications. In: Siggraph 2003 sketches and applications. ACM Press, New York. DOI 10.1145/095400.965509

Benson AJ, Hutt ECB, Brown SF (1989) Thresholds for the perception of whole body angular movement about a vertical axis. Aviat Space Environ Med 60:205–213

Cheung B, Money K, Wright H, Bateman W (1995) Spatial disorientation - implicated accidents in Canadian forces, 1982–1992. Aviat Space Environ Med 66(6):579–585

Dichgans J, Held R, Young LR, Brandt T (1972) Moving visual scenes influence the apparent direction of gravity. Science 178:1217–1219

Einstein A (1907) Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen. Jahrbuch der Radioaktivität und Elektronik 4:411–462

Ercoline WR (1997) Classification of the USAF SD mishaps. In: Braithwaite MG, DeRoche SL, Alvarez EA, Reese MA (eds) Proceedings of the first tri-service conference on rotary-wing spatial disorientation: spatial disorientation in the operational rotary-wing environment. USAARL report no. 97-15

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433

Gepshtein S, Banks MS (2003) Viewing geometry determines how vision and haptics combine in size perception. Curr Biol 13:483–488

Ghahramani Z, Wolpert DM, Jordan MI (1997) Computational models of sensorimotor integration. In: Morasso PG, Sanguineti V (eds) Self-organization, computational maps, and motor control. Elsevier, Amsterdam, pp 117–147

Gibson JJ (1950) The perception of the visual world. Houghton-Miffin, Boston

Gibson JJ (1966) The senses considered as perceptual systems. Houghton-Mifflin, Boston

Gillingham KK, Previc FH (1993) Spatial orientation in flight. AL-TR-1993–0022. Brooks Air Force Base, Armstrong Laboratory, Texas

Glasauer S (1995) Linear acceleration perception: frequency dependence of the hilltop illusion. Acta Otolaryngol Suppl 520:37–40

Glasauer S, Mittelstaedt H (1998) Perception of spatial orientation in microgravity. Brain Res Brain Res Rev 28(1–2):185–193

Goldberg J, Fernandez C (1971) Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. I. Resting discharge and response to constant angular accelerations. J Neurophysiol 34:635–660

Groen EL, Valenti Clari MSV, Hosman RJAW (2001) Evaluation of perceived motion during a simulated takeoff run. J Aircraft 38(4):600–606

Harris LR, Jenkin M, Zikovitz DC (2000a) Vestibular capture of the perceived distance of passive linear self motion. Arch Ital Biol 138(1):63–7

Harris LR, Jenkin M, Zikovitz DC (2000b) Visual and non-visual cues in the perception of linear self motion. Exp Brain Res 135:12–21

Helmholtz H (1866) Treatise on physiological optics. Ed. trans. JPC Southall. Thoemmes Press, Bristol, 2000

Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: mandatory fusion within, but not between, senses. Science 298:1627–1630

Howard IP (1982) Human visual orientation. Wiley, New York

Howard IP, Childerson L (1994) The contribution of motion, the visual frame, and visual polarity to sensations of body tilt. Perception 23:753–762

Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Ann Rev Psychol 55:271–304

Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? Vis Res 43:2539–2558

Körding KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427(6971):244–247

Landy MS, Kojima H (2001) Ideal cue combination for localizing texture-defined edges. J Opt Soc Am A Opt Image Sci Vis 18:2307–2320

Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. Vis Res 35:389–412

Lessard CS, Matthews R, Yauch D (2000) Effects of rotation on somatogravic illusions. IEEE Eng Med Biol Mag 19:59–65

Longuet-Higgins HC, Prazdny K (1980) The interpretation of a moving retinal image. Proc R Soc Lond B 208:385–397

Malcolm R, Melvill-Jones G (1970) A quantitative study of vestibular adaptation in humans. Acta Otolaryngol 70:126–135

Maybeck PS (1979) Stochastic models, estimation and control. Academic, NewYork

Merfeld DM, Zupan L, Peterka RJ (1999) Humans use internal models to estimate gravity and linear acceleration. Nature 398:615–618

Merfeld DM, Zupan LH, Gifford CA (2001) Neural processing of gravito-inertial cues in humans. II. Influence of the semicircular canals during eccentric rotation. J Neurophysiol 85(4):1648–1660

Merfeld DM, Park S, Gianna-Poulin C, Black FO, Wood S (2005) Vestibular perception and action employ qualitatively different mechanisms. I. Frequency response of VOR and perceptual response during translation and tilt. J Neurophysiol 94:186–198

Mittelstaedt H (1983) A new solution to the problem of the subjective vertical. Naturwissenschaften 70(6):272–281

Mittelstaedt ML, Mittelstaedt H (2001) Idiothetic navigation in humans: estimation of path length. Exp Brain Res 139:318–332

Oruç I, Maloney LT, Landy MS (2003) Weighted linear cue combination with possibly correlated error. Vis Res 43:2451–2468

Otakeno S, Matthews RSJ, Folio L, Previc FH, Lessard CS (2002) The effects of visual scenes on roll and pitch thresholds in pilots versus nonpilots. Aviat Space Environ Med 73:98–101

Previc FH, Varner DC, Gillingham KK (1992) Visual scene effects on the somatogravic illusion. Aviat Space Environ Med 63:1060–1064

Reymond G, Kemeny A (2000) Motion cueing in the Renault driving simulator. Vehicle Syst Dynam 34(4):249–259

Reymond G, Droulez J, Kemeny A (2002) Visuovestibular perception of self-motion modeled as a dynamic optimization process. Biol Cybern 87(4):301–314

Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. Proc R Soc B Biol Sci 273(1598):2159–2168

Royden CS, Crowell JA, Banks MS (1994) Estimating heading during eye movements. Vis Res 34(23):3197–3214

Seidman SH, Telford L, Paige GD (1998) Tilt perception during dynamic linear acceleration. Exp Brain Res 119(3):307–314

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9(4):578–585

Stone LS, Thompson P (1992) Human speed perception is contrast dependent. Vis Res 32(8):1535–1549

Tokumaru O, Kaida K, Ashida H, Mizumoto C, Tatsuno J (1998) Visual influence on magnitude of somatogravic illusion evoked on spatial disorientation demonstrator. Aviat Space Eviron Med 69:111–116

van Boxtel JJ, Wexler M, Droulez J (2003) Perception of plane orientation from self-generated and passively observed optic flow. J Vis 3(5):318–332

Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598–604

Wexler M, Panerai F, Lamouret I, Droulez J (2001) Self-motion and the perception of stationary objects. Nature 409(6816):85–88

Wichmann FA, Hill NJ (2001) The psychometric function: I. Fitting, sampling, and goodness-of-fit. Perc Psychophys 63:1293–1313

Yuille AL, Bülthoff HH (1996) Bayesian decision theory and psychophysics. In: Richards W, Knill DC (eds) Perception as Bayesian inference. Cambridge University Press, Cambridge, pp 123–162

Zupan LH, Merfeld DM (2003) Neural processing of gravito-inertial cues in humans. IV. Influence of visual rotational cues during roll optokinetic stimuli. J Neurophysiol 89(1):390–400

Zupan LH, Merfeld DM (2005) Human ocular torsion and perceived roll responses to linear acceleration. J Vestib Res 15(4):173–183

Zupan LH, Merfeld DM, Darlot C (2002) Using sensory weighting to model the influence of canal, otolith and visual cues on spatial orientation and eye movements. Biol Cybern 86(3):209–230