



Integration of disparity and velocity information for haptic and perceptual judgments of object depth

Rachel Foster^{a,b}, Carlo Fantoni^b, Corrado Caudek^{b,c}, Fulvio Domini^{a,b,*}

^a Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI, USA

^b Center for Neurosciences and Cognitive Systems, Italian Institute of Technology, Rovereto, Italy

^c Department of Psychology, Università degli Studi di Firenze, Firenze, Italy

ARTICLE INFO

Article history:

Received 16 March 2010

Received in revised form 9 December 2010

Accepted 10 December 2010

Available online 14 January 2011

PsycINFO codes:

2330 Motor Processes

2323 Visual Perception

Keywords:

Prehension

Motor control

Somatosensory

Haptic

Visual feedback

Shape perception

Cue integration

ABSTRACT

Do reach-to-grasp (prehension) movements require a *metric representation* of three-dimensional (3D) layouts and objects? We propose a model relying only on *direct* sensory information to account for the planning and execution of prehension movements in the absence of haptic feedback and when the hand is not visible. In the present investigation, we isolate relative motion and binocular disparity information from other depth cues and we study their efficacy for reach-to-grasp movements and visual judgments. We show that (i) the amplitude of the grasp increases when relative motion is added to binocular disparity information, even if depth from disparity information is already veridical, and (ii) similar distortions of derived depth are found for haptic tasks and perceptual judgments. With a quantitative test, we demonstrate that our results are consistent with the *Intrinsic Constraint* model and do not require 3D metric inferences (Domini, Caudek, & Tassinari, 2006). By contrast, the linear cue integration model (Landy, Maloney, Johnston, & Young, 1995) cannot explain the present results, even if the flatness cues are taken into account.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

It is commonly believed that visually guided behavior relies on a three-dimensional (3D) *metric* representation of the environment and the objects in it (Glover, 2004; Greenwald & Knill, 2009). It is also believed that this 3D depth map is found by reversing the physics of image generation to infer the outside world from sensory data (Helmholtz, 1867/1962; Landy, Maloney, Johnston, & Young, 1995; Landy, Banks, & Knill, in press; Poggio, Torre, & Koch, 1985). The solution of the so-called “inverse-optics” problem by a biological system, however, is extremely difficult because of the underdetermination of the required information. Horizontal binocular disparities, for instance, are not sufficient to recover an object’s depth unless the viewing distance is known (Mayhew & Longuet-Higgins, 1982; Fantoni, 2008). Similarly, optic flow is not sufficient to recover surface slant unless additional parameters are known (i.e., the angular displacement between the observer and the surface and the amount of surface rotation) — see Fantoni, Caudek, and Domini (2010).

Moreover, even sufficient constraints provided by multiple cues do not guarantee unique percepts (Todd, 2004).

For these reasons, some researchers have questioned the assumption that visuomotor processes rely on metric representations of target distances. Instead, they have hypothesized that (1) the brain relies mainly on image measurements that specify 3D properties directly, without building an explicit metric representation of the environment, and (2) appropriate body–environment interactions emerge as a consequence of adaptive mechanisms, not as the solution of the “inverse-optics” problem (Braunstein, 1994; Domini & Caudek, 2003; Robert, Zeller, Faugeras, & Hébert, 1997; Thaler & Goodale, 2010; Todd, 2004). In prehension movements aimed at reaching and grasping visual objects, for instance, the (haptic and/or visual) feedback resulting from the contact between the hand and the target provides an error signal for calibration that improves the accuracy of subsequent reaches (e.g., Mon-Williams & Bingham, 2007). Thus, visuomotor actions (such as prehension and pointing) may not require the recovery of the full 3D metric depth map, but instead be based on simpler mechanisms of conditional associative learning. If this is true, we should expect that perceptual metric judgments and motor actions in novel stimulus situations *with no haptic feedback* should be systematically distorted, which indeed has been found to be the case (e.g., Cuijpers, Brenner, & Smeets, 2008).

* Corresponding author. Tel.: +1 401 863 1356.

E-mail addresses: rachel.mary.foster@gmail.com (R. Foster), carlo.fantoni@iit.it (C. Fantoni), corrado.caudek@unifi.it (C. Caudek), fulvio.domini@brown.edu (F. Domini).

In the current investigation, we carried out a *cue combination experiment* in which human performance was measured in three stimulus conditions: with disparity-only information, motion-only information, or both (see also Tittle, Norman, Perotti, & Phillips, 1998). In different blocks of trials, participants either performed a grasping task or provided a perceptual judgment.

Two models of cue integration are considered here. In the first model, image measurements, diagnostic of 3D depth, but insufficient for metric reconstruction, are utilized (*intrinsic constraints*). The second model, instead, is based on the assumption that the brain uses metric structure (i.e., distance and direction) to represent locations (*linear cue integration*). In the next sections, we will describe the two models and show how it is possible to empirically validate their predictions by using the results of the present experiments.

2. Intrinsic constraints

The *intrinsic constraint* (IC) model proposes that, rather than deriving the full *metric* depth map, it is more advantageous for an organism to derive the best estimate of the *local affine* structure and use haptic feedback to calibrate ordinally scaled distance estimates (Di Luca, Domini, & Caudek, 2007; Domini & Caudek, 2010; Domini & Caudek, in press; Domini, Caudek, & Tassinari, 2006; Tassinari, Domini, & Caudek, 2008; see also Bingham & Pagano, 1998; Thaler & Goodale, 2010).

Retinal signals like relative disparity d are direct measures of the local affine structure, because $d \propto z$, where z is the depth map. The precision of the estimate of the affine structure is given by the signal-to-noise ratio (SNR) d/σ_d . We have shown that the best estimate of the affine structure is found through a linear combination of the retinal signals (not the depth estimates recovered from the signals) that maximizes the “information content” of the combined signal (i.e., the SNR – see also MacKenzie, Murray, & Wilcox, 2008). Once retinal signals are combined through this optimal combination rule, they determine a composite signal that encodes the affine structure, but with better precision (i.e., larger SNR) than either would have in isolation. This composite signal has been termed ρ . We propose that visually guided behavior depends upon this combined signal, which is scaled through calibration and perceptual learning from haptic feedback.

In the absence of haptic feedback, we also hypothesize that both perceptual judgments and motor actions are a monotone function of ρ (Domini et al., 2006). Consequently, we expect both perceptual judgments and motor actions to be systematically distorted, because unbiased estimations of 3D properties and target locations cannot be derived from ρ .

2.1. Disparity and motion integration

In the present investigation, we study the integration of disparity and motion information for both a motor task and a perceptual judgment. In both cases, according to IC, in the absence of haptic feedback, the amount of recovered depth z' should be a monotone function of the combined signal ρ :

$$z' = f_\rho(\rho). \tag{1}$$

Domini et al. (2006) showed that ρ is equal to the scores of the first principal component computed from the standardized retinal signals. Consequently, flatness cues are disregarded.

When only one signal is present, ρ is equal to the standardized value of that signal. For disparity-only stimuli, therefore, $\rho_d = \frac{d}{\sigma_d}$, where d is the relative disparity and σ_d is the measurement noise. The amount of depth recovered from disparity is

$$z'_d = f_\rho(\rho_d). \tag{2}$$

For motion-only stimuli, $\rho_v = \frac{v}{\sigma_v}$, where v is the relative velocity and σ_v is the measurement noise. The amount of depth that is recovered from motion information is given by

$$z'_v = f_\rho(\rho_v). \tag{3}$$

When both cues are present, we have that

$$\rho_c = \sqrt{\frac{v^2}{\sigma_v^2} + \frac{d^2}{\sigma_d^2}} \tag{4}$$

and

$$z'_c = f_\rho(\rho_c). \tag{5}$$

If we assume that the function $f_\rho(\rho)$ is linear for the range of depth magnitudes used in the present experiment, then

$$f_\rho(\rho) \approx a_\rho + k_\rho \rho. \tag{6}$$

Therefore,

$$z'_d = a_\rho + k_\rho \frac{d}{\sigma_d}, \tag{7}$$

$$z'_v = a_\rho + k_\rho \frac{v}{\sigma_v}, \tag{8}$$

$$z'_c = a_\rho + k_\rho \rho_c. \tag{9}$$

Considering that

$$d \approx \frac{IOD}{z_f^2} z = k_d z, \tag{10}$$

$$v \approx \frac{\omega}{z_f^2} z = k_v z, \tag{11}$$

where z is the distal relative depth, IOD is the observer's interocular distance, z_f is the fixation distance, and ω is the object's rotation velocity about a vertical axis, it follows that

$$z'_d = a_\rho + k_\rho \frac{k_d}{\sigma_d} z = a_\rho + K_d z, \tag{12}$$

$$z'_v = a_\rho + k_\rho \frac{k_v}{\sigma_v} z = a_\rho + K_v z, \tag{13}$$

where $K_d = k_\rho \frac{k_d}{\sigma_d}$ and $K_v = k_\rho \frac{k_v}{\sigma_v}$. For the disparity–motion (combined) condition, we can thus write

$$\begin{aligned} z'_c &= a_\rho + k_\rho \sqrt{\frac{v^2}{\sigma_v^2} + \frac{d^2}{\sigma_d^2}} z \\ &= a_\rho + k_\rho \sqrt{\frac{k_v^2 z^2}{\sigma_v^2} + \frac{k_d^2 z^2}{\sigma_d^2}} \\ &= a_\rho + z \sqrt{\frac{k_v^2 k_\rho^2}{\sigma_v^2} + \frac{k_d^2 k_\rho^2}{\sigma_d^2}} \\ &= a_\rho + z \sqrt{K_v^2 + K_d^2}. \end{aligned} \tag{14}$$

Eq. (14), therefore, provides a criterion for testing the IC model. K_d and K_v are the slopes of the linear functions relating recovered and distal depth magnitudes for the disparity-only and motion-only conditions, respectively. If the IC model is consistent with human performance, then the slope (K_c) of this linear relation in the

Download English Version:

<https://daneshyari.com/en/article/920131>

Download Persian Version:

<https://daneshyari.com/article/920131>

[Daneshyari.com](https://daneshyari.com)