Vision Research 49 (2009) 2131–2139

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

How long do intrinsic and extrinsic visual cues take to exert their effect on the perceptual upright?

Bahar Haji-Khamneh *, Laurence R. Harris

Dept. Psychology, York University, 4700 Keele St., Toronto, ON, Canada M3J 1P3

ARTICLE INFO

Article history:
Received 10 October 2008
Received in revised form 14 April 2009

Keywords:
Orientation perception
Scene perception
Object perception
Visual frame
Horizon

ABSTRACT

We determined the amount of time it took for intrinsic and extrinsic visual cues to determine the perceptual upright. The perceptual upright was measured using a probe, the identity of which depended on its perceived orientation (the Oriented Character Recognition Test). A visual background that filled the field of view and contained both intrinsic and extrinsic cues was presented in different orientations and for presentation times of between 50 and 500 ms followed by a mask. The contribution of each class of cue was identified by exploiting their different degrees of ambiguity. Intrinsic cues include scene structure (e.g., walls, floor and ceiling of an indoor scene) which indicates four potential up directions, and the horizon which indicates two possibilities. Extrinsic cues, which rely on information not in the image such as a surface acting as a support structure for an object, signal the direction of up unambiguously. The contribution of each class of visual cue could thus be identified from the number of cycles its effect showed as the background was presented in all orientations round the clock. While the more high-level extrinsic cues to up exerted a larger influence on the perceptual upright than the intrinsic cues, the magnitude of each cue's effect increased with presentation time at approximately the same rate with a time constant of about 60 ms. This finding poses a challenge for bottom–up theories of scene perception and suggests that low-level and high-level information are processed in parallel at least insofar as they indicate orientation.

1. Introduction

Vision tells us about the identity of objects ('seeing') but also carries proprioceptive information about the body's orientation relative to the world. Orientation is fundamental to perception and the recognition of objects depends on their orientation. The perceived direction of 'up' has conventionally been measured using the subjective visual vertical (e.g., Mittelstaedt, 1983). However, the orientation at which objects appear upright (the perceptual upright) is not always the same as the orientation of the subjective visual vertical because the perceptual upright is more heavily influenced by orientation of the visual background. Dyde, Jenkin, and Harris (2006) define the perceptual upright (PU) as being the orientation at which objects are recognized as being "the right way up". The right way up is the orientation at which objects are most readily and accurately identified and is fundamental to our ability to interact with the environment. The perceptual upright is conceptually distinct from the 'canonical orientation' which defines 'the right way up' as the orientation at which objects are most accurately and speedily recognized (see for e.g., Jolicoeur, 1985; McMullen & Jolicoeur, 1992). While the perceptual upright and the 'canonical orientation' are closely related concepts, and would likely both be influenced to the same extent by background scene orientation, the canonical orientation is derived from reaction time data whereas the PU is derived from a character recognition task. The perceptual upright is derived from a combination of visual and vestibular cues, together with an internal representation of the orientation of the body (Asch & Witkin, 1948a; Dyde et al., 2006; Mittelstaedt, 1986, 1999). Here we investigate specifically the contribution of the visual cue to the perceptual upright.

A typical scene contains both intrinsic and extrinsic visual cues to orientation. The overall frame or structure of the scene (floor or ground plane, walls, ceiling or sky) and the orientation of the horizon (even if not directly visible) are intrinsic to a scene. By contrast, the spatial-relationships between and within objects (that a table can act as a support surface for an object; that a lampshade is at the top of a lamp standard) are not intrinsic to scenes and have to be learned through familiarity with statistical regularities in the environment (Schwarzkopf & Kourtzi, 2008) and an internalization of the laws of physics (McIntyre, Zago, Berthoz, & Lacquaniti, 2001). These learned relationships constitute an axis of polarity that does not change when the overall scene changes in orientation. Such extrinsic cues will be referred to as polarizing cues. The knowledge that light comes from above (Mamassian & Goutcher, 2001; Ramachandran, 1988) can also be used to specify the orientation of an object or scene using shading and shadows. The interpretation of this cue can be altered by experience suggesting

---

* Corresponding author. Fax: +1 416 736 5814.
E-mail address: bahar.haji@gmail.com (B. Haji-Khamneh).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.06.003

that the light cue is also at least partially extrinsic (Adams, Graf, & Ernst, 2004). Whether intrinsic and extrinsic cues are processed by the same or different mechanisms is unknown.

Intrinsic and extrinsic cues both contribute to determining the PU. However, as Fig. 1 shows, some of these cues have different degrees of ambiguity and indicate more than one direction of up. The fact that different cues are differentially ambiguous can be used to identify their contributions in a given scene. The intrinsic cue that comes from the structure of a room provides four potential directions of up: as the scene is rotated, each of these directions aligns with gravity every 90° of rotation. Likewise, the line specifying the elevation of the horizon simultaneously indicates two directions of upright separated by 180°. In contrast to these ambiguous intrinsic cues, extrinsic cues specify a unique direction of up. Each of these cues is able to influence the orientation of the PU. Thus when a scene filling the visual field is presented at all orientations, the effect induced by the three classes of visual components within it can be distinguished by the number of cycles of shift of the perceptual upright that the tilted scene induces: the effect of the frame cues will complete four cycles, the horizon's effect will complete two and extrinsic cues will always indicate a unique direction.

While much is known about various properties of the global context such as color (Oliva & Schyns, 2000; Steeves et al., 2004) and spatial frequency (Rousselet, Joubert, & Fabre-Thorpe, 2005), relatively little is known about the influence of the orientation of the global context on the perception of self and object orientation (Rousselet, Macé, & Fabre-Thorpe, 2003; Vuong, Hof, Bülthoff, & Thornton, 2006). Extracting the gist of a scene can be done in less than 150 ms (Hegde, 2008) but is the time it takes to extract a gist comparable to the time it takes for a scene to exert an influence on the perception of objects within it? Here we measured the time course with which each class of cue present in the scene exerted its effect, expecting that differential processing systems would be reflected in different amounts of time needed for each type of cue to exert its effect. If higher-level extrinsic polarizing cues require more semantic and spatial processing than relatively low-level frame and horizon cues, then we might expect that such cues would exert their effect at a later stage than low-level intrinsic cues and should take longer. Conversely, if low-level and high-level information were processed in parallel, we would expect no differences in the time course of intrinsic and extrinsic cues.

To test these hypotheses we used the Oriented CHAracter Recognition Test (OCHART) (Dyde et al., 2006) which exploits the notion that the letters ‘p’ and ‘d’ rely on their orientation for their identity. By identifying the orientation at which the letter's identity is least certain (i.e., when either identify is equally likely to be perceived) we can obtain an estimate of the orientation at which its orientation is most certain: the perceptual upright. The influence of the orientation of the visual background was obtained by repeating OCHART with the background at different orientations. Each background was presented for a fixed period of time between 50 and 500 ms followed immediately by a pattern mask that limited the processing time to the presentation duration.

2. Methods

2.1. Subjects

Three females and five males between the ages of 24 and 45 participated in these experiments. All observers had normal or corrected-to-normal vision. All observers gave informed consent as required by the Ethics Guidelines of York University which complies with the 1964 Declaration of Helsinki. Six of the participants were volunteers and the other two were compensated at a rate of $10 per session. All participants took part in all experiments.

2.2. Apparatus

Stimuli were presented on a 21 in. Dell P1110 Trinitron monitor with a resolution of 28.3 pixels/cm and a mean luminance of 43.15 cd/m^2 at a refresh rate of 120 Hz (i.e., 8.33 ms/frame). Stimuli were composed one frame at a time and presented using Psyscope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993; MacWhinney, Cohen, & Provost, 1997). Because the timing of the stimulus and mask presentation on the computer screen was

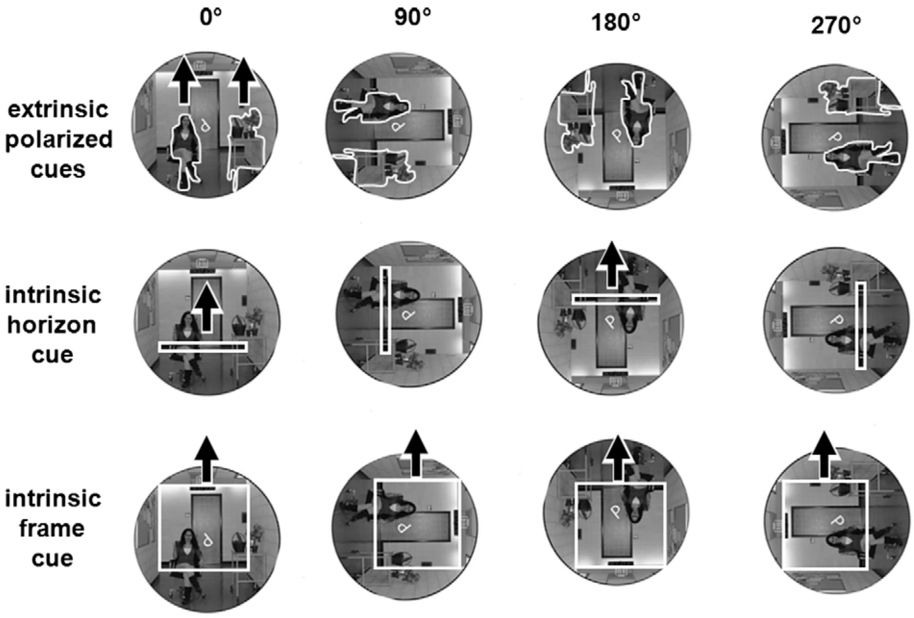

Fig. 1. A visual scene contains several cues to orientation including high-level extrinsic polarizing cues (highlighted in the top row) and low-level intrinsic cues from the horizon and visual frame (highlighted in the middle and bottom rows, respectively). When the picture is rotated through 360°, the direction of up specified by the polarizing cues rotates through one cycle (top row); the direction specified by the horizon cue rotates two cycles (middle row) and the direction indicated by the frame cue (the square formed by the edges of the walls and the floor and ceiling) rotates through four cycles (bottom row).