Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 145–156
Vision Research
www.elsevier.com/locate/visres

Visual learning by cue-dependent and cue-invariant mechanisms

Volodymyr Ivanchenko, Robert A. Jacobs *

Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY 14627, USA

Received 15 December 2005; received in revised form 13 September 2006

Abstract

We examined learning at multiple levels of the visual system. Subjects were trained and tested on a same/different slant judgment task or a same/different curvature judgment task using simulated planar surfaces or curved surfaces defined by either stereo or monocular (texture and motion) cues. Taken as a whole, the results of four experiments are consistent with the hypothesis that learning takes place at both cue-dependent and cue-invariant levels, and that learning at these levels can have different generalization properties. If so, then cue-invariant mechanisms may mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient. We claim that learning takes place at multiple levels of the visual system, and that a comprehensive understanding of visual perception requires a good understanding of learning at each of these levels.

Keywords: Visual learning; Cue-invariance

1. Introduction

Despite decades of research, perceptual learning is a poorly understood phenomenon. Perhaps the most important lesson that research has taught us is that our current theories and experiments are too simple and too narrowly focused. It is likely that perceptual learning takes place at multiple levels of the human perceptual system, and that a comprehensive understanding of perception will require a good understanding of learning at each of these levels. Unfortunately, the study of perceptual learning at multiple levels is nearly unexplored in the scientific literature (see Ahissar & Hochstein, 1997, 2002, for a notable exception). This lack of understanding of learning at multiple levels is, we believe, a major reason why the literature on perceptual learning often contains seemingly confusing (and contradictory) results.

This article reports the results of experiments investigating learning at two levels of the visual system, namely the levels of visual cue-dependent and visual cue-invariant mechanisms (e.g., shape-from-visual-texture or shape-from-visual-motion mechanisms versus a mechanism for perceiving shape that is independent of the visual cue used to define the shape).[1] Within the vision sciences, the study of visual cue-invariant mechanisms is relatively unusual. These mechanisms ought to be of fundamental interest to scientists because visual perception of natural environments must integrate information provided by multiple cues. In this sense, these mechanisms can be regarded as among the "highest level" mechanisms of our visual systems.

* Corresponding author. Fax: +1 585 442 9216.
E-mail addresses: vivanchenko@bcs.rochester.edu (V. Ivanchenko), robbie@bcs.rochester.edu (R.A. Jacobs).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.09.028

[1] We hypothesize that visual cue-invariant mechanisms are constructed from cue-dependent mechanisms. For example, a cue-invariant mechanism for representing visual shape might receive inputs from both a mechanism that represents shape-from-visual texture and a mechanism that represents shape-from-visual-motion (and, perhaps, inputs from several other cue-dependent mechanisms for representing shape). If this mechanism's output at any moment in time does not depend on which mechanism provided an input, then its output would be cue-invariant. To our knowledge, the vision sciences literature does not contain any studies directly evaluating this hypothesis.

Do visual cue-invariant mechanisms exist? Recent psychophysical data suggests that the human visual system may contain neural mechanisms that represent object shape or depth independent from the visual cue(s) specifying the shape or depth. For example, Poom and Börjesson (1999) reported that prolonged viewing of an adaptation surface caused a test surface to appear to slant in the direction opposite to that of the adaptation surface regardless of whether the two surfaces were defined by the same cue (either motion parallax or binocular disparity) or different cues. Other behavioral studies suggesting visual cue-invariant mechanisms are Bradshaw and Rogers (1996) and Domini et al. (2001).

Related data have been found in neuroscientific studies using monkeys. For example, Sakata et al. (1999) showed that some visually responsive neurons in the macaque anterior intraparietal area encode surface tilt regardless of whether the tilt is specified by disparity alone, monocular cues alone, or both. Other neuroscientific studies indicating visual cue-invariant mechanisms in monkeys are Sary, Vogels, and Orban (1993), Sereno, Trinath, Augath, and Logothetis (2002), Tsutsui, Sakata, Naganuma, and Taira (2002).

Brain-imaging studies using human observers have reported similar data. Grill-Spector, Kushnir, Edelman, Itzchak, and Malach (1998) found that a region located on the lateral aspect of the occipital lobe was preferentially activated during a visual object recognition task relative to control conditions irrespective of whether the object shape was defined by luminance, motion, or texture cues. Kourtzi and Kanwisher (2000) reported overlapping activations in the lateral and ventral occipital cortex for objects depicted by different visual formats (grayscale images and line drawings), and a reduced response when objects were repeated, independent of whether they recurred in the same or a different format.[2] Other relevant brain-imaging studies using human observers are reported in Kourtzi, Betts, Sarkhei, and Welchman (2005) Welchman, Deubelius, Conrad, Bülthoff, and Kourtzi (2005).

Although the studies cited above suggest the existence of visual cue-invariant mechanisms, they did not examine the nature of these mechanisms in a detailed way and, importantly for our purposes, they did not examine the role of these mechanisms in perceptual learning. To date, we are aware of only one study on cue-invariant mechanisms and perceptual learning. Rivest, Boutet, and Intrilligator (1996) trained different sets of observers to visually discriminate the orientations of color-defined bars, of luminance-defined bars, or of motion-defined bars. A similar improvement from pre-test to post-test was found regardless of whether the bars seen after training were defined by the same or by a different cue as the cue seen during training. The authors concluded that training changed the sensitivity of cells that represent visual orientation in a cue-invariant manner.

This article studies the hypothesis that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient. For example, if an observer learns to make more accurate depth-from-visual-texture judgments, then it would be advantageous to the observer to generalize this gained knowledge so that it can be used when estimating depth from cues other than texture, such as when making depth-from-visual-motion judgments. An important goal of the reported experiments is to evaluate this hypothesis. A secondary goal is to compare the generalization properties of visual cue-dependent versus cue-invariant mechanisms. We hypothesize that the "lower level" cue-dependent mechanisms tend to use local representations that lead to stimulus-specific learning (i.e., learning effects are limited to the specific stimulus conditions used during training), whereas the "higher level" cue-invariant mechanisms tend to use global representations that lead to stimulus-general learning (i.e., learning effects generalize to novel stimulus conditions). To our knowledge, there are currently no studies comparing the properties of cue-dependent versus cue-invariant mechanisms.

The results of four experiments are reported. In the first experiment, subjects were trained to judge the 3D orientations of planar surfaces slanted in depth when surfaces were defined by a training cue and when slants were centered near a training slant. Subjects were tested on the same task when surfaces were defined by either the training cue or a novel cue, and when slants were centered either near the training slant or near a novel slant. Because subjects showed improved performance when tested both with the training cue and with the novel cue, the results suggest that training produced modifications to both cue-dependent and cue-invariant mechanisms. Furthermore, these two sets of mechanisms seem to have different properties—cue-dependent mechanisms of visual slant are slant-specific whereas cue-invariant mechanisms are not. Experiment 2 was similar to Experiment 1, but it required subjects to judge the slants of cylinders. As in Experiment 1, its results suggest that training produced modifications to both cue-dependent and cue-invariant mechanisms, thereby producing transfer of learning from training to novel cue conditions. In addition, this experiment found that both sets of mechanisms either ignored or generalized over an irrelevant shape attribute. Experiment 3 required subjects to judge the curvature-in-depth of cylinders. The results again demonstrate learning by both cue-dependent and cue-invariant mechanisms. Experiment 4 found that learning

[2] It is interesting to note that cue-invariance may take place across sensory modalities, not just within the visual modality. Brain-imaging studies with humans have provided evidence for neural mechanisms which are modality-invariant. Amedi, Malach, Hendler, Peled, and Zohary (2001) found preferential activation in the lateral occipital complex when observers viewed objects and also when they grasped the same objects. Pietrini et al. (2004) found that visual and tactile recognition of man-made objects evoked category-related patterns of responses in a ventral extrastriate visual area in the inferior temporal gyrus that were correlated across sensory modality.