Cognition 126 (2013) 135–148

Contents lists available at SciVerse ScienceDirect

Cognition

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

Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies

Ilker Yildirim *, Robert A. Jacobs

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, United States

ARTICLE INFO

Article history:
Received 2 January 2012
Revised 16 August 2012
Accepted 19 August 2012
Available online 25 October 2012

Keywords:
Multisensory perception
Vision
Touch
Learning
Categorization
Experimentation
Computational modeling

ABSTRACT

We study people's abilities to transfer object category knowledge across visual and haptic domains. If a person learns to categorize objects based on inputs from one sensory modality, can the person categorize these same objects when the objects are perceived through another modality? Can the person categorize novel objects from the same categories when these objects are, again, perceived through another modality? Our work makes three contributions. First, by fabricating Fribbles (3-D, multi-part objects with a categorical structure), we developed visual-haptic stimuli that are highly complex and realistic, and thus more ecologically valid than objects that are typically used in haptic or visual-haptic experiments. Based on these stimuli, we developed the *See and Grasp* data set, a data set containing both visual and haptic features of the Fribbles, and are making this data set freely available on the world wide web. Second, complementary to previous research such as studies asking if people transfer knowledge of object identity across visual and haptic domains, we conducted an experiment evaluating whether people transfer object category knowledge across these domains. Our data clearly indicate that we do. Third, we developed a computational model that learns multisensory representations of prototypical 3-D shape. Similar to previous work, the model uses shape primitives to represent parts, and spatial relations among primitives to represent multi-part objects. However, it is distinct in its use of a Bayesian inference algorithm allowing it to acquire multisensory representations, and sensory-specific forward models allowing it to predict visual or haptic features from multisensory representations. The model provides an excellent qualitative account of our experimental data, thereby illustrating the potential importance of multisensory representations and sensory-specific forward models to multisensory perception.

1. Introduction

When recording neural activity in the human medial temporal lobe, Quiroga, Kraskov, Koch, and Fried (2009) found individual neurons that explicitly encode multisensory percepts. For example, one neuron responded selectively when a person viewed images of the television host Oprah Winfrey, viewed her written name, or heard her spoken name. (To a lesser degree, the neuron also responded to the actress Whoopi Goldberg.) Another neuron responded selectively when a person saw images of the former Iraqi leader Saddam Hussein, saw his name, or heard his name. Clearly, our brains encode abstract representations of objects that are multisensory in the sense that these representations are activated by perceptual inputs, but these inputs span multiple sensory formats or modalities.

Why would our brains acquire abstract representations that are activated by inputs from a variety of sensory modalities? One possible answer to this question is that these representations facilitate the transfer of knowledge across modalities. Consider, for instance, a person that

* Corresponding author. Tel.: +1 585 275 7187; fax: +1 585 442 9216.
E-mail addresses: iyildirim@bcs.rochester.edu (I. Yildirim), robbie@bcs.rochester.edu (R.A. Jacobs).

0010-0277/$ - see front matter
http://dx.doi.org/10.1016/j.cognition.2012.08.005

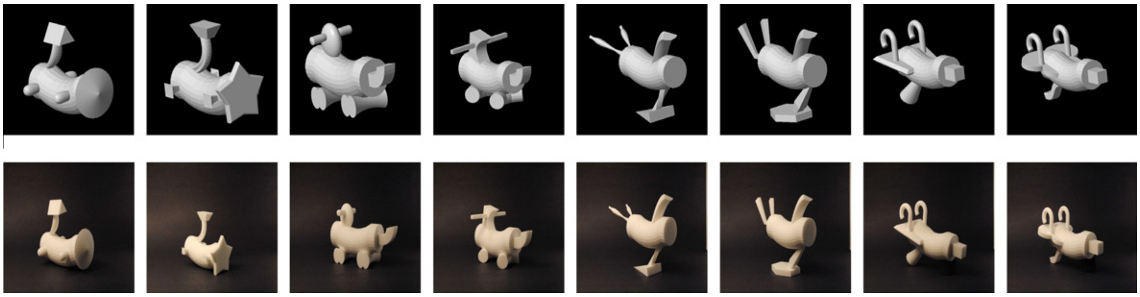

Fig. 1. The top row shows computer-generated images of Fribbles which are rendered using the Fribbles' 3-D object models. The bottom row shows photographs of the physical objects corresponding to these same Fribbles which were fabricated via a 3-D printing process using the same 3-D object models. Pairs of columns illustrate exemplars from different categories (e.g., columns 1–2 illustrate exemplars from category A).

learns to categorize a set of objects based solely on tactile or haptic inputs. Would the person be able to categorize these same objects when the objects are viewed but not grasped? Would the person be able to view novel objects from the same categories and be able to categorize these?

Here, we report experimental and computational studies of the acquisition of multisensory representations of object category, and the role these representations play in the transfer of knowledge across visual and haptic modalities. Our work includes three contributions. First, our experiment used an unusual set of visual-haptic stimuli known as "Fribbles". Fribbles are complex, 3-D objects with multiple parts and spatial relations among the parts (see Fig. 1). Moreover, they have a categorical structure—that is, each Fribble is an exemplar from a category formed by perturbing a category prototype. Fribbles have previously been used in the study of visual object recognition (Hayward & Williams, 2000; Tarr, 2003; Williams, 1997). An innovation of our work is that we have fabricated a large set of Fribbles using a 3-D printing process and, thus, our Fribbles are physical objects which can be both seen and grasped. Based on this set of stimuli, we have created a data set, referred to as the *See and Grasp* data set, containing both visual and haptic features of the Fribbles. We are making this data set freely available on the world wide web with the hope that it will encourage quantitative research on computational models of visual-haptic perception.

Second, we conducted an experiment evaluating whether people can transfer knowledge of object category across visual and haptic modalities. Previous researchers have considered the transfer of knowledge of object identity across visual and haptic modalities (e.g., Lacey, Peters, & Sathian, 2007; Lawson, 2009; Norman, Norman, Clayton, Lianekhammy, & Zielke, 2004). They have also compared similarity and categorization judgements based solely on visual input with those based solely on haptic input (Gaißert & Wallraven, 2012; Gaißert, Bülthoff, & Wallraven, 2011; Gaißert, Wallraven, & Bülthoff, 2008, 2010). To our knowledge, our experiment is the first focused on the transfer of object category knowledge across visual and haptic modalities.

Lastly, we developed a computational model, referred to as the MVH (Multisensory-Visual-Haptic) model, accounting for how multisensory representations of prototypical 3-D shape might be acquired, and of the role these representations might play in the transfer of category knowledge across visual and haptic modalities. Like some previous models in the literature (Biederman, 1987; Marr & Nishihara, 1978), the model makes use of part-based representations of prototypes. However, it goes beyond previous work by introducing a learning mechanism for the acquisition of these representations. Using its acquired multisensory representations along with sensory-specific forward models for predicting visual or haptic features from multisensory representations, the model transfers object category knowledge between visual and haptic modalities, thereby providing a qualitative account of our experimental data.

2. Previous research on visual-haptic object perception

Previous research has shown that knowledge of object identity transfers (at least in part) across visual and haptic domains (e.g., Lacey, Peters, et al., 2007; Lawson, 2009; Norman et al., 2004). For example, Lacey, Peters, et al. (2007) trained subjects to identify objects either visually or haptically. Following training, subjects were tested on the same task using the untrained sensory modality. Subjects showed excellent transfer to the novel modality when objects were presented at the same orientation as experienced during training, and still showed good transfer when objects were rotated to a new viewpoint.

Researchers have also compared people's vision-only and haptic-only similarity judgements. For example, Gaißert and colleagues collected people's unisensory similarity judgements for naturalistic objects resembling sea shells (Gaißert and Wallraven, 2012; Gaißert, Bülthoff, et al., 2011; Gaißert et al., 2008, 2010). Analyses based on multidimensional scaling showed that people's vision-only and haptic-only similarity spaces were nearly identical. Gaißert and colleagues also examined people's vision-only and haptic-only categorization judgements. Analyses showed that these categorizations were highly similar to each other, and that they were consistent with people's similarity judgements (also see Haag, 2011).

Additional research has compared the acquisition of haptic concepts by blind individuals and sighted controls. Homa, Kahol, Tripathi, Bratton, and Panchanathan (2009) found that blind subjects learned the categories quickly and comparably with sighted subjects. Other research has