BRAIN RESEARCH 1242 (2008) 4–12

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Review

Linking neurons to behavior in multisensory perception: A computational review

Wei Ji Ma[*], *Alexandre Pouget*

Department of Brain and Cognitive Sciences, University of Rochester, Rochester NY 14627, USA

ARTICLE INFO

Article history:
Accepted 27 April 2008
Available online 21 June 2008

Keywords:
Multisensory integration
Cue combination
Computational modeling
Population coding
Neural variability
Superadditivity
Segregation

ABSTRACT

A large body of psychophysical and physiological findings has characterized how information is integrated across multiple senses. This work has focused on two major issues: how do we integrate information, and when do we integrate, i.e., how do we decide if two signals come from the same source or different sources. Recent studies suggest that humans and animals use Bayesian strategies to solve both problems. With regard to how to integrate, computational studies have also started to shed light on the neural basis of this Bayes-optimal computation, suggesting that, if neuronal variability is Poisson-like, a simple linear combination of population activity is all that is required for optimality. We review both sets of developments, which together lay out a path towards a complete neural theory of multisensory perception.

Contents

Accurate perception frequently relies on combining uncertain information from multiple senses. Imagine that you are trying to locate a predator hiding in the bushes. You hear a faint sound of the predator's footsteps and at the same time you see a movement of the leaves. That movement could be caused by the animal, but also by a gust of wind. If the predator caused the movement, the visual information will help you localize it with greater precision.

This example illustrates several general aspects of multisensory perception. Combining information across senses can

---

[*] *Corresponding author.* Department of Brain and Cognitive Sciences University of Rochester Rochester, NY 14627, USA. Fax: +1 585 442 9216.

E-mail address: weijima@gmail.com (W.J. Ma).

0006-8993/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2008.04.082

be of critical importance to an animal's survival, making it plausible that evolutionary pressure has optimized the neural circuits that serve this purpose. Moreover, those circuits have to solve two problems simultaneously: figuring out whether two cues had the same source (the predator) or different sources (the predator and the wind), and in the former case, how to combine them. Finally, cues can come with different reliabilities. Visual information will be more reliable on a sunny day than on a foggy day, and you can trust auditory information more if there is little background noise. These aspects have guided the theoretical developments we discuss in this review.

1. Optimal cue integration

When a common source is assumed, a systematic strategy to quantify cue combination is to introduce a small discrepancy (also called conflict, disparity, or incongruency) between the cues. The conflict must be small in order to not violate the common-source assumption. In such a paradigm, the percept (estimate of the stimulus) inferred from both cues presented together will lie somewhere in between the percepts inferred from each cue individually. The intuition is that higher weight will be given to the most reliable cue, and that therefore the multi-cue percept will be closest to the percept obtained from that cue. Recent psychophysical studies have quantified this intuition, both across (Alais and Burr, 2004; Battaglia et al., 2003; Ernst and Banks, 2002; van Beers et al., 1996; Wolpert et al., 1995) and within sensory modalities (Jacobs, 1999; Knill and Saunders, 2003). As an example, we consider a laboratory version of the ventriloquist effect (Alais and Burr, 2004), the well-known illusion in which a performer makes a puppet appear to speak (Howard and Templeton, 1966; Welch and Warren, 1980). This experiment involved spatial localization along the azimuthal dimension, based on brief visual flashes and auditory clicks. Importantly, observers were instructed to regard each pair of multisensory signals as being caused by a single, well-localized event, for instance a ball hitting the screen. The investigators found that the mean auditory–visual estimates of location, locations $\hat{s}_{\mathrm{AV}}$, could be expressed as a linear combination of the auditory and visual s_{A} and s_{V}:

$$\hat{s}_{\mathrm{AV}} = \frac{w_{\mathrm{A}} s_{\mathrm{A}} + w_{\mathrm{V}} s_{\mathrm{V}}}{w_{\mathrm{A}} + w_{\mathrm{V}}} \tag{1}$$

In this expression, the weights are given by the inverse variances of estimates in the respective modalities: $w_{\mathrm{A}} = \frac{1}{\sigma_{\mathrm{A}}^2}$ and $w_{\mathrm{V}} = \frac{1}{\sigma_{\mathrm{V}}^2}$. For example, if in a certain condition the visual variance is larger than the auditory variance (and therefore vision is less reliable than audition), vision will be given less weight than audition in the combination.

Moreover, the inverse variance of the auditory–visual estimates was found to be

$$\frac{1}{\sigma_{\mathrm{AV}}^2} = \frac{1}{\sigma_{\mathrm{A}}^2} + \frac{1}{\sigma_{\mathrm{V}}^2} \tag{2}$$

This indicates that using two cues led to higher precision than using any one cue. The right-hand side of Eq. (2) also gives the highest possible precision that can be achieved by an unbiased estimator, given σ_{A} and σ_{V}. Eqs. (1) and (2) state what is meant by statistical optimality in this task. Although they summarize average human behavior over many trials (they give the mean and variance of maximum-likelihood estimates), it is commonly assumed that they reflect regularities that hold on a trial-by-trial basis. On a single trial, we can think of a sensory cue as providing a probability distribution over the stimulus. If we denote the auditory-only distribution by $p(s \mid A)$, the visual-only distribution by $p(s|V)$ and the multisensory distribution by $p(s|A, V)$, then the optimal multisensory distribution is the product distribution (Clark and Yuille, 1990; Yuille and Bulthoff, 1996)

$$p(s|A, V) \propto p(s|A)p(s|V), \tag{3}$$

where the proportionality is such that $p(s \mid A,V)$ is normalized to 1. We have assumed that the auditory and visual distributions are independent given the stimulus (this is called conditional independence). When the distributions in Eq. (3) are Gaussian, Eqs. (1) and (2) directly follow from Eq. (3). As human behavior follows Eqs. (1) and (2) in a wide variety of paradigms, (multisensory) cue integration has become a poster child of Bayes-optimal computation.

Several years ago, a review article stated that these findings of approximate Bayes-optimal cue integration in humans raised two central questions (Banks, 2004): "1. how does the brain know the variances of its sensory estimates to make the correct weight assignments; 2. how does the brain know when sensory estimates are coming from the same source and not different sources, so that combining makes sense?" Since then, significant progress has been made on both these questions, in particular in the theoretical domain.

2. Optimal cue integration with neural populations

When studying how neuronal circuits implement near-optimal cue integration, an important fact to take into account is that the responses of cortical neurons are typically very variable (Compte et al., 2003; Dean, 1981; Holt et al., 1996; Tolhurst et al., 1982). Presenting the same stimulus repeatedly will give rise to many different population responses. A first sight, such variability is a nuisance that could compromise optimality. Recent work, however, has argued that the presence of variability is not the problem. If we experience uncertainty about a stimulus, this stimulus must generate variability in the brain, otherwise there would be no uncertainty. However, the format of the neural variability is important in the neural implementation of the optimal cue integration (Eq. (3)) (Ma et al., 2006). If the statistics of the variability are known (either to the experimenter or to downstream neurons), then Bayes' rule can be used to convert the population pattern of activity on a single trial into a probability distribution over the stimulus. To be precise, if the population activity on a single trial is denoted by a vector $\mathbf{r}=(r_1, r_2, \ldots, r_N)$, where r_i is the activity of the ith neuron and N is the number of neurons, then one can obtain the so-called posterior distribution through

$$p(s|r) \propto p(\mathbf{r}|s)p(s), \tag{4}$$

where $p(\mathbf{r}|s)$ is the response distribution and $p(s)$ is the prior distribution (Foldiak, 1993; Sanger, 1996). The posterior