MIT Press ESForum

15

Coordination in Sensory Integration

Jochen Triesch, Constantin Rothkopf,
and Thomas Weisswange

Abstract

Effective perception requires the integration of many noisy and ambiguous sensory signals across different modalities (e.g., vision, audition) into stable percepts. This chapter discusses some of the core questions related to sensory integration: Why does the brain integrate sensory signals, and how does it do so? How does it learn this ability? How does it know when to integrate signals and when to treat them separately? How dynamic is the process of sensory integration?

Introduction

Merriam-Webster defines coordination as "the harmonious functioning of parts for effective results" and as such it is omnipresent in brain and behavior. During perception, the brain has to make sense of sensory signals from a number of different modalities (e.g., vision, audition, olfaction, touch, proprioception). These signals need to be processed and integrated to compute an (usually correct) interpretation of the environment. How this happens is a fundamental problem. In this chapter we raise a number of central questions regarding how sensory integration pays special attention to dynamic coordination in the brain. A more detailed review of sensory integration is provided elsewhere (Rothkopf et al. 2010).

Why Integrate?

Perception is a difficult computational problem. The state of the world must be inferred from noisy and ambiguous sensory signals. To reach a solution, the brain must rely on making use of all available sources of information—from the different sensory modalities mentioned above (Stein and Meredith 1993),

From "Dynamic Coordination in the Brain: From Neurons to Mind,"
C. von der Malsburg, W. A. Phillips, and W. Singer, eds. 2010. Strüngmann Forum Report, vol. 5,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-01471-7.

to different so-called *cues* within one modality. There are, for example, many so-called *depth cues* in visual perception that are thought to contribute to the perception of an object's distance from the observer (Landy et al. 1995). Next to the integration of various sources of evidence, perception utilizes, to a large extent, previously acquired knowledge about the world. Such prior information will be particularly important when the sensory data are very ambiguous (Weiss et al. 2002) and may be either innate or the result of learning, and thus subject to constant adaptation.

The benefit of combining several sources of information is twofold. First, our estimates of the state of the world will become more accurate as we integrate several noisy sources of information. To date, most work on sensory integration has focused on this aspect, and has been demonstrated amply in humans and various animal species. Second, we may be able to respond more quickly; that is, processing time is reduced when stimuli are presented in more than one modality. Both aspects are of obvious relevance for an organism's survival and well-being and can be closely related. Thus, when several sources of noisy evidence are available, under certain assumptions we can obtain the same amount of information from these sources if we observe a single source for a long time or several of them for a correspondingly shorter time.

How to Integrate?

A natural starting point for determining how the brain might integrate sensory information from different cues or modalities is to ask: What is the *optimal solution* to the problem? Such questions can be answered in the popular framework of *Bayesian inference* (Pearl 1988), for which many reviews are available (Kersten et al. 2004; Kersten and Yuille 2003; Yuille and Kersten 2006). In this framework one can construct so-called "ideal observers" that use all the available sensory information in an optimal fashion according to the laws of probability and statistics. After the ideal observer has been constructed and its behavior has been analyzed, it can be compared to that of human subjects or animals. In many (simple) situations, human behavior has been well-modeled by an appropriate ideal observer model, and this is usually taken as evidence that the brain performs Bayesian inference.

Unfortunately, however, solving the Bayesian inference problem and constructing appropriate ideal observer models can be a very difficult task. In the most general setting, Bayesian inference belongs to a class of computational problems that requires an exponentially increasing amount of processing as the problem size gets bigger (e.g., the more sensory variables are involved). In these situations, it may be infeasible to construct the ideal observer, and thus approximations have to be made. As a consequence, it is impossible to judge whether human behavior is optimal. However, since the brain will also have to use approximations to solve the Bayesian inference problem, it is important

From "Dynamic Coordination in the Brain: From Neurons to Mind,"
C. von der Malsburg, W. A. Phillips, and W. Singer, eds. 2010. Strüngmann Forum Report, vol. 5,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-01471-7.

to ask what kinds of approximations it is using. In principle, this question can be answered within the Bayesian inference framework but we do not know of specific examples where this has been demonstrated.

Finally, viewing cue integration (and more generally perception) as Bayesian inference happens entirely on a computational level. It is still unclear how the neural implementation of Bayesian inference (or approximations of it) would look at the level of groups of neurons exchanging action potentials. This constitutes one of the most important and pressing questions in the field of computational neuroscience (Deneve 2005; Ma et al. 2006). A promising view is put forward by Phillips, von der Malsburg, and Singer (this volume), who argue that dynamic coordination is relevant to Bayesian inference because the distinction between driving and modulatory interactions is implied by the way in which posterior probabilities are computed from current data and prior probabilities.

How Does the Brain Learn How to Integrate?

The concept of Bayesian inference provides a powerful framework for studying sensory integration, but it does not address how the brain acquires the necessary probabilistic models: How does it decide what sensory variables to represent? How does it learn their statistical relationships? In the machine learning and statistics communities, progress has been in understanding how such models and their parameters can be learned, but optimal Bayesian learners are even harder to construct than ideal observers, and human learning can deviate strongly from the ideal case.

Experimental evidence regarding the acquisition of sensory integration abilities stems from developmental studies with children and learning experiments with adults. Interestingly, recent experiments with children suggest that it may take many years before children exhibit appropriate sensory integration abilities consistent with ideal observer models (Gori et al. 2008; Nardini et al. 2008; Neil et al. 2006). Initially, they may not be integrating different modalities at all (Gori et al. 2008).

In adult learning experiments, a relatively simple case is the one where the set of different cues is fixed and only their relative weighting changes. In visual cue integration, for example, Ernst et al. (2000) and Atkins et al. (2001) showed that when two conflicting visual cues are paired with a haptic cue, subjects will, over the course of a few days, learn to increase the weight of the visual cue that is consistent with the haptic cue and decrease the weight of the inconsistent cue. Thus, it appears that the haptic cue serves as a reference model for adjusting the visual cues.

When to Integrate?

Another fundamental issue concerns the timing of signals from different modalities or cues: when should they be combined or when should they be

From "Dynamic Coordination in the Brain: From Neurons to Mind,"
C. von der Malsburg, W. A. Phillips, and W. Singer, eds. 2010. Strüngmann Forum Report, vol. 5, series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-01471-7.

considered separately (Koerding et al. 2007)? An interesting problem in this context is that of audiovisual source localization. Imagine your task is to estimate the location of one or two target objects that are presented simultaneously in the auditory and/or visual domain through brief light flashes and sounds. When one auditory and one visual signal are received, but they seem to be very far apart, then it is prudent to assume that they did not originate from the same object and thus should not be integrated into a single percept. In contrast, if the two sources are sufficiently close, we assume that there is only a single object giving rise to both the auditory and the visual signal. In this case, it may be better to integrate the position estimates to arrive at a single, more precise estimate.

In Bayesian terms, the brain considers two different models to explain observed sensory signals. The first posits that there are two distinct objects: one producing the visual signal; the other producing the auditory signal. The second model posits that there is only one object giving rise to both the visual and auditory stimulus. How, then, might the brain make the appropriate determination? One obvious strategy is to evaluate both models and choose the one judged most probable, a technique called *model selection*. A plausible alternative is to evaluate both models and average their interpretations; different weights are given to both models in the averaging process, depending on how likely they appear. This technique is called *model averaging*. Recent research has started to address which strategy human subjects use (Shams and Beierholm 2009). However, thus far, evidence has been mixed. People appear to behave differently in different tasks, and large individual differences between subjects have been observed.

How does the brain *learn* when to integrate signals from different modalities and when to treat them separately? Recently, Weisswange et al. (2009) demonstrated that this ability could be acquired through generic reinforcement learning mechanisms (Sutton and Barto 1998). In their model, an agent needs to make orienting movements toward objects and is rewarded for localizing them precisely. In the situation where a visual and an auditory input are close together, the model will integrate them into a single position estimate. When they are far apart, the model will orient toward either the visual or the auditory stimulus without trying to integrate them. Although this model cannot prove that the brain acquires the ability to select the appropriate model in a similar way, it shows that the underlying reinforcement learning mechanism is sufficient to produce this behavior.

How Dynamic Is Cue Integration?

Whether or not different pieces of sensory information are integrated depends on the current situation (e.g., stimuli, context, behavioral goals). For a decision to be reached, different modalities and cues need to be *dynamically coordinated*.

From "Dynamic Coordination in the Brain: From Neurons to Mind,"
C. von der Malsburg, W. A. Phillips, and W. Singer, eds. 2010. Strüngmann Forum Report, vol. 5,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-01471-7.

Though the need to determine dynamically what aspects of the current sensory input should be integrated and what should be segregated has been well studied within submodalities such as motion perception (Braddick 1993), very little work has addressed to date this issue in relation to multicue or multimodal integration. Instead, most work has assumed a fixed situation, with a fixed set of sensory cues, and with fixed reliabilities that are known to the subject. In situations where different cues are in conflict with each other, subjects are known to suppress and/or recalibrate discordant cues flexibly (Murphy 1996). How exactly this occurs is an issue that has received little attention.

One exception can be found in research on self-organized cue integration. In many real-world situations, the usefulness of different sensory cues or modalities for a certain task will change over time. Cues may sometimes be conflicting or may require recalibration. Unfortunately, it is usually not clear a priori which cues are to be trusted and which ones should be suppressed or recalibrated. Triesch and von der Malsburg (2001) proposed the idea of *democratic cue integration*. In democratic integration, several cues are merged into an estimate of the state of the world, and the result is fed back to all individual cues to drive quick adaptations. Cues that conflict with the agreed-upon result are suppressed and/or recalibrated. The system is simply driven to maintain agreement among the different cues.

While initial work on democratic integration explored the benefits of such a scheme in the context of a computer vision problem of tracking people in video sequences, more recent work has studied the topic psychophysically. Triesch et al. (2002) showed that human subjects who track objects among distractors quickly reweight different cues (e.g., size, color, and shape of the tracked object), depending on the reliability of the cues. This reweighting occurs within one second. The neural basis of this flexible reweighting of different information sources remains a promising topic for future studies.

Conclusion

The integration of different sensory modalities and cues poses a central problem in perception. Although the Bayesian framework has proved very useful in understanding subjects' behavior on a range of tasks, more research is needed to understand the neural implementation of Bayesian inference processes and the approximations that the brain may be using. Furthermore, the learning mechanisms that set up the system to perform in a near-optimal fashion require investigation. Since much of this learning takes place in the context of goal-directed actions, the concept of reinforcement learning can be used to frame enquiry into these issues.

From "Dynamic Coordination in the Brain: From Neurons to Mind,"
C. von der Malsburg, W. A. Phillips, and W. Singer, eds. 2010. Strüngmann Forum Report, vol. 5,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-01471-7.