Review

A model of the temporal dynamics of multisensory enhancement

Benjamin A. Rowland *, Barry E. Stein

Wake Forest School of Medicine, Winston-Salem, NC 27157, United States

ABSTRACT

The senses transduce different forms of environmental energy, and the brain synthesizes information across them to enhance responses to salient biological events. We hypothesize that the potency of multisensory integration is attributable to the convergence of independent and temporally aligned signals derived from cross-modal stimulus configurations onto multisensory neurons. The temporal profile of multisensory integration in neurons of the deep superior colliculus (SC) is consistent with this hypothesis. The responses of these neurons to visual, auditory, and combinations of visual–auditory stimuli reveal that multisensory integration takes place in real-time; that is, the input signals are integrated as soon as they arrive at the target neuron. Interactions between cross-modal signals may appear to reflect linear or nonlinear computations on a moment-by-moment basis, the aggregate of which determines the net product of multisensory integration. Modeling observations presented here suggest that the early nonlinear components of the temporal profile of multisensory integration can be explained with a simple spiking neuron model, and do not require more sophisticated assumptions about the underlying biology. A transition from nonlinear "super-additive" computation to linear, additive computation can be accomplished via scaled inhibition. The findings provide a set of design constraints for artificial implementations seeking to exploit the basic principles and potency of biological multisensory integration in contexts of sensory substitution or augmentation.

© 2014 Elsevier Ltd. All rights reserved.

Contents

1. Introduction

The evolution of multiple sensory systems has enhanced the likelihood of survival for organisms living in a wide variety of environments. This is not only because the senses substitute for one another when necessary, but because they can interact synergistically, thereby providing far more information about external events than would otherwise be possible. This is because the different senses are not corrupted by the same sources of noise, and combining their conditionally independent estimates of the same event yields a better analysis of its features (Ernst and Banks, 2002). This advantage manifests physiologically as enhancements in the speed and robustness of reactions to concordant cross-modal stimuli (Rowland et al., 2007a; Rowland and Stein, 2008), which in turn lead to faster and more accurate behavioral responses to the originating event (Meredith and Stein, 1983; Gielen et al., 1983; Perrott et al., 1990; Hughes et al., 1994; Frens et al., 1995; Wilkinson et al., 1996; Goldring et al., 1996; Jiang et al., 2002). Such enhancements are particularly beneficial when

* Corresponding author. Tel.: +1 3365779513.
E-mail addresses: browland@wakehealth.edu, benjaminarowland@gmail.com
(B.A. Rowland), bestein@wakehealth.edu (B.E. Stein).

the information provided by the inputs is otherwise impoverished and/or unreliable; that is, circumstances in which their individual utilities are minimized (Stein and Meredith, 1993).

The best studied system in which this occurs is the mammalian superior colliculus (SC), which mediates the detection, localization, and orientation toward environmental targets (Meredith et al., 1987; Stein and Meredith, 1993). Individual neurons within the SC are sensitive to cues derived from different sensory modalities (e.g., vision, audition, and somatosensation) within circumscribed and overlapping regions of space (Stein and Arigbede, 1972). When stimulated by cross-modal cues within their respective receptive fields (RFs), their net evoked response magnitude (i.e., total number of impulses) is elevated above the response magnitude evoked by only one of the cues individually ("multisensory enhancement"). For robust stimuli, this enhancement typically reflects the sum of the net unisensory response magnitudes, but can be greater than this sum when the unisensory responses are less robust.

However, recent analyses examining the temporal profile of multisensory enhancement suggest that this enhancement is not uniform over the duration of the response (i.e., the entire discharge train). As the multisensory response rises and falls, its instantaneous firing rate (IFR) rarely reflects a simple addition of the component unisensory firing rates, even when the overall enhancement in the net response magnitude is consistent with an additive model (Rowland et al., 2007a). Rather, response enhancements are proportionally largest at the beginning of the response, which leads to earlier-than-expected response onsets (Rowland et al., 2007a; Rowland and Stein, 2008). The timing and magnitude of these multisensory enhancements, especially when occurring early in the discharge train, have the potential to greatly influence downstream circuits responsible for overt behavioral responses, as well as other targets involved in more higher-order perceptual processes. The operational principles of these neurons are a subject of great interest to basic scientists and researchers in applied domains seeking to engineer devices for sensory augmentation and substitution. However, most computational approaches to understanding multisensory integration in the SC have been restricted to describing its net products (e.g., Anastasio et al., 2000; Rowland et al., 2007b; Cuppini et al., 2010), not its moment-to-moment operations.

The purpose of this paper is to describe how the nonlinearities evident at the beginning of the multisensory response can be explained by a simple spiking model of SC multisensory integration, and do not require more complex assumptions about the biological substrate. At a coarse temporal resolution, the behavior of this model is similar to those described previously. However, at the level of resolution addressed here, the timing and "shape" of the inputs are revealed as key determinants of the integrated multisensory response. It thereby makes the neurobiological computations underlying the multisensory response more explicit.

2. Results

2.1. Empirical observations

In multisensory SC neurons, concordant cross-modal signals typically evoke responses containing more impulses (i.e., enhanced net response magnitude), higher firing rates, longer durations, and shorter latencies than do their individual component stimuli (Stein and Meredith, 1993). The magnitude of the total multisensory response is generally related to the efficacy of the component stimuli: typically greater than the sum of these constituent unisensory response magnitudes when they are individually weak, and equal to their sum when they are more robust (Meredith and Stein, 1986). Fig. 1A and B provide typical examples from the cat SC. In Fig. 1A, both unisensory responses are very weak, engaging a net

superadditive computation. In Fig. 1B, the unisensory responses are more robust, revealing a net additive integrative computation. These examples are consistent with the "principle of inverse effectiveness", which specifies an inverse relationship between the proportional multisensory enhancement and the magnitude of the response to the most effective modality-specific component stimulus. However, despite the difference in the net products of these two examples, an examination of their temporal profiles (captured by the instantaneous firing rate and cumulative impulse count trace comparisons) reveals similarities in the underlying computational schematic. Importantly, both responses have aspects in which the multisensory response is more robust, and more dynamic, than an additive model would predict.

2.2. Enhancements in magnitude

On a moment-by-moment basis, most enhanced multisensory responses evidence "instantaneously" superadditive computations in portions of the response when the signals are weak or modest (typically at the beginning or end of one of the component responses), additive computations when they are modest, and subadditive computations if and when they are very potent. The net computation evident in the overall response magnitude or firing rate reflects the sum of these linear and nonlinear instantaneous computations; thus, the difference between a net superadditive (Fig. 1A) and a net additive (Fig. 1B) product is related to not just the potency, but the relative incidence of superadditive, additive, and subadditive computations that took place during each.

2.3. Enhancements in timing

Most enhanced multisensory responses are more dynamic than their component unisensory responses. In both examples in Fig. 1, the multisensory response rises to its maximum firing rate more rapidly than predicted by the summed unisensory responses. However, it also descends from its peak earlier than expected, prior to the time at which the summed unisensory responses are expected to peak. Later, in its declining phase, the multisensory response can have a slower dynamic than expected, leading to a longer-than-expected response duration, although this result is variable across samples.

These examples illustrate the computational schematic that underlies the multisensory responses of most neurons in the SC, one in which the relationship between the unisensory and multisensory responses is neither linear nor strictly fixed in time. Clearly, the temporal alignment of the unisensory inputs (reflected in the unisensory responses) is a critical determinant in the instantaneous and overall products. Aligning weak or modest portions of the unisensory responses leads to large, typically superadditive enhancements, while aligning robust portions yields additive or subadditive enhancements. Multisensory responses change quickly when they are robust, and slowly when less robust. These underlying dynamics, only visible at a fine temporal resolution, can have profound consequences for the overall multisensory product achieved in any particular circumstance.

Fig. 2 illustrates the generality of these observations across a population of 324 samples of multisensory and unisensory responses recorded from the cat SC (a subset of the data originally published in Stanford et al., 2005). This analysis is restricted to multisensory neurons that are overtly responsive to brief presentations (50 ms) of visual and auditory stimuli individually and exhibit an enhanced response when the auditory stimulus is presented between 30 and 100 ms after the visual stimulus onset. This window of time typically yields the greatest likelihood for cross-modal interactions (Meredith et al., 1987). The restriction of the analysis to neurons overtly responsive to both modalities reduces