Relative Unisensory Strength and Timing Predict Their Multisensory Product

Ryan L. Miller, Scott R. Pluta, Barry E. Stein, and Benjamin A. Rowland

Department of Neurobiology and Anatomy, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, North Carolina 27157

Understanding the principles by which the brain combines information from different senses provides us with insight into the computational strategies used to maximize their utility. Prior studies of the superior colliculus (SC) neuron as a model suggest that the relative timing with which sensory cues appear is an important factor in this context. Cross-modal cues that are near-simultaneous are likely to be derived from the same event, and the neural inputs they generate are integrated more strongly than those from cues that are temporally displaced from one another. However, the present results from studies of cat SC neurons show that this “temporal principle” of multisensory integration is more nuanced than previously thought and reveal that the integration of temporally displaced sensory responses is also highly dependent on the relative efficacies with which they drive their common target neuron. Larger multisensory responses were achieved when stronger responses were advanced in time relative to weaker responses. This new temporal principle of integration suggests an inhibitory mechanism that better accounts for the sensitivity of the multisensory product to differences in the timing of cross-modal cues than do earlier mechanistic hypotheses based on response onset alignment or response overlap.

Key words: cross modal; enhancement; integration; superior colliculus; temporal

Introduction

The operational principles by which the brain integrates signals from various senses ensure that they are combined in useful ways. For example, the responses of multisensory neurons in the mid-brain superior colliculus (SC), and the detection/localization behaviors they mediate, are markedly enhanced by cross-modal cues that are colocalized and are unaffected or depressed when those cues are spatially disparate (Meredith and Stein, 1986). Colocalized cues from different senses are most likely derived from the same event, whereas disparate cues most likely derive from different, unrelated events. Therefore, sensitivity to the spatial proximity of cross-modal cues is a useful operational principle for multisensory integration in this context and it significantly affects behavioral performance (Stein et al., 1989; Jiang et al., 2002; Burnett et al., 2004; Gingras et al., 2009; Stevenson et al., 2012, but also see Fiebelkorn et al., 2011).

A similar principle for multisensory integration can be intuited for the dimension of time: temporal proximity, like spatial proximity, is a powerful indicator of relatedness. Consistent with this idea, SC multisensory responses are most enhanced by cues

that are near-simultaneous and unaffected or depressed by those that are more disparate (Meredith et al., 1987). The aggregate population results from Meredith et al. (1987) illustrate a relationship between multisensory response enhancement in visual-auditory neurons and the stimulus onset asynchrony (SOA) of these cross-modal cues (i.e., an “SOA tuning function”) that is approximately symmetric, albeit with some variability among individual samples and a bias noted toward larger enhancements when the visual stimulus preceded the auditory. A similar bias has been identified in behavior: visual-auditory stimulus pairs are more quickly and more reliably detected when the visual precedes the auditory (Hershenson, 1962; Diederich and Colonius, 2004). Because auditory transmission delays are shorter than visual delays, this response bias has been interpreted as reflecting a preference of multisensory SC neurons for cross-modal inputs that arrive in a temporally aligned (and thus overlapping) fashion (Diederich and Colonius, 2004). This range of SOAs that are integrated at the single-cell (or behavioral) level defines the “temporal window of integration” (Meredith et al., 1987) and has been assumed to be relatively static (but see Discussion). This is our current mechanistic understanding of why temporal proximity is an effective predictor of the physiological and behavioral measures of multisensory enhancement.

The objective of the present study was to examine this temporal principle of multisensory integration more systematically by determining the impact of variations in the timing and efficacy of visual and auditory stimuli on the responses of cat SC neurons. The results confirm that temporal proximity is a critical factor; however, they also suggest a novel temporal principle of multisensory integration that would not be predicted based on our current understanding of this phenomenon: that multisensory integration is more effective when stronger inputs are advanced

Received Nov. 21, 2014; revised Jan. 29, 2015; accepted Feb. 16, 2015.

Author contributions: R.L.M., B.E.S., and B.A.R. designed research; R.L.M. and S.R.P. performed research; R.L.M. contributed unpublished reagents/analytic tools; R.L.M. analyzed data; R.L.M., B.E.S., and B.A.R. wrote the paper.

This work was supported by the National Institutes of Health (Grants NS036916 and EY016716). We thank Nancy London for technical assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Ryan Miller, Department of Neurobiology and Anatomy, Wake Forest School of Medicine, Grey 4110, 1 Medical Center Blvd, Winston-Salem, NC 27157. E-mail: ryamille@wakehealth.edu.

S. Pluta's current affiliation is Department of Molecular and Cell Biology, Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720.

DOI:10.1523/JNEUROSCI.4771-14.2015

Copyright © 2015 the authors 0270-6474/15/355213-08\$15.00/0

in time relative to weaker inputs. This principle gives an accurate accounting for both the present observations and those made previously, but points to a different underlying mechanism by which the system operates in real time to synthesize inputs from different sensory channels with very different temporal signatures.

Materials and Methods

Protocols. Protocols were in accordance with the National Institutes of Health's *Guide for the Care and Use of Laboratory Animals*, eighth edition (2011). They were approved by the Animal Care and Use Committee of Wake Forest School of Medicine, an Association for the Assessment and Accreditation of Laboratory Animal Care International-accredited institution. Two male cats were used in this study.

Surgical procedure. After administering the anesthetic ketamine hydrochloride (25–30 mg/kg, IM) and the preanesthetic tranquilizer acepromazine maleate (0.1 mg/kg, IM), the animal was transported to a surgical preparation room, given presurgical antibiotics (5 mg/kg enrofloxacin, IM) and analgesics (0.01 mg/kg buprenorphine, IM), and prepared for surgery. The animal was intubated and transferred to the surgical suite, where a surgical level of anesthesia was induced and maintained (1.5–3.0% inhaled isoflurane), and placed in a stereotaxic head holder. During surgery, expired CO₂, oxygen saturation, blood pressure, and heart rate were monitored with a vital signs monitor (VetSpecs VSM7) and body temperature was maintained with a hot water heating pad. A craniotomy was made dorsal to the SC and a stainless steel recording chamber (McHaffie and Stein, 1983) was placed over the craniotomy and secured with stainless steel screws and dental acrylic. The skin was sutured closed, the inhalation anesthetic was discontinued, and the animal was allowed to recover. When mobility was reinstated the animal was placed back in its home pen and given the analgesics ketoprofen (2 mg/kg, IM, sid) and buprenorphine (0.01 mg/kg, IM, bid) for up to 3 d.

Recording procedure. After allowing 7 or more days of postsurgical recovery, weekly experimental recording sessions began. In each session, the animal was anesthetized with ketamine hydrochloride (20 mg/kg, IM) and acepromazine maleate (0.1 mg/kg IM), intubated, and artificially respired. It was maintained for recording in a recumbent position and, to preclude introducing wounds or pressure points, two horizontal head-posts held the head by attaching the recording chamber to a vertical bar. Respiratory rate and volume were adjusted to keep the end-tidal CO₂ at ~4.0%. Expiratory CO₂, heart rate, and blood pressure were monitored continuously to assess and, if necessary adjust, depth of anesthesia. Neuromuscular blockade was induced with an injection of rocuronium bromide (0.7 mg/kg, IV) to preclude movement artifacts, prevent ocular drift, and maintain the pinnae in place. Contact lenses were placed on the eyes to prevent corneal drying and focus the eyes on a tangent screen. Anesthesia, paralysis, and hydration were maintained by intravenous infusion of ketamine hydrochloride (5–10 mg/kg/h), rocuronium bromide (1–3 mg/kg/h), and 5% dextrose in sterile saline (2–4 ml/h). Body temperature was maintained at 37–38°C using a hot water heating pad.

A glass-coated tungsten electrode (tip diameter: 1–3 μm, impedance: 1–3 MΩ at 1 kHz) was lowered to the surface of the SC and then advanced by a hydraulic microdrive to search for single neurons in the multisensory (i.e., deep) layers. The neural data were sampled at ~24 kHz, band-pass filtered between 500 and 7000 Hz, and spike-sorted online and/or offline using a recording system (Tucker-Davis Technologies or TDT). When a neuron was isolated so that it had an impulse height at least 4 SDs above noise (determined online using TDT software), its visual and auditory receptive fields (RFs) were manually mapped using white light-emitting diodes (LEDs) and broadband noise bursts. These were generated from a grid of LEDs and speakers ~60 cm from the animal's head. Testing stimuli were presented at the approximate center of each RF. Stimulus intensity was adjusted to produce weak, but consistent responses from each neuron for each stimulus modality. Stimuli for testing included visual alone (V, 75 ms duration white LED flash), auditory alone (A, 75 ms broadband (0.1–20 kHz noise with a square-wave envelope), and 11 cross-modal combinations of these stimuli with varying SOAs. SOAs varied from A75V (auditory 75 ms before visual) to V175A (visual 175 ms before auditory) in 25 ms steps. In cases in which neurons were maintained for long enough periods, multiple test blocks were run consecutively using different stimulus intensities to create different levels of balance between the two unisensory response magnitudes.

At the end of a recording session, the animal was injected with ~50 ml of saline subcutaneously to ensure postoperative hydration. Anesthesia and neuromuscular blockade were terminated and, when the animal was able to breath without assistance, it was removed from the head-holder, extubated, and monitored until mobile. Once mobile, it was returned to its home pen.

Data analysis. A total of 226 tests were conducted on 143 SC neurons, with some neurons being tested with multiple sets of modality-specific stimulus intensities. Response magnitudes were evaluated as the number of impulses elicited within 500 ms after stimulus onset minus the spontaneous rate (i.e., the number of impulses within 500 ms before stimulus onset). Response onset latency was determined by the three-step geometric method (Rowland et al., 2007).

For each neuron, the relative difference between the response magnitude (i.e., mean number of impulses per trial) to the visual (V) and auditory (A) stimuli was used to quantify its unisensory imbalance (UI) according to a contrast function (Eq. 1). A neuron was classified as having “balanced” sensitivity if the visual and auditory response magnitudes did not significantly differ (two-tailed paired *t* test) and “imbalanced” if they did differ significantly.

$$\text{Unisensory Imbalance} = \frac{V - A}{V + A} \quad (1)$$

The efficacy of multisensory integration as evidenced by a multisensory response (MS) was quantified in two ways. The first method evaluated the proportionate difference between the magnitudes of the multisensory and best (i.e., largest) unisensory responses (“multisensory enhancement,” ME, Eq. 2). A second method evaluated the proportionate difference between the multisensory response magnitude and the sum of the two unisensory responses (“additivity index,” AI, Eq. 3).

$$\text{ME} (\%) = \frac{\text{MS} - \max(V, A)}{\max(V, A)} \times 100 \quad (2)$$

$$\text{AI} (\%) = \frac{\text{MS} - (V + A)}{V + A} \times 100 \quad (3)$$

Relationships between ME and SOA (i.e., the enhancement SOA tuning function) and between AI and SOA (the additivity SOA tuning function) were derived for each test for each neuron. To control for variability in response latencies associated with different sensory inputs, ME and AI were also related to response onset asynchrony (ROA), which is the difference between the expected unisensory response onsets at a particular SOA (e.g., ROA = 0 indicates that the SOA is such that the visual and auditory response onsets cooccur). Because these relationships parallel those involving SOA, they are referred to as ROA tuning functions. For the purposes of averaging data across the samples, the value of the tuning functions between sampled ROA values was derived using linear interpolation between adjacent sampled points.

In some cases, tuning functions peaked at an “optimal” asynchrony value or range, and decreased symmetrically around it. In other cases, the decrease was asymmetric and often fell off far more rapidly near one of the extremes of the range tested. This asymmetry was determined by the slope of a line fit to the tuning function that minimized least-squares error. This slope indicates whether multisensory enhancement was biased to be larger in auditory-first configurations (negative values), visual-first (positive values), or had no preference (values close to zero). The interaction between unisensory imbalance and tuning function asymmetry was studied both across neurons and, in some cases, within neurons tested at multiple stimulus intensities.

A final analysis examined the effect of an interaction between the imbalance of the unisensory responses and their order of occurrence (i.e., stronger response first vs stronger response second) on multisensory

Figure 1. Balanced unisensory activation yields the greatest integrated multisensory product. **A**, Top, Three exemplar neurons illustrate the trend in which the level of imbalance in a neuron's unisensory visual and auditory responses predicts the relative magnitude of its multisensory response to their combination. In each example, the sum of visual and auditory responses is approximately equivalent (horizontal dashed lines labeled "additive"), but as the unisensory response imbalance (shown at the top of each series of bars) grows, multisensory response magnitude decreases. Bottom, This is also evident as decreases in multisensory enhancement (brown) and the additivity index (black) as levels of unisensory response imbalance increase. **B**, Population averages reflect the same relationship. The average multisensory enhancement obtained (left) was significantly higher in neurons with balanced than imbalanced unisensory responses. So too was the additivity index, revealing that the incidence of superadditive integration was far higher in neurons with balanced unisensory responses. $**p < 0.005$.

enhancement. For each test in which the unisensory responses could be categorized as imbalanced (see above), multisensory responses for each ROA between ± 50 and 100 ms were selected and categorically designated as "balanced," "stronger first," or "weaker first" based on the significance and direction of their imbalance scores. ME and AI were compared for each of these groups.

Of the 226 SOA test blocks conducted, 116 (from 76 multisensory SC neurons) met the following criteria for inclusion in this study: recording "isolation" was maintained long enough to present a minimum of 20 trials (typically 30) for each stimulus configuration and both unisensory responses were significantly greater than baseline firing. For purposes of evaluating SOA tuning curve asymmetry (see Fig. 2C), an additional criterion was added: the neuron had to demonstrate significant multisensory enhancement at one or more of the SOAs tested (paired t test, Šidák correction for multiple comparisons; Šidák, 1967), removing an additional 41 tests of the 116 for this particular analysis. This was necessary because, for neurons that did not integrate at any SOA, the slope of the SOA tuning function was assumed to be randomly determined and therefore not meaningful.

Results

Balanced unisensory responses yielded the strongest multisensory enhancement

Approximately half ($n = 59$) of the sample of neurons exhibited unisensory response magnitudes that were not significantly different from one another ($UI \approx 0$, t test) and were thus categorized as "balanced." The remainder ($n = 57$) were categorized as "imbalanced" and further categorized as "visual-dominant" (UI significantly > 0 , $n = 32$) or "auditory-dominant" (UI significantly < 0 , $n = 25$).

The balance between a multisensory neuron's responses to visual and auditory stimuli individually proved to be a powerful predictor of its response to their combination. It also proved to be a critical variable in understanding the neuron's sensitivity to their relative timing (i.e., temporal offset). Although not previously described, simple mathematical reasoning leads one to expect the products of multisensory integration to be sensitive to the balance of a neuron's unisensory responses. Because multisensory enhancement is evaluated relative to the strongest unisensory response (Eq. 2), increasing imbalance can be viewed as a relative reduction in the impact of the weaker modality-specific input and an absolute reduction in the total excitatory input. However, the present findings show that the neural sensitivity to UI is greater than predicted by this mathematical reasoning: SC neurons integrated "balanced" cross-modal inputs significantly more efficaciously than "imbalanced" inputs even when the two configurations produced the same number of impulses (Fig. 1A). This was the case across a wide range of response magnitudes.

Figure 1A illustrates the main effect of unisensory balance in three exemplar neurons: increasing the degree of unisensory imbalance (moving left to right in the figure) was coupled with disproportionate decreases in the multisensory response and the magnitude of the multisensory enhancement produced.

On average (Fig. 1B, left), the balanced samples exhibited ~ 2.5 times the multisensory enhancement obtained in the imbalanced samples (104% vs 39%, respectively, $p < 0.001$, Mann–Whitney U test). This difference remained significant even after controlling for differences in their net unisensory effectiveness ($p < 0.001$, ANCOVA).

Surprisingly, significant ($p = 0.004$, Mann–Whitney U) differences between balanced and imbalanced samples were also evident when multisensory enhancement was calculated relative to the sum of the unisensory response magnitudes (Eq. 2), with average AI scores of 21% versus 2%, respectively. Again, this difference remained significant after controlling for differences in net unisensory effectiveness ($p = 0.02$, ANCOVA). The difference between the AI measurements for the balanced and imbal-

anced samples underscores the inherent nonlinearity of the multisensory computation and demonstrates sensitivity beyond that expected from the mathematical reasoning described above. This reveals a principle of multisensory integration based on response efficacy that operates in tandem with other previously described principles such as inverse effectiveness (Stein and Stanford, 2008). This principle has also recently been documented in the psychophysical domain (Otto et al., 2013).

Unisensory balance determines the sensitivity of multisensory enhancement to timing

The SOA tuning function quantifies the sensitivity of each sample's multisensory product to the relative timing of the visual and auditory components of the cross-modal stimulus. The SOA tuning function, averaged across all samples (Fig. 2A), was approximately Gaussian in shape (least-squares Gaussian fit: peak height: 11.6 imp/trial; peak center: V23A; RMS width: 68 ms) and strongly resembled that earlier reported for "canonical" exemplar neurons (dashed line) and the overall population by Meredith et al. (1987). However, grouping samples by unisensory balance category (auditory-dominant, balanced, or visual-dominant) revealed that the population-averaged function was actually a composite of groups with very different sensitivities.

Neurons with balanced responses, represented by the middle exemplar in Figure 2B, exhibited approximately symmetric SOA tuning functions most similar to the population-averaged function. However, the SOA tuning functions for the imbalanced samples (i.e., either visual-dominant or auditory-dominant) were markedly asymmetric. For the auditory-dominant group (see exemplar, Fig. 2B, left), multisensory enhancement was greatly diminished when the auditory response was delayed relative to the visual (e.g., V150A). For the visual-dominant group (see exemplar, Fig. 2B, right), multisensory enhancement was greatly diminished when the visual response was delayed relative to the auditory (e.g., A50V).

A quantitative analysis of these trends was conducted by relating the UI to the slope of a least-squares linear fit to the SOA tuning function. This provided a measure of its asymmetry, with more negative values indicating greater auditory-before-visual preference and more positive values indicating greater visual-before-auditory preference. These scores were well correlated (adjusted Pearson correlation, $r = 0.45$, $p < 0.001$) at the population level (Fig. 2C). Individual neurons tested with multiple stimulus efficacy levels (dotted connecting lines, Fig. 2C) produced results consistent with the population trend. Therefore, it did not appear that individual neurons were tuned to integrate visual-auditory cues in a particular timing relationship; rather, a neuron's SOA tuning curve could be easily changed by manipulating the stimulus features that altered the balance between those unisensory component responses.

In general terms, the observed correlation between unisensory imbalance and SOA tuning function asymmetry suggests that reducing the effectiveness of one unisensory component in a pair will cause it to be integrated more efficaciously when the stronger stimulus is "early" rather than "late." The remainder of the analysis focused on this novel observation in more detail.

Principle of "stronger first"

The individual exemplars illustrating the sensitivity of the SOA tuning functions to unisensory imbalance (Fig. 2B) were representative of the averaged functions for each group: balanced, visual-dominant, and auditory-dominant (Fig. 3A). For the balanced group, the SOA that produced the maximum enhance-

Figure 2. Sequencing the cross-modal component stimuli in order of their effectiveness yields the greatest multisensory product. **A**, Population-averaged responses (purple line) show the relationship between SOA and response magnitude, thereby illustrating the temporal window of integration. These data correspond well to the exemplar neuron published in Meredith et al. (1987) (black dashed line). **B**, The population average obscures the presence of SOA tuning functions that have very different profiles, as shown by three exemplar neurons. These illustrate the relationship between unisensory response magnitudes and the effect of SOA on the integrated multisensory response. For neuron 1, the unisensory auditory response was stronger than the visual (i.e., imbalanced), and the SOA function was asymmetric. Normalized multisensory responses were strongest when the auditory stimulus preceded the visual (e.g., A50V), reaching its optimum at A25V. It progressively decreased with shorter SOAs, continuing to decrease as the visual preceded the auditory by greater amounts (e.g., V150A). Neuron 3 was also imbalanced, but responded more robustly to the visual stimulus. Therefore, it showed the opposite asymmetric trend: the multisensory responses were weakest when the auditory stimulus preceded the visual and became progressively stronger as the SOA became shorter, continuing to increase as the visual stimulus preceded the auditory by increasing amounts. In the case of neuron 2, the unisensory responses were balanced and there was no asymmetry in its SOA function. Vertical dashed lines are plotted at V50A for comparison between neurons. **C**, Evaluating these two metrics at the population level revealed the consistency of this trend (Pearson correlation, $r = 0.45$, $p < 0.001$), in which neurons provided with the more effective stimulus first yielded the greatest multisensory responses. For reference, the exemplars from **B** are numbered here (1–3) and color coded. Multiple tests within the same neuron are connected by dotted lines and demonstrate that the effects described were evident within, as well as across, neurons.

Figure 3. Similar trends are evident when assessing the integrated multisensory product based on relative timing of the two stimuli (SOA) or the responses they elicit (ROA). **A**, The population of balanced neurons (i.e., $V \approx A$) demonstrated a higher level of additivity (i.e., superadditivity) at each SOA and a more symmetrical function than either set of imbalanced neurons. **B**, Distribution of the differences between visual and auditory response onset latencies. Median difference is 44 ms. **C**, The same trends as seen in **A** are evident when the ROA is evaluated. SOA and ROA curves are plotted on the same graph to demonstrate their similarities (the SOA function is shifted 44 ms to the left according to the median difference in visual and auditory onset latencies). Note that the best response alignment produces equivalent additivity as the best stimulus alignment. **D**, Evaluating the mean additivity index of multisensory responses from imbalanced neurons with an ROA of at least ± 50 ms shows that response order is a significant determinant of multisensory integration: strong-first produces more robust responses than strong-second (t test, $p < 0.005$).

ment was when visual stimulus onset preceded auditory onset by 25 ms (V25A), which matches the maximum identified in the averaged population function (Fig. 2A). Interestingly, this delay also matches the crossing point for the averaged SOA timing functions for the visual-dominant and auditory-dominant groups (Fig. 3A). Therefore, in the absence of any neuron-specific information, V25A is a good “rule of thumb” to maximize multisensory enhancement.

Prior work suggested that the bias in this function toward visual-before-auditory configurations results from substantial intermodality differences in the neural transmission delays before signals reach the SC (Meredith et al., 1987; Diederich and Colonius, 2004). The intermodality difference between these delays can be estimated for each sample by the difference between the visual and auditory response onsets (Fig. 3B). Adjusting each sample’s SOA tuning curve according to this onset latency difference (see Materials and Methods) and then averaging these curves produces average ROA tuning curves (Fig. 3C). For reference, the SOA tuning curves are replotted on the axis after shift-

ing by 44 ms, the median V-A latency difference. Note the close agreement between the two methods, perhaps due to the relative narrowness of the distribution of V-A latency differences. The optimal ROA for balanced samples was near zero (A9V), which is consistent with earlier observations that integration is maximized when the visual stimulus occurs first because of differences in neural transmission time. However, the ROA tuning functions also illustrate the complexities of the temporal principle described here: if unisensory response magnitudes are imbalanced, then multisensory enhancement is maximized when stronger responses are advanced in time relative to weaker responses (“stronger first”) and minimized when stronger responses are delayed (“stronger second”). For balanced samples, the ROA tuning function is symmetric (i.e., there is no modality-specific order preference). Figure 3D summarizes this principle by comparing the mean AI scores obtained from all imbalanced samples in which the response onsets are separated by 50–100 ms, which represent the best test conditions for the hypothesis. These data indicate that the difference in the AI scores between placing the stronger response first versus second (8%) is significant (t test, $p = 0.003$).

Mechanisms underlying the “stronger first” principle

The most intuitive mechanistic hypothesis for the sensitivity of SC multisensory integration to the timing of the stimuli, and the one most strongly advanced based on prior work (Meredith et al., 1987), is that multisensory products are highly sensitive to the degree of temporal overlap of the component unisensory inputs. In this theory, more overlap in the inputs would provide more opportunity for interactions between them and thus greater multisensory enhancement. This “overlap hypothesis” might also account for the observed “stronger first” principle. Because response efficacy is generally correlated with response duration (i.e., stronger responses are often longer duration), overlap would typically be greater when the stronger (longer) response begins first and minimized when the weaker (shorter) response begins first (Fig. 4A). This hypothesis has fundamental merit in that unisensory stimuli presented at very long delays will produce inputs that do not overlap and thus do not interact, whereas those presented at shorter delays will produce inputs that do overlap and will usually produce enhancement. Using the unisensory responses as estimators for the timing of their respective inputs, the overlap hypothesis predicts a positive correlation between the area of overlap between the unisensory spike density functions (i.e., the integral of the overlapping region, in units of impulses, when responses are aligned according to the appropriate SOA) and the number of impulses in the multisensory response above those predicted by an additive computation.

However, the results are not consistent with these predictions. A correlation calculated between these variables across the population (Fig. 4B) and within subgroups (balanced, strong first, strong second) shows that there is no greater degree of additivity when there is greater overlap in the (estimated) unisensory inputs and these have, in fact, a significantly negative correlation (adjusted Pearson correlation; population: $r = -0.24$ strong first: $r = -0.30$, balanced: $r = -0.13$; weak first: $r = -0.23$; all $p < 0.05$). One might expect that this is the result of inverse effectiveness, that is, that stronger responses will tend to have larger areas of overlap and will also tend to integrate less. However, accounting for this by normalizing the response magnitudes still fails to produce a positive correlation, though it does render the negative correlation nonsignificant.

These data suggest that the overlap hypothesis, despite having validity on a fundamental level, fails to appreciate some key factors that quantitatively determine the products of SC multisensory integration. One of these factors is the temporal dynamics of the unisensory inputs and their respective alignments in different samples. Figure 4C illustrates that the mean enhancement magnitudes observed for the imbalance subgroups (balanced, stronger-first, stronger-second) vary in different time ranges of the multisensory response. In all three groups, enhancement is larger around the initial window of overlap between the unisensory inputs (i.e., 20 ms before to 30 ms after second response onset; Fig. 4C, left), previously termed the initial response enhancement or “IRE” (Rowland et al., 2007) and lower in a later window centered about the end of the overlapping portions of the unisensory responses (100–150 ms after second response onset; Fig. 4C, right). In the early window (Fig. 4C, left), stronger-second samples generate more enhancement than stronger-first samples; in the later window, this trend is reversed: stronger-first samples generate more enhancement than stronger-second samples (Fig. 4C, right). When enhancements across all response phases are combined, the stronger first yields a greater response than the weaker first, as noted above.

These results can be described by a conceptual model in which multisensory enhancement results from an interaction between the excitatory and inhibitory inputs activated by the cross-modal stimulus components (Fig. 4D). In this basic model, each component stimulus produces an excitatory input followed by an inhibitory input to the target SC neuron, with the inhibitory dynamics scaling disproportionately with the excitatory dynamics. In imbalanced cases, the timing of the larger inhibitory dynamic associated with the stronger response is critical in determining the enhancement observed within each window

Figure 4. An inhibitory input is consistent with the order effect. **A**, A prevailing hypothesis that may explain the order effect presented here is that placing the stronger response first maximizes overlap between the two unisensory inputs and therefore maximizes multisensory integration. **B**, If true, a positive correlation should be observed between degree of overlap between the unisensory responses (illustrated by the purple overlapping areas of the blue and red “responses”) and superadditivity in the multisensory response. This was not the case, as more overlap actually produced slightly weaker multisensory integration. **C**, A closer look at timing reveals a dichotomy. In the early period around the onset of the second response (“IRE”, left), stronger-first actually produces less enhancement than when the stronger response is second. Later on in the response (right), the roles reverse and stronger first produces significantly more enhancement. **D**, This can be explained by lagging inhibition that follows (and scales disproportionately with) the excitatory input. When the stronger response comes first (left, top), much of the inhibitory input is not able to interact with the second response and is decaying over the course of the multisensory response. In contrast, with the strong second (left, bottom), the early multisensory period experiences strong enhancement due to the inhibition not yet arriving, but the responses in later periods are strongly suppressed when it arrives. This model inherently leads to a “stronger-first” rule (right) that closely resembles the empirical results (Fig. 3C).

(IRE vs Late). In the stronger-first configuration (Fig. 4D, top left), the effect of this inhibition is maximum within the early window (IRE), but substantially reduced in the late window as the inhibitory input subsides. In contrast, in the stronger-second configuration (Fig. 4D, bottom left), there is minimal effect of this inhibition in the early window, but a maximal effect soon thereafter as the inhibitory input arrives. To determine whether this basic schema had predictive validity relative to empirical data (Fig. 3), a simple model was constructed whereby the overlap (evaluated using cross-correlation between each pairing of input) between the three important components (strong excitation, strong inhibition, weak excitation) was calculated and summed at each ROA using the response shapes drawn in Figure 4D (left;

i.e., excitatory is positive, inhibitory is negative). The results are presented in Figure 4D (right), which shows remarkable similarity to the empirical data (Fig. 3C). In the model, the asymmetry seen in both the visual-dominant and auditory-dominant traces result from the differences in the magnitudes of inhibition produced by the two stimuli and their timing; that is, how much of each inhibitory trace overlaps with the period of multisensory integration. The basic results obtained (i.e., balanced inputs are best and stronger first is better than stronger second) are not dependent on the use of specific model parameters. Rather, they are directly derived from the model's assumption that the inhibitory component scales nonlinearly with, and is delayed relative to, the excitatory component.

Discussion

As noted earlier, the process of integrating information across multiple sensory modalities is sensitive to the likelihood that the inputs are derived from a common event (and are thus related) or from different events (and are thus unrelated). However, the determination of interrelatedness is complicated by the fact that different sensory systems have very different operational parameters and there is some debate as to which cue features are useful in making this determination in any particular circumstance (Stein and Meredith, 1993; Murray et al., 2005; Senkowski et al., 2008; Shams and Beierholm, 2010). However, in the context of the SC and its detection/localization computations, it can be inferred that temporal and spatial concordance are powerful indicators of interrelatedness because unrelated stimuli are unlikely to be simultaneous or colocalized. Empirical support for this basic logic has been identified at physiological and behavioral levels. Cross-modal cues that are spatially concordant (i.e., fall within a neuron's overlapping receptive fields) enhance responses, whereas discordant cues either have no effect or depress responses (Meredith and Stein, 1986; Kadunce et al., 2001). Similarly, temporally concordant cues that produce near-simultaneous input traces enhance responses, whereas disparate cues do not (Meredith et al., 1987; Diederich and Colonius, 2004). Although some heterogeneity in these sensitivities has been observed (Kadunce et al., 2001; Carriere et al., 2008), the predictions derived from the basic logic have carried the most predictive power for whether responses to cross-modal cues will be enhanced or depressed. Understanding the normal operation of these processes will provide insights into the abnormalities that might be present when it is disrupted, as it is in individuals with autism spectrum disorder (Brandwein et al., 2013) or dyslexia (Hairston et al., 2005; Blau et al., 2009; Kronschnabel et al., 2014).

According to the basic logic, the temporal sensitivity of multisensory integration is a function of the absolute temporal disparity of the two unisensory neural inputs and the order of arrival should not significantly affect their integrative product. The present data reveal that it is only when the unisensory inputs are "balanced" that their multisensory product depends solely on their absolute temporal offset and generally produce a robust multisensory response (see also Otto et al., 2013). In this special case, there is no need to consider the order of inputs. A more general principle is that larger enhancements are more reliably achieved when the stronger (i.e., more effective) input reaches the neuron first. This principle of "stronger first" most accurately predicts the magnitude of the integrative products of all the cross-modal samples examined, including those with both balanced or imbalanced inputs as well as all possible magnitudes of imbalance. It thereby provides far more predictive power than temporal proximity alone.

The temporal sensitivity of SC multisensory integration also proved to have a fundamental similarity to its spatial sensitivity. As shown previously (Kadunce et al., 2001), spatial concordance between visual and auditory inputs to a given neuron is a requirement for their integration, but this only means that they must fall within their respective receptive fields. There is no systematic relationship between the amount of overlap and the multisensory product. Similarly, as shown here, the cross-modal stimuli must have temporal concordance for them to be integrated, but there is no systematic relationship between the amount of overlap and the magnitude of the integrated product. Although this may at first seem counterintuitive, it is understood to result from the fact that many temporal configurations of two component responses can yield equally high overlap, but the integrative product will often be highly variable. Indeed, the present results underscore the dynamic nature of the SC multisensory integrative process and expand our understanding of the underlying mechanisms. Net multisensory products are derived from the sum of interactions that take place between the input signals on a moment-by-moment basis as the response evolves. These interactions change according to the state of the cross-modal input alignments at each moment in time. Therefore, any predictive model of multisensory integration must also evaluate the interaction of these cross-modal inputs on a moment-by-moment basis.

The critical feature of the current model assumed to underlie these temporal dynamics is an assumption that each sensory cue evokes in the SC nonlinearly scaled and temporally offset excitatory and inhibitory input traces. The timing of the excitatory traces of the cross-modal inputs relative to one another and, importantly, to the inhibitory traces, determines the multisensory computation at each given moment. When the stronger input is advanced in time (relative to the weaker input), its trailing inhibition suppresses enhancement at the beginning of the multisensory response, but is relinquished toward its end. Conversely, when the stronger input is delayed in time, enhancement is stronger at the beginning of the response (before the strong inhibition), but greatly suppressed when the strong inhibition arrives.

The present results also reveal that, contrary to prior assumptions, an individual SC neuron is not committed, or "tuned," to integrate cross-modal cues at a specific temporal offset. When the physical parameters of the stimuli are changed, the neuron's multisensory product also changes and does so in accordance with the principle of stronger-first regardless of its particular sensitivities to those physical parameters. Stated another way: the temporal window of integration for a given neuron (and presumably at the level of behavior as well) is not a static feature, but one that is highly contingent upon the relative potency of the two stimuli. This observation can be extrapolated to make additional empirical predictions. Gross changes in stimulus features that make neurons more responsive as a population (e.g., raising their intensity) should not only change the aggregate computation in predictable fashion, but should have a similar effect on the behavioral consequences (e.g., detection/localization): greater performance benefits should occur when stronger stimuli are advanced in time relative to weaker stimuli.

It is interesting to consider these observations in the context of the impact of cross-modal experience on multisensory integration. Yu et al. (2009) found that exposure to an asynchronous visual-auditory stimulus increased the magnitude and duration of SC responses to the first stimulus (regardless of modality), but not to the second. Therefore, repeated exposure to a particular cross-modal stimulus pairing effectively leads to the stronger-

first arrangement that the present results show to be maximally effective. This provides a potential mechanism whereby neurons can adapt and become maximally responsive to those cross-modal cue relationships that are most frequently encountered in their particular environment and they would likely do so quite readily early in life when these relationships are first encountered (Yu et al., 2010; Xu et al., 2012; Stein et al., 2014) and possibly throughout life as a mechanism for temporal recalibration (Fujisaki et al., 2004; Vatakis et al., 2007; Mégevand et al., 2013). However, these possibilities remain to be explored.

References

- Blau V, van Atteveldt N, Ekkebus M, Goebel R, Blomert L (2009) Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr Biol* 19:503–508. [CrossRef Medline](#)
- Brandwein AB, Foxe JJ, Butler JS, Russo NN, Altschuler TS, Gomes H, Molholm S (2013) The development of multisensory integration in high-functioning autism: high-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cereb Cortex* 23:1329–1341. [CrossRef Medline](#)
- Burnett LR, Stein BE, Chaponis D, Wallace MT (2004) Superior colliculus lesions preferentially disrupt multisensory orientation. *Neuroscience* 124:535–547. [CrossRef Medline](#)
- Carriere BN, Royal DW, Wallace MT (2008) Spatial heterogeneity of cortical receptive fields and its impact on multisensory interactions. *J Neurophysiol* 99:2357–2368. [CrossRef Medline](#)
- Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys* 66:1388–1404. [CrossRef Medline](#)
- Fiebelkorn IC, Foxe JJ, Butler JS, Molholm S (2011) Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. *Exp Brain Res* 213:167–174. [CrossRef Medline](#)
- Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7:773–778. [CrossRef Medline](#)
- Gingras G, Rowland BA, Stein BE (2009) The differing impact of multisensory and unisensory integration on behavior. *J Neurosci* 29:4897–4902. [CrossRef Medline](#)
- Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT (2005) Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res* 166:474–480. [CrossRef Medline](#)
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289–293. [CrossRef Medline](#)
- Jiang W, Jiang H, Stein BE (2002) Two corticotectal areas facilitate multisensory orientation behavior. *J Cogn Neurosci* 14:1240–1255. [CrossRef Medline](#)
- Kadunce DC, Vaughan JW, Wallace MT, Stein BE (2001) The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. *Exp Brain Res* 139:303–310. [CrossRef Medline](#)
- Kronshabel J, Brem S, Maurer U, Brandeis D (2014) The level of audiovisual print-speech integration deficits in dyslexia. *Neuropsychologia* 62:245–261. [CrossRef Medline](#)
- McHaffie JG, Stein BE (1983) A chronic headholder minimizing facial obstructions. *Brain Res Bull* 10:859–860. [Medline](#)
- Mégevand P, Molholm S, Nayak A, Foxe JJ (2013) Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS One* 8:e71608. [CrossRef Medline](#)
- Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365:350–354. [CrossRef Medline](#)
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215–3229. [Medline](#)
- Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15:963–974. [Medline](#)
- Otto TU, Dassy B, Mamassian P (2013) Principles of multisensory behavior. *J Neurosci* 33:7463–7474. [CrossRef Medline](#)
- Rowland BA, Quessy S, Stanford TR, Stein BE (2007) Multisensory integration shortens physiological response latencies. *J Neurosci* 27:5879–5884. [CrossRef Medline](#)
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401–409. [CrossRef Medline](#)
- Shams L, Beierholm UR (2010) Causal inference in perception. *Trends Cogn Sci* 14:425–432. [Medline](#)
- Šidák Z (1967) Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* 62:626–633.
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255–266. [Medline](#)
- Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J Cogn Neurosci* 1:12–24. [CrossRef Medline](#)
- Stein BE, Stanford TR, Rowland BA (2014) Development of multisensory integration from the perspective of the individual neuron. *Nat Rev Neurosci* 15:520–535. [CrossRef Medline](#)
- Stein BE, Meredith MA (1993) *The merging of the senses*. Cambridge, MA: MIT.
- Stevenson RA, Fister JK, Barnett ZP, Nidiffer AR, Wallace MT (2012) Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Exp Brain Res* 219:121–137. [CrossRef Medline](#)
- Vatakis A, Navarra J, Soto-Faraco S, Spence C (2007) Temporal recalibration during asynchronous audiovisual speech perception. *Exp Brain Res* 181:173–181. [CrossRef Medline](#)
- Xu J, Yu L, Rowland BA, Stanford TR, Stein BE (2012) Incorporating cross-modal statistics in the development and maintenance of multisensory integration. *J Neurosci* 32:2287–2298. [CrossRef Medline](#)
- Yu L, Stein BE, Rowland BA (2009) Adult plasticity in multisensory neurons: short-term experience-dependent changes in the superior colliculus. *J Neurosci* 29:15910–15922. [CrossRef Medline](#)
- Yu L, Rowland BA, Stein BE (2010) Initiating the development of multisensory integration by manipulating sensory experience. *J Neurosci* 30:4904–4913. [CrossRef Medline](#)