UNDERSTANDING THE PROCESS OF MULTISENSORY INTEGRATION

by

RYAN MILLER

A Dissertation Submitted to the Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Neurobiology and Anatomy

May 2016

Winston-Salem, North Carolina

Approved By:

Barry E. Stein, Ph.D., Advisor

Benjamin A. Rowland, Ph.D., Co-Advisor

Ann M. Peiffer, Ph.D., Chair

Thomas J. Perrault Jr., Ph.D.

Emilio Salinas, Ph.D.

Terrence R. Stanford, Ph.D.
# TABLE OF CONTENTS

**LIST OF ILLUSTRATIONS AND TABLES** ......................................................... iv

**LIST OF ABBREVIATIONS** .......................................................................... vi

**ABSTRACT** ................................................................................................... vii

**INTRODUCTION** .......................................................................................... ix

**CHAPTER ONE** – Relative Unisensory Strength and Timing Predict Their Multisensory Product ................................................................. 1

*Published in The Journal of Neuroscience, April, 2015*

Abstract .......................................................................................................... 2

Introduction .................................................................................................... 3

Methods ......................................................................................................... 6

Results .......................................................................................................... 13

Discussion .................................................................................................... 28

**CHAPTER TWO** – Multisensory integration uses a real-time unisensory-multisensory transform ................................................................. 36

*Submitted to Neuron, April, 2016.*

Abstract .......................................................................................................... 37

Introduction .................................................................................................... 38

Methods ......................................................................................................... 41

Results .......................................................................................................... 56

Discussion .................................................................................................... 70
CHAPTER THREE – Immature corticotectal influences initiate the development of multisensory integration in the midbrain ........................................ 79

In preparation

Introduction ........................................................................................................ 80
Methods .......................................................................................................... 81
Results ........................................................................................................... 83
Discussion ..................................................................................................... 89

SUMMARY & CONCLUSIONS ..................................................................... 92

CURRICULUM VITAE .................................................................................. 118
LIST OF ILLUSTRATIONS AND TABLES

CHAPTER ONE

Figure 1. Balanced unisensory activation yields the greatest integrated multisensory product………………………………………………………………… 15

Figure 2. Sequencing the cross-modal component stimuli in order of their effectiveness yields the greatest multisensory product…………………………………… 18

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)…………………………………………………………………………… 22

Figure 4. An inhibitory input is consistent with the order effect………………. 27

CHAPTER TWO

Figure 1. The CTM schematic………………………………………………………. 50

Table I. List of parameters of the CTM model…………………………………….. 53

Figure 2. Variation in multisensory products across neurons and SOAs………. 59

Figure 3. Correlations between multisensory and summed unisensory responses………………………………………………………………………………… 60

Figure 4. Changes in the multisensory products over time at different SOAs…. 62

Figure 5. CTM accuracy at the single neuron level………………………………… 65

Figure 6. CTM accuracy at the population level………………………………….. 66

Figure 7. The CTM is highly accurate in predicting the higher-order features of multisensory integration…………………………………………………………….. 68
Figure 8. The CTM explains variations in enhancement magnitude among neurons.................................................................70

Sup. Figure 1. Parameter sensitivity.......................................................... 71

CHAPTER THREE

Figure 1. Anatomical framework............................................................. 85
Figure 2. Exemplars of neural development with age................................. 86
Figure 3. Multisensory neurons appear earlier in SC than in AES............. 87
Figure 4. Unisensory response latencies decrease in both regions............. 88
Figure 5. Unisensory response latencies.................................................. 89
Figure 6. SC and AES unisensory responses become more robust............ 89
Figure 7. SC and AES unisensory responses become more reliable.......... 90
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>Anterior ectosylvian sulcus</td>
</tr>
<tr>
<td>AI</td>
<td>Additivity index</td>
</tr>
<tr>
<td>CTM</td>
<td>Continuous-time multisensory</td>
</tr>
<tr>
<td>DR</td>
<td>Dynamic Range</td>
</tr>
<tr>
<td>ETOC</td>
<td>Estimated time of convergence</td>
</tr>
<tr>
<td>IM</td>
<td>Intramuscular</td>
</tr>
<tr>
<td>IRE</td>
<td>Initial response enhancement</td>
</tr>
<tr>
<td>ME</td>
<td>Multisensory enhancement</td>
</tr>
<tr>
<td>RF</td>
<td>Receptive Field</td>
</tr>
<tr>
<td>ROA</td>
<td>Response onset asynchrony</td>
</tr>
<tr>
<td>SC</td>
<td>Superior colliculus</td>
</tr>
<tr>
<td>SDF</td>
<td>Spike density function</td>
</tr>
<tr>
<td>SOA</td>
<td>Stimulus onset asynchrony</td>
</tr>
<tr>
<td>UI</td>
<td>Unisensory imbalance</td>
</tr>
</tbody>
</table>
Multisensory integration, the process by which the brain integrates inputs from different senses, has enormous survival value, and the search for its guiding principles has yielded substantial insight into this remarkable process. At the single neuron level, responses are more robust to spatiotemporally concordant modality-specific sensory cues (likely derived from the same event) than to either cue alone – an effect that is strongest when the cues are weakest. This multisensory enhancement effect increases event detectability and the likelihood of adaptive responses. Spatially discordant cues are either not integrated, or are integrated to yield depression.

We extended our understanding of the governing principles by finding that the relative strength and timing of modality-specific cues in SC neurons is also a key response predictor (Chapter 1). Multisensory integration is strongest when the component responses are balanced, weaker when they are unbalanced but the stronger is initiated first, and weakest when the strongest is second. The relative strength and timing of the component responses proved to be an important
determinant of the integrated product. This simple finding provided the insight to reexamine this process. The result, detailed in Chapter 2, was a new model that can accurately predict a neuron’s multisensory response on a moment-by-moment basis as it evolves, with only knowledge of its responses to the individual component cues.

Chapter 3 deals with how this process develops during early life. Despite the existence of numerous cross-modal inputs to the SC, SC neurons specifically require convergent unisensory inputs from an area of association cortex (the anterior ectosylvian sulcus, AES) in order to engage this process. However, this area, like all other cortical areas develops much more slowly than their midbrain counterparts. How the complex, adult-like process of multisensory integration develops rapidly in the midbrain, while depending on inputs from a slowly maturing region of cortex was a seeming paradox. This was resolved by finding that the AES unisensory inputs to the SC can be surprisingly immature, yet still provide the critical inputs needed to facilitate SC multisensory integration. Apparently, only their functional, reliable input is required to enable this midbrain process.
INTRODUCTION

One feature shared by virtually all living organisms is the use of multiple sensory systems to appreciate the environment. From single-celled organisms to humans, the ability to transduce multiple forms of energy improves the probability of detecting food, danger and conspecifics. It has strong survival benefits for the individual and the species. This capacity is markedly enhanced by the ability to integrate the information derived from these different senses.

One of the best-studied neural circuits in which multisensory integration occurs is the midbrain superior colliculus (SC) of mammals which participates in the detection and localization of salient events. Individual neurons in the intermediate and deep layers of the SC are responsive to multiple sensory modalities (e.g., vision, audition, and somatosensation) (Stein et al., 1973), and integrate signals from these modalities, so that their response to cross-modal cues (e.g., a visual-auditory pair) can be more/less robust than their response to one of the cues alone ("enhancement"/"depression", respectively) (Meredith and Stein, 1983). Some of the earliest studies of this phenomenon identified several "principles" that described which configurations of cross-modal combinations would be most likely to elicit enhanced or depressed responses (Meredith and Stein, 1983, 1986; Meredith et al., 1987). These include the "spatial principle" (spatially aligned cues yield enhancement, spatially disparate cues yield depression), the "temporal principle" (temporally aligned cues yield enhancement, temporally disparate cues yield depression), and "inverse effectiveness" (less effective cues elicit more proportional
enhancement when combined than more effective cues). These principles are logically consistent with the functional role of the SC: cues that are in spatiotemporal concordance are the most likely to originate from a common source event, and the informational gain conferred by the joint consideration of multiple reports is largest when the individual reports are most ambiguous. These principles served as our understanding of the computational basis of multisensory integration for many years (Stein, 2012).

However, these principles only describe general trends and dependencies of enhancement and depression, they do not describe the internal computations taking place when cross-modal signals are being synthesized. Consequently, while they are useful as general heuristics, they do not provide a way of quantitatively predicting what multisensory product will be elicited by any particular cross-modal pair. More recently we came to the realization that in order to understand these internal computations, the phenomenon of multisensory integration had to be evaluated in a new way. In the past, enhancement and depression were quantified in terms of the total number of neural impulses evoked by cues presented in isolation or combination (Stein and Meredith, 1993). To understand the underlying computation, the process of multisensory integration had to be studied in finer detail – on a moment-by-moment basis, as unisensory inputs were being transformed into a multisensory output. We found in this examination that we could not only understand how multisensory computations were effected in the adult, but gain
greater insight into how the process of multisensory integration is developed in the neonate, both within the SC and in the cortex.

Chapter 1 describes our first foray into this effort, in which our new analytic approach identified a previously undescribed interaction between time and cue efficacy in the determination of the multisensory product. The largest products are achieved when the effectiveness of cues is balanced across the modalities. But when they are imbalanced, a new dependency appears: multisensory products are maximized when inputs conveying the more effective cue are the first to arrive at the target neuron. This observation provided one of the first indications that the temporal structure of the cross-modal inputs were of critical importance in determining the multisensory product, and specifically, that a delayed inhibition was a major factor in the computation.

Chapter 2 describes a continuation of this work, where we used a robust statistical analysis of the moment-by-moment multisensory and unisensory responses of SC neurons to derive two basic principles of the multisensory transform: (1) that inputs from different modalities were integrated continuously and in real-time, and (2) that later portions of the response were indeed shaped by the delayed inhibition suggested by our earlier work. To test these principles for descriptive sufficiency, we embedded them in a continuous time neural network model and demonstrated how they could be used to quantitatively predict the moment-by-moment multisensory response given only knowledge of a neuron's response to the individual component
cues. The predictions of the model were highly accurate and precise, suggesting that the derived principles have, for the first time, sufficiently captured the computational bases of multisensory integration in the SC. Furthermore, the model we developed can be used as a tool to quantitatively identify the way in which individual neurons integrate cross-modal cues (i.e., by defining for the neuron a "multisensory signature"), which will allow future researchers to identify possible sources of variation in multisensory products observed in different preparations, circuits, and circumstances.

Chapter 3 describes the most recent work, in which the focus has been on how this process develops. Neurons in the brain do not integrate cues "by default"; rather, multisensory integration capabilities must be developed postnatally. This development is contingent on experience with cross-modal cues, and happens in different brain regions at different times. While the experiential antecedents of multisensory integration are well-described, the relevant circuitry is not. Is the development of multisensory integration capabilities simply tied to the maturation of the neuron’s unisensory inputs, or is there some higher-order organizational principle in place? To address this question, we recorded neurons from the SC and a multisensory region of the cortex, the anterior ectosylvian sulcus (AES), in the same animals at different stages of development. The SC and the multisensory regions of AES receive input from unisensory neurons within other regions of the AES; and in fact, these unisensory neurons have been shown to be critical for the development and expression of multisensory integration in the SC. However, we
show that multisensory integration capabilities in the SC and AES mature at different times (AES after SC), and that SC multisensory integration capabilities develop while the critical corticotectal inputs from AES are not yet functionally mature. Thus, the SC appears to have special machinery in place to develop multisensory integration capabilities using immature inputs.
References


CHAPTER ONE

RELATIVE UNISENSORY STRENGTH AND TIMING PREDICT THEIR MULTISENSORY PRODUCT

Ryan L. Miller¹, Scott R. Pluta¹, Barry E. Stein¹, and Benjamin A. Rowland¹

¹ Department of Neurobiology and Anatomy
Wake Forest School of Medicine
Medical Center Blvd.
Winston-Salem, NC, 27157, USA

The following manuscript was published in The Journal of Neuroscience 2015, 35(13):5213–5220, and is reprinted here with permission. Some changes have been made to the manuscript in consultation with graduate committee members. Stylistic variations are due to the requirements of the journal. Ryan L. Miller collected and analyzed the data, and prepared the manuscript. Dr. Scott R. Pluta also collected some of the data included here. Drs. Barry E. Stein and Benjamin A. Rowland assisted in planning the experiments, analyzing the data, and preparing the manuscript.
Abstract

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.
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 midbrain superior colliculus (SC), and the detection/localization behaviors they mediate, are markedly enhanced by cross-modal cues that are co-localized and are unaffected or depressed when those cues are spatially disparate (Meredith and Stein, 1986). Co-localized cues from different senses are most likely derived from the same event, while disparate cues most likely derive from different, unrelated events. Thus, 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 (Burnett et al., 2004; Gingras et al., 2009; Jiang et al., 2002; Stein et al., 1989; Stevenson et al., 2012, though see also 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 of these cross-modal cues (i.e., an "SOA tuning function") that is roughly symmetric, albeit with some variability among individual samples, and a bias noted towards 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 (Diederich and Colonius, 2004; Hershenson, 1962). 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 (though 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 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.
Methods

Protocols were in accordance with the NIH Guide for the Care and Use of Laboratory Animals Eighth Edition (NRC, 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 pre-anesthetic tranquilizer acepromazine maleate (0.1 mg/kg, IM), the animal was transported to a surgical preparation room, given pre-surgical 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 days.

Recording Procedure

After allowing seven or more days of post-surgical 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 resired. 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 CO$_2$ at ~4.0%. Expiratory CO$_2$, 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 (lens on eye ipsilateral to recording side was opaque) 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, bandpass filtered between 500 and 7000 Hz, and spike-sorted online and/or offline using a TDT (Tucker-Davis Technologies, Alachua, FL, USA) recording system. When a neuron was isolated so that it had an impulse height at least 4 standard deviations 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 approximately 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 (0.1-20kHz) noise with a square-wave envelope), and 11 cross-modal combinations of these stimuli with varying stimulus onset asynchronies (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.

\[
\text{Unisensory Imbalance} = \frac{V - A}{V + A}
\]

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).

\[
ME (\%) = \frac{MS - \max(V, A)}{\max(V, A)} \times 100 \tag{2}
\]

\[
AI (\%) = \frac{MS - (V + A)}{V + A} \times 100 \tag{3}
\]

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 co-occur). 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 enhancement. For each test in which the unisensory responses could be categorized as imbalanced (see above), multisensory responses for each ROA between ±50-100 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 (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 which 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) the sample of neurons exhibited unisensory response magnitudes that were not significantly different from one another (UI≈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 unisensory response imbalance (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. 1). 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 approximately 2.5 times the multisensory enhancement obtained in the imbalanced samples (104% vs. 39% respectively, $p < 0.001$, Mann-Whitney U). 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% vs. 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 imbalanced 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).
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 roughly 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 (green) as levels of unisensory response imbalance increase. B) The 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.
Unisensory balance determines the sensitivity of multisensory enhancement to timing

The stimulus-onset asynchrony (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 roughly 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 Fig. 2B, exhibited roughly symmetric SOA tuning functions most similar to the population-averaged function. However, the SOA tuning functions for the imbalanced samples (i.e., either visual- 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 unisensory imbalance score (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. Thus, 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.
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 stimulus onset asynchrony (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 which have very different profiles, as shown by 3 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. Consequently, 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.

The 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 enhancement 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). Thus, 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 towards visual-before-auditory configurations results from substantial inter-modality differences in the neural transmission delays before signals reach the SC (Diederich and Colonius, 2004; Meredith et al., 1987). The inter-modality 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 Methods), and then averaging these curves produces average ROA (response onset asynchrony) tuning curves (Fig. 3C). For reference, the SOA tuning curves are re-plotted on the axis after shifting 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, 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). Fig. 3D summarizes this principle by comparing the mean AI scores obtained from all imbalanced samples where 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 vs. second (8%) is significant (t-test, $p = 0.003$).
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 response onset asynchrony (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 $\pm 50$ ms shows that response order is a significant determinant of multisensory integration: strong-first produces more robust responses than strong-second (t-test, $p<0.005$).

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 non-significant.
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. Fig. 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., 2007a) 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 (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 if this basic
schema had predictive validity relative to empirical data (i.e., 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 Fig. 4D (left; i.e., excitatory is positive, inhibitory is
negative). The results are presented in Fig. 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: i.e., 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 non-linearly with, and is delayed relative to, the excitatory
component.
Figure 4: An inhibitory input is consistent with the order effect. A) A prevailing hypothesis which 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 1st 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 isn’t 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) which closely resembles the empirical results (Fig 3C).
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 (Murray et al., 2005; Senkowski et al., 2008; Shams and Beierholm, 2010; Stein and Meredith, 1993). 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 co-localized. 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, while discordant cues either have no effect or depress responses (Kadunce et al., 2001; Meredith and Stein, 1986). Similarly, temporally concordant cues that produce near-simultaneous input traces enhance responses, while disparate cues do not (Diederich and Colonius, 2004; Meredith et al., 1987). Although some heterogeneity in these sensitivities has been observed (Carriere et al., 2008; Kadunce et al., 2001), 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 (Blau et al., 2009; Hairston et al., 2005; 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 impact 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. But, 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 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. Though 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. For example, two highly robust unisensory responses might be separated in time so as to barely interact. Because the responses themselves are so strong, however, this slight interaction could yield an equal area of overlap to a pair of very weak responses which perfectly co-occur. In the first case, strong responses which are separated in time will typically give very weak integration, while in the second case weak responses aligned in time will typically give strong integration. 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. Thus, 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 towards 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. Additionally, stronger stimuli themselves tend to produce stronger and faster responses than do weak stimuli. 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 (Stein et al., 2014; Xu et al., 2012; Yu et al., 2010) and possibly throughout life as a mechanism for temporal recalibration (Fujisaki et al., 2004; Mégevand et al., 2013; Vatakis et al., 2007). However, these possibilities remain to be explored.
References


CHAPTER TWO

MULTISENSORY INTEGRATION USES A REAL-TIME UNISENSORY-MULTISENSORY TRANSFORM

Ryan L. Miller\textsuperscript{1}, Barry E. Stein\textsuperscript{1}, and Benjamin A. Rowland\textsuperscript{1}

\textsuperscript{1} Department of Neurobiology and Anatomy  
Wake Forest School of Medicine  
Medical Center Blvd.  
Winston-Salem, NC, 27157, USA

The following manuscript is currently in submission. Ryan L. Miller collected and analyzed the data, and prepared the manuscript. Drs. Barry E. Stein and Benjamin A. Rowland assisted in planning the experiments, evaluating the data, and preparing the manuscript.
Abstract

The manner in which the brain integrates its different sensory inputs in order to facilitate perception and behavior has been the subject of numerous speculations. Using multisensory neurons in the superior colliculus, this study demonstrates that two operational principles are sufficient to understand how this remarkable result is achieved: (1) unisensory signals are integrated continuously as they arrive at their common target neuron, and (2) the resultant multisensory computation is modified in shape and timing by a delayed inhibition. These principles were tested for descriptive sufficiency by embedding them in an artificial neural network model and using it to predict a neuron’s moment-by-moment multisensory response given only knowledge of its responses to the individual modality-specific component cues. The predictions proved to be highly accurate, reliable, and unbiased, and were in most cases not statistically distinguishable from the neuron’s actual instantaneous multisensory response at any phase throughout its entire duration.
Introduction
Multisensory neurons in the superior colliculus (SC) enhance their sensory processing by synthesizing information from multiple senses (Stein, 2012; Stein and Meredith, 1993). When cross-modal (e.g., visual-auditory) signals are in spatiotemporal concordance, as when derived from the same event, they elicit enhanced responses and the originating event is more robustly detected and localized (Meredith and Stein, 1983; Meredith et al., 1987).

The brain normally develops this capability for “multisensory integration” by acquiring experience with cross-modal signals early in life (Rowland et al., 2014; Stein et al., 2014; Wallace and Stein, 1997; Wallace et al., 2006; Xu et al., 2015). In the absence of these antecedent experiences and the changes they induce in the underlying neural circuitry, the net response to concordant cross-modal stimuli is no more robust than to the most effective individual component stimulus; i.e., the neuron’s "default" multisensory computation reflects a maximizing or averaging of those unisensory inputs (Alvarado et al., 2008; Jiang et al., 2006; Stein et al., 2014). The epigenetic acquisition of this capacity, its nonlinear scaling, and its functional utility have attracted much attention (Alais and Burr, 2004; Anastasio and Patton, 2003; Bürck et al., 2010; Colonius and Diederich, 2004; Cuppini et al., 2012; Ernst and Banks, 2002; Morgan et al., 2008; Ohshiro et al., 2011; Rowland et al., 2007b), but its biological bases remains poorly understood.

In part, this is because efforts to understand this process have focused on the generalized, "canonical" relationship between net multisensory and unisensory
response magnitudes. These are abstract quantities calculated by averaging together many neurons' responses to stimuli (i.e., impulse counts) measured over long temporal windows. This general relationship is not a direct indicator of the actual multisensory transform as it occurs on a moment-by-moment basis, and as individual neurons communicate their multisensory products to downstream neurons. It merely reflects aggregate relationships quantified in an empirically convenient fashion. It is also sufficiently abstract as to be reproducible by any number of "biologically plausible" models. But because they are based on these abstract, averaged quantities, such proposals are loosely constrained, have limited predictability, and fail to capture the inherent variation in multisensory products among neurons at different times. A comprehensive analysis of the statistical properties and dynamic features of the multisensory computation is needed to appreciate the actual operating constraints of the biological mechanism.

The present effort sought to do just that. The operating principles of the multisensory transform were inferred from empirical data gathered here, and to determine whether they fully described this transform, they were put to a stringent test. They were used to predict the moment-by-moment multisensory responses of individual neurons given only knowledge of each neuron's response to its modality-specific component stimuli. This was accomplished by embedding the principles in a simple neural network model, the continuous-time multisensory model (CTM, Fig. 1). Despite being highly constrained by a small number of relatively inflexible parameters (and even when fixing those parameters across the population), the
model proved to be highly accurate and precise in predicting the moment-by-moment multisensory responses; thus, it demonstrated that these operating principles provide a complete description of the multisensory transform as it operates in real time. Although this approach was developed for describing SC multisensory integration, the principles identified here are likely to be common among neurons throughout the nervous system for integrating their inputs, whether within a given sensory modality or across different sensory modalities.
Methods

**Electrophysiology:**

Protocols were in accordance with the NIH Guide for the Care and Use of Laboratory Animals Eighth Edition (NRC, 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 Procedures:** Each animal was anesthetized and tranquilized with ketamine hydrochloride (25-30 mg/kg, IM) and acepromazine maleate (0.1 mg/kg, IM). It was then transported to the surgical preparation room, where it was given prophylactic antibiotics (5 mg/kg enrofloxacin, IM) and analgesics (0.01 mg/kg buprenorphine, IM) and intubated. It was then transferred to the surgical suite. Deep anesthesia was induced and maintained (1.5-3.0% inhaled isoflurane), and it was placed in a stereotaxic frame. During the surgery expired CO$_2$, 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 covered with a stainless steel recording cylinder (McHaffie and Stein, 1983), which was secured to the skull with stainless steel screws and dental acrylic. The skin was sutured closed around the implant, the anesthetic discontinued, and the animal was allowed to recover. When mobility was regained the animal was returned to its home pen and given analgesics (2 mg/kg ketoprofen, IM, sid; 0.01 mg/kg buprenorphine, IM, bid) for up to 3 days.
Recording Procedures: After at least seven days of recovery, weekly experimental recording sessions began. In each session the animal was anesthetized and tranquilized with ketamine hydrochloride (20 mg/kg, IM) and acepromazine maleate (0.1 mg/kg IM), intubated and artificially resired, and secured to a stereotaxic frame in a recumbent position by attaching two head posts to the recording chamber. Respiratory rate and volume were adjusted to keep the end-tidal CO$_2$ at ~4.0%. Expiratory CO$_2$, 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 and fix the orientations of the eyes and ears. Contact lenses (lens on eye ipsilateral to recording side was opaque) were placed on the eyes to prevent corneal drying and focus them on the tangent plane where LEDs were positioned. Anesthesia, neuromuscular blockade, and hydration were maintained by intravenous infusion of ketamine hydrochloride (5 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 pad.

A glass-coated tungsten electrode (1–3 MΩ impedance at 1 kHz) was manually lowered with an electrode manipulator to the surface of the SC and then advanced by hydraulic microdrive to search for single neurons in its multisensory (i.e., intermediate and deep) layers. The neural data were sampled at ~24 kHz, bandpass filtered between 500 and 7000 Hz, and spike-sorted online and offline using a TDT
(Tucker-Davis Technologies, Alachua, FL, USA) recording system and OpenSorter software. When a visual-auditory neuron was isolated (amplitude 4 or more standard deviations above background), 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 approximately 60 cm from the animal’s head.

Testing stimuli were presented at the approximate center of each RF. Stimulus intensities were adjusted to evoke weak but reliable responses from each neuron for each stimulus modality. Because the goal of this study is to use unisensory responses to predict multisensory integration, only neurons that were reliably responsive to both visual and auditory cues individually were included in the study. This required that each unisensory response was at least 1 impulse/trial above the spontaneous activity and that this difference was statistically significant (paired t-test). Stimuli for testing included a visual stimulus alone (V, 75 ms duration white LED flash), an auditory stimulus alone (A, 75 ms broadband [0.1-20kHz] noise with a square-wave envelope), and their cross-modal combinations at varying stimulus onset asynchronies (SOAs). From these, 3 SOAs were chosen as representing the range of most robust integration across the population: V0A (simultaneous), V25A (auditory lagging visual by 25ms) and V50A. 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 demonstrably able to breathe without assistance, it was removed from
the head-holder, extubated, and monitored until ambulatory. Once ambulatory, it was returned to its home pen.

**Data Analysis:**

Analyses were divided into two sections. The first analyzed the empirical data to reveal the operational principles of the moment-by-moment multisensory transform. The second embedded these principles within an artificial neural network model (the CTM) and tested whether they could be used to accurately predict the moment-by-moment multisensory responses of any given neuron.

**Evaluation of the Multisensory Transform:** The computations underlying the SC multisensory response (i.e., the manner in which various unisensory inputs are combined to evoke a multisensory discharge train) are typically inaccessible to direct probes. However, the operational principles of the moment-by-moment multisensory transform can be inferred from the overt responses of the neuron to modality-specific cues presented individually (i.e., its unisensory responses) and in combination (i.e., its multisensory response). This is because the impulse generator of the SC neuron that transforms its inputs to its outputs (i.e., its responses) is reasonably assumed to be source-agnostic; that is, it is a transform that is consistently applied to all inputs, regardless of whether they are unisensory or multisensory. Consistently-applied transforms do not alter the statistical relationships between variables (Howell, 2014); thus, the relationships between the moment-by-moment unisensory and multisensory responses reflect, on the other
side of the impulse generator, relationships between the moment-by-moment unisensory inputs and their multisensory synthesis (viz. the multisensory transform).

Thus, the basic objective was to quantify the relationships between the unisensory and multisensory responses at a high temporal resolution ("moment-by-moment", operationally: 1 ms resolution) relative to stimulus onset. These comparisons required shifting each of the recorded unisensory responses to align stimulus onsets with those in the matching multisensory condition. For example, in comparisons involving the V0A (i.e., simultaneous) condition, the multisensory response at t=60 ms after stimulus onset was compared with the sum of the unisensory visual and auditory responses at t=60 ms. In comparisons involving the V50A condition, where the auditory stimulus was delayed by 50 ms relative to the visual stimulus, the multisensory response at t=60 ms was compared with the sum of the unisensory visual response at t=60 ms and the unisensory auditory response at t=10 ms after the auditory stimulus onset.

For population analyses, samples were grouped by SOA condition and synchronized to the "estimated time of convergence" (ETOC). The ETOC was an estimate of when input from both the visual (V) and auditory (A) modalities first converged onto the target neuron in the cross-modal stimulus condition, and thus, when the multisensory transform could begin. To find this time, the two unisensory response latencies (LV and LA) were added to the two stimulus onset times (SV and SA) and the maximum selected (i.e., ETOC=max(SV+LV,SA+LA)).
To obtain quantitative measures of "instantaneous" response efficacy, the binary impulse raster matrices (1 ms resolution) for each response were converted to spike density functions by convolving each trial with a narrow Gaussian kernel (N(0,8 ms)) and calculating mean and variance across trials. Response onsets and offsets (relative to stimulus onset) were calculated from each impulse raster using a three-step geometric method (Rowland et al., 2007). In addition, a linear splining technique was applied to each response trace (slope at time t estimated as the slope of a line connecting the mean spike density function at t-10 ms and t+10 ms) to generate an estimate of its "instantaneous" slope.

A temporal correlation between each mean multisensory response and the summed unisensory responses was calculated as the $R^2$ between firing rate traces paired by timestep (1 ms resolution). Because conventional statistical thresholds require that observations be independent and identically distributed (invalid for data sampled on adjacent timesteps), a bootstrapping procedure was used to evaluate the statistical significance of this $R^2$ value. The sampling distribution of $R^2$ expected if the multisensory and summed unisensory responses were uncorrelated was calculated by repeatedly (10,000 times): (1) randomizing the times of the impulses in the multisensory impulse raster, (2) calculating the resultant spike density function, and (3) calculating the $R^2$ between the now-randomized multisensory and unisensory spike density functions as before (i.e., paired by timestep). The Z-score of the actual $R^2$ value relative to this bootstrapped distribution was evaluated using conventional
thresholds (Z=1.96 for alpha=0.05). Another correlation was calculated between the multisensory and summed unisensory responses on each timestep. For this, samples were grouped by SOA and synchronized by the ETOC, then on each timestep $R^2$ was computed between the firing rates, paired by neuron. Significance was determined using conventional standards (i.e., alpha=0.05). Identical analyses were applied to the estimated instantaneous slopes of the responses.

To examine trends in the multisensory/unisensory magnitude difference, the multisensory and summed unisensory response traces were aligned and directly subtracted from one another. The resulting difference trace (which indicated nonlinearity in the multisensory product) was assessed for amplitude within different windows of time (t-tests) and changes in amplitude over time (regression). This often exposed a phenomenon referred to as the IRE (“Initial Response Enhancement”) (Rowland et al., 2007a). Multisensory amplification was further quantified by calculating the proportionate difference between the mean total multisensory response magnitude (#impulses) and either the largest associated unisensory response magnitude (“Multisensory Enhancement,” ME) or the sum of the unisensory response magnitudes (“Additivity Index,” AI).

\[
ME = 100 \times \frac{VA - \text{max}(V,A)}{\text{max}(V,A)}
\]  

(1)

\[
AI = 100 \times \frac{VA - (V+A)}{V+A}
\]  

(2)
All response magnitudes were corrected for levels of spontaneous activity as in earlier studies (Alvarado et al., 2009). ME and AI were calculated for the entire response and also within two windows, one representing the early window of multisensory interaction which contained the IRE (ETOC-20ms to ETOC+30ms) and the other the remaining portion of the response. The deviation of ME and AI from zero and differences between them across these windows of time were evaluated statistically using t-tests.

The CTM: The principles inferred from the empirical analyses above were tested for validity and completeness by embedding them in an artificial neural network model and evaluating whether this model could accurately and reliably predict the multisensory response of a given neuron in an unbiased fashion given only knowledge of the associated unisensory responses of the neuron (Fig. 1). This type of test is conceptually parallel to other statistical regressions. For example, just as a simple linear regression between two variables is used to support a (linear) model mapping from variable X to variable Y according to some free parameters (slope and intercept), here a (nonlinear) model (the CTM) maps from unisensory responses to multisensory responses according to some internal parameters (n=3). In both cases the validity and completeness of the model is supported when the mapping errors are small and un-patterned.
**Fig. 1: The CTM schematic.** Visual (bird) and auditory (accompanying song) inputs are integrated continuously as they arrive at their common target neuron. The multisensory input drives the neuron’s impulse generator (dark gray ratchet) to produce its multisensory outputs (purple arrows and graph at the bottom). Note that the moment-by-moment products (impulses) of this integrated output decrease over time. This is due to the influence of a delayed inhibition (orange) which suppresses later portions of the response.

A leaky integrate-and-fire single neuron model (I/F) is the simplest model of an impulse generator. The output of a model I/F neuron ($S(t)$) produced by a continuous "net input signal" ($I(t)$) and stochastic zero-mean Gaussian-distributed "noise input" ($N(0, \sigma)$) is determined by a membrane potential ($V(t)$) with time.
constant and a rule that clamps \( V(t) \) to rest (0) for some refractory period (1 ms) after it exceeds an arbitrarily defined threshold (\( V(t) > 1 \)):

\[
\tau \frac{dV}{dt} = -V(t) + I(t) + N(0, \sigma)
\]

(3)

if \( V(t) > 1 \) \{\( V(t + r) = 0 \) for \( r \leq (0: R) \); \( S(t) = 1 \); \} else \( S(t) = 0 \)  

(4)

The "noise input" \( N(0, \sigma) \) represents random fluctuations of input and is randomly sampled on each moment in time. Eqs. 3 and 4 can be numerically simulated from an initial condition to produce an output impulse train in response to any arbitrary input \( I(t) \). It can be simulated many (10,000) times from randomly-selected initial conditions \( (V(0) \in [0,1]) \) and the resulting impulse raster convolved with a narrow Gaussian kernel \( N(0,8 \text{ ms}) \) and averaged across trials to estimate a mean response \( S \) to arbitrary input \( I \). This is referred to as the "forward model transform": \( S = M(I|\tau, \sigma) \). The prediction for the multisensory response \( S_{VA} \) is generated from \( I_{VA} \), which is calculated from the responses of the neuron to the modality-specific components using an equation with three components:

\[
I_{VA} = (I_V + I_A) \times H_{VA}
\]

(5)

The first two components represent the input from the individual modalities \((I_V \text{ and } I_A)\) combined according to the simplest possible rule: summation. The third component \((H_{VA})\) represents the influence of a delayed inhibition. Due to the way
they are estimated (see below), \( I_V \) and \( I_A \) account for both excitatory and inhibitory influences. \( H_{VA}(t) \) is an additional inhibition present in the multisensory condition that scales with (i.e., counteracts) the more robust overt multisensory response. In reality, the magnitude of this component might depend on the preceding overt response magnitude and reflect a recurrent inhibition. However, to maintain a conceptual parallel with other regression techniques and mitigate concerns of overfitting, a simple formula for its \textit{a priori} calculation is provided based only on the observed unisensory responses (i.e., without knowledge of the multisensory response):

\[
H_{VA} = \left( 1 + \frac{h \times (\alpha(15\text{ms}) \ast (M(I_V+I_A | \tau,\sigma)-(M(I_V | \tau,\sigma)+M(I_A | \tau,\sigma))))}{I_V+I_A} \right)^{-1}
\]

(6)

In Eq. 6, \( \alpha \) represents an alpha function with a rate parameter of 15 ms used to filter (i.e., smooth) a predictive error calculation in the numerator (\( \ast \) denotes the operation of convolution). Parameter \( h \) controls the strength of this inhibitory component. Eqs. 3-6 represent a complete system for predicting the multisensory response given estimated unisensory inputs \( I_V \) and \( I_A \) and according to parameters \( \tau, \sigma, \) and \( h \). It is numerically simulated with temporal resolution \( \Delta t=0.1 \text{ ms} \) using the exact solution of Eq. 3 for a single time step (Koch and Segev, 2003).
<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Description</th>
<th>Optimal Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ</td>
<td>Time constant controlling the speed with which the model reacts to changing input</td>
<td>5-10 ms</td>
</tr>
<tr>
<td>σ</td>
<td>Random fluctuations of input</td>
<td>1.5-2.5</td>
</tr>
<tr>
<td>h</td>
<td>Strength of delayed inhibition</td>
<td>0.0064-0.0128</td>
</tr>
</tbody>
</table>

Table I. List of parameters of the CTM model, a description of what each represents, and the range of their optimal values.

Conceptually, the unisensory inputs are estimated by applying the inverse model transform to the overt unisensory spike density functions:

\[ I_V = M^{-1}(IFR_V|\tau, \sigma) \]  

\[ I_A = M^{-1}(IFR_A|\tau, \sigma) \]

This is accomplished numerically by applying an iterative algorithm to the recorded unisensory spike density functions. In generic form, this algorithm infers input \( I(t) \) from observed spike density function \( IFR(t) \) according to parameters \( \tau \) and \( \sigma \).

Step 1 is to derive an initial estimate of the spontaneous input \( (I_{spont}) \). The average firing rate in the time window preceding stimulus onset is calculated \( (R_{spont}) \). For each value of \( \tau \) and \( \sigma \), the most likely value of \( I_{spont} \) to produce \( R_{spont} \) is found using a search algorithm that estimates the model's f-I curve. Starting from an initial
estimate for \( I_{\text{spont}} \), Eqs. 3-4 are simulated many (100) times from random initial conditions to estimate the firing rate it produces (estimated in a 1 s window). The estimate of \( I_{\text{spont}} \) is then raised or lowered depending on whether the model prediction overestimated or underestimated \( R_{\text{spont}} \), and the simulations are repeated. This procedure is repeated until the model-predicted firing rate is within 0.2 impulses/s of \( R_{\text{spont}} \).

\( I_{\text{spont}} \) provides the initial condition for the estimated input current at stimulus onset \((I(0)=I_{\text{spont}})\). Then, for each future time point \((T>0)\), there is an initial estimate for \( I(T)=I(T-1) \), Eqs. 3-4 are simulated from \( t=0 \) to \( t=T \) many (10,000) times from random initial conditions, and the mean value of \( S(T) \) is calculated. \( I(T) \) is raised or lowered based on whether the mean \( S(T) \) under- or over-predicts IFR(T). The value of \( I(T) \) is fixed when the mean of \( S(T) \) is within either 1% or 0.5 imp/s of IFR(T), then the procedure moves to the next timestep.

Applying this procedure to the unisensory visual and auditory responses derives the currents \( I_V(t) \) and \( I_A(t) \) for each value of \( \tau \) and \( \sigma \). For each value of \( \tau \), \( \sigma \), and \( h \), Eqs. 3-6 are then numerically simulated. This is done repeatedly (10,000 times) with random initial conditions in order to derive a model-predicted multisensory response trace based only on the unisensory responses.

**Evaluation of the CTM:** The model was evaluated for accuracy, reliability, and bias. Each evaluation compared the actual multisensory response to the model prediction.
of it (based on the unisensory responses and best parameters) on each (1 ms) timestep in the response window. In order to provide a referent baseline against which to evaluate "good" vs. "poor" performance, we simultaneously compared the actual multisensory response to an "additive model" formed by summing the unisensory responses. Like the CTM, this model predicts an instantaneous correlation between the multisensory and unisensory responses; thus, it does not represent anything like a null hypothesis. Its invalidation does not necessarily support CTM, nor does a failure to invalidate it call the CTM into question, since in certain circumstances the CTM's predictions are equivalent to those of the additive model. Rather, its purpose here is to provide an easily-understandable referent in order to interpret the scale of the performance metrics.

Accuracy was assessed by comparing, on a neuron-by-neuron and moment-by-moment basis, the model-predicted firing rate to the sampling distribution of the mean firing rate estimated from the actual multisensory response (the error t-score). The mean and variation of the error t-scores were computed across the population at each moment in time and compared between the CTM and the additive model. For an individual sample, if a model prediction was within the central 95% of this distribution (i.e., "statistically indistinguishable" from the actual multisensory response) it was considered to be "practically equivalent". Reliability was assessed by the standard error of the distribution of error t-scores at each moment in time. Bias (i.e., whether a model consistently over-predicts or under-predicts) was determined at each moment by calculating whether the model-predicted response
was greater, lesser, or within 10% of the actual multisensory response. These classifications were assigned values of +1, -1, and 0 respectively, and the average of these scores across neurons (for each moment) revealed any relevant trends in bias.

Successful prediction of the moment-by-moment dynamics of the multisensory response should, in principle, also mean successful prediction of net or total response metrics (e.g., total response magnitude) and thus, any higher-order features based on them: ME, AI, and sensitivity of the multisensory response magnitude to SOA for individual samples. This was evaluated by correlating these values calculated for individual samples with the model predictions of them.

The next analysis identified how many free parameters were necessary to achieve "good" model fits to the data (based on minimization of sum-of-squares [SOS] error). This involved "fixing" different sets of parameters (e.g., $\tau, \sigma$; $\sigma$ and $h$; all three) to their "population-best" values (i.e., minimizing total error over the population) while allowing other parameter(s) to be freely fitted to individual samples. The additive model was again used to provide a referent for accuracy. Alternative methods of choosing the best parameters were explored (e.g., using best parameters for same neuron but different SOA). The final analysis examined whether the parameters optimally-fit for different neurons could be used to explain observed variation in the magnitude of the multisensory response.
Results

Overview: The responses of individual visual-auditory multisensory SC neurons (n=86) were recorded and analyzed to determine each neuron’s moment-by-moment unisensory-to-multisensory transform. These 86 neurons were tested at three SOAs each, yielding 258 multisensory responses. Of these 258 responses, 172 (67%) exhibited multisensory integration, defined as a multisensory response significantly greater than the best unisensory response. These 172 tests were used for the majority of analyses, which revealed two operating principles of the multisensory transform: (1) unisensory inputs are integrated continuously and in “real-time” (i.e., virtually instantaneously), and (2) there is a delayed inhibition. These principles were embedded in an artificial neural network model (continuous-time multisensory model, CTM) which used the unisensory responses of a neuron to visual and auditory cues to predict the neuron’s moment-by-moment multisensory response to the combination of those cues. The model was then used to explain observed variation in multisensory products.
Fig. 2: Variation in multisensory products across neurons and SOAs. In these 3 exemplar neurons (left-right), the relationship between the timing of the visual and auditory inputs and their synthesized multisensory product was examined. Three SOAs were employed with the visual stimulus preceding the auditory: 0 ms (top), 25 ms (middle) and 50 ms (bottom). In each of the panels, the unisensory response spike density functions (V, A) were plotted, as was their arithmetic sum (V+A), and the actual multisensory response they yielded in combination (VA). Note that even the smallest change in the relative timing of the cross-modal component stimuli substantially altered the alignment of their inputs, and significantly changed the magnitude of their integrated product (i.e., ME, “Multisensory Enhancement”; and Al, the “Additivity Index”). In each case, the enhanced product was greatest during the period of “Initial Response Enhancement” (i.e., the IRE, gray shading, see Rowland et al., 2007). Together, these observations suggest that cross-modal inputs are synthesized in real-time, but that the scale of the multisensory products they generate changes during the course of the response.
Fig. 3: Correlations between multisensory and summed unisensory responses. **A:** Population distribution of z-scores of the correlation between the multisensory and summed unisensory responses collapsed over time (see methods). All were significantly correlated with Z=1.96 or greater (α=0.05 or greater). Gray points indicate locations of the exemplars that were used in Fig. 2. **B:** Same as (A) but for correlations between the "instantaneous" slopes of the multisensory and summed unisensory responses. Virtually all were significant. **C:** Samples were grouped by SOA and synchronized by “Estimated Time of Convergence” (ETOC). The unisensory-multisensory correlation was then assessed within each of the 3 groups (V0A, V25A, and V50A). The coefficient of determination (R²) was significant and high throughout the duration of the multisensory response (the gray shading indicates the region of the IRE. The dashed line indicates the threshold for significance (R²=0.045). **D:** Significant (albeit diminished) correlations were also observed within each cohort between the slopes of the multisensory and summed unisensory responses throughout the course of the response.
**Empirical Analyses:** Each of the neurons was studied with a battery of randomly interleaved tests that included the component stimuli individually and their spatially concordant cross-modal combination at three different ecologically common SOAs: synchronous (V0A), visual 25ms before auditory (V25A) and 50ms before auditory (V50A). The unisensory response latencies (visual=68 ± 10 ms (mean ± std), auditory=22 ± 8 ms) and magnitudes (visual=6.3± 4.3 impulses/trial, auditory=4.8± 3.4 impulses/trial), and estimates of the net multisensory enhancement (ME=105 ± 98%) and additivity index (AI=36 ± 46%), were all consistent with previous reports (Pluta et al., 2011; Stanford et al., 2005; Yu et al., 2009).

The fundamental operating principles of the moment-by-moment multisensory transform were immediately evident from inspection of the unisensory and multisensory responses (spike density functions, see Methods) of several representative exemplars (Fig. 2). First, when responses were aligned by stimulus onset, there was a strong correlation between the dynamics of the summed unisensory and multisensory responses: when the former rose or fell, there were similar changes in the latter, even though their magnitudes differed. This suggested that the unisensory inputs were being synthesized almost immediately upon arrival at the target neuron. Secondly, the enhancement visible in the multisensory response was largest near the “Estimated Time of Convergence”, or ETOC (producing the "Initial Response Enhancement", or IRE, see Rowland et al., 2007), but rapidly diminished thereafter. While there are several ways to interpret the post-
IRE downward trend (see Discussion), the most plausible is that it is due to a delayed inhibitory influence.

Fig. 4: Changes in the multisensory products over time at different SOAs. A: A heat map (rows = neurons, columns = time) illustrates consistency of the IRE (delineated by solid vertical lines) and the transition in the magnitude of the multisensory products across samples grouped by SOA. Red = superadditive multisensory products, Green = subadditive products. Samples were synchronized to the ETOC. Note the high degree of superadditivity within the IRE. B: Normalized, averaged multisensory (VA) and summed unisensory (V+A) responses within, and immediately after the IRE (gray shading) were plotted. As shown in the heat map, the superadditive enhancement was largely confined to the IRE. C: Comparisons of the average firing rate differences between the multisensory and best unisensory component responses (enhancement), and between the multisensory and summed unisensory responses (superadditivity) demonstrated that both were significantly (p<0.001) greater within, than after, the IRE.
These two principles were robustly represented in the population. Across the entire response window, significant correlations (bootstrap test, see Methods) were identified between the magnitudes (Fig. 3A) and slopes (Fig. 3B) of the mean multisensory and associated summed unisensory responses for virtually every sample. At the moment-by-moment level (over neurons grouped by SOA), significant \((p<0.05)\) correlations in response magnitude were evident at all timesteps following the ETOC (Fig 3C), and these did not change significantly at different stages of the response. Significant \((p<0.05)\) correlations in the slopes of the responses were also identified throughout, but these decreased significantly \((p<0.05)\) as responses became less robust and more variable late in the discharge train. There was also high consistency at the different SOAs, so that no significant differences in any of these metrics were noted across SOA groups. Taken together, these observations reveal that the multisensory and unisensory response traces are reliable, that multisensory responses are not \textit{de novo} quantities with arbitrary dynamics, and most importantly, that the multisensory transform synthesizes unisensory inputs virtually instantaneously.

The difference in magnitude between the multisensory and summed unisensory responses was greatest around the ETOC where the IRE was greatest (85\% of samples were net-superadditive during the IRE, see Fig. 4A). This feature was consistent across neurons despite variation among the temporal profiles of unisensory responses, did not depend on SOA, and remained robust at the population level (Fig. 4B). Within the IRE, multisensory responses averaged 40
imp/s on median above the best unisensory response, 15 imp/s above the additive response (Fig. 4C). After the IRE, multisensory responses dropped to 19 imp/s above the best unisensory response and only 5 imp/s above the additive, a trend consistent with the influence of a delayed inhibition.

Indeed, only these two operational principles (real-time integration and delayed inhibition) are required to understand the multisensory transform. This was established quantitatively by embedding them in the simplest possible artificial neural network model (the CTM, see Methods) and using it to predict each individual moment-by-moment multisensory response given only knowledge of the associated unisensory responses.

**CTM Evaluation:** The CTM predictions are shown in Fig. 5 for a highly-representative neuronal exemplar tested at multiple SOAs (Fig. 5A), as well as for neurons with more idiosyncratic response profiles (Fig. 5B). In all cases the model was highly accurate in predicting the temporal dynamics of the multisensory response. As a consequence it also accurately predicted the total magnitude of each response, and thus the net enhancement of the response (ME and AI), as well as variations in its magnitude at different SOAs.
Fig. 5: CTM accuracy at the single neuron level. A: The multisensory responses of this exemplar neuron was matched by the predicted responses at each SOA (left-right), whereas the additive model (dashed gray) had very poor performance near the ETOC (vertical line) when the most robust integration was taking place (i.e., the IRE). Note that model accuracy was high for both total multisensory enhancement (ME) and additivity index (AI). B: Model accuracy remained high even for neurons with atypical response profiles.

The CTM performance was examined for the entire population of neurons studied. To provide an informative benchmark against which to appreciate the scale and pattern of the residual error in the model, its performance was compared with that of a foil "additive model" derived from the sum of the unisensory responses. This foil captured the moment-by-moment correlation between the multisensory and unisensory responses, but did not accurately predict the magnitude of their
difference or the transition from superadditivity to additivity/subadditivity during the response.

Fig. 6: CTM accuracy at the population level. A: Mean error t-score (across neurons) at each moment in time after the estimated time of convergence (ETOC) was quantified as a t-score (larger is worse). The model error (green) is much lower than that of the additive model (gray) throughout the multisensory response, but particularly within the IRE (gray box) where the multisensory enhancement was greatest. **A, inset:** The standard error of the mean t-score for the CTM was also lower than that of the additive model throughout the response, indicating that CTM more precisely captured the variation within the population. **B:** The percentage of CTM predictions (green) and additive model predictions (gray) across the population that are "practically equivalent" (not statistically distinguishable) from the actual multisensory response at each moment in time. Note the high effectiveness of the CTM predictions, especially during the IRE, where the performance of the additive model is particularly poor. **C:** Bias of the CTM (green) and additive model (gray) at each moment in time indicate when model predictions of the multisensory response were consistently too high (positive) or too low (negative). The CTM predictions varied around 0. However, the additive model significantly underestimated the magnitude of the multisensory response, particularly during the IRE. **D:** By regressing the total error of each model's predictions against the sum of the total unisensory magnitudes a pattern of errors was noted for the additive model (gray) that changed as the magnitude of the multisensory response changed. Such a pattern of errors was not present in the CTM predictions.
The mean error of the CTM was low (mean error t-score = 1.1± 0.9) throughout the response window, but lowest (0.9± 0.7) around the ETOC, when the strongest multisensory products were observed (Fig. 6A). The foil additive model behaved quite differently: its error was not only higher throughout the response, but particularly high around the ETOC (mean error t-score = 2.3 ± 1.4), when the integration process was most productive. As shown in Fig 6B, the CTM-predicted multisensory responses were often statistically indistinguishable from the actual multisensory responses, with 85% of them being practically equivalent around the ETOC and 78% practically equivalent across the entire response (versus 46% and 61% for the additive model). Furthermore, the small errors of the CTM were generally unbiased (Fig. 6C), even around the ETOC (-0.01 versus -0.53 for the additive model), and the model was equally accurate for all response efficacies, whereas the additive model was especially poor when responses were weak (Fig. 6D).

Because the CTM could reliably predict the moment-by-moment firing rate of the multisensory response, it could also reliably predict its total response magnitude (R²=0.98). Thus, it could also reliably predict the traditionally-quantified products of multisensory integration (ME [R²=0.91] and AI [R²=0.89], Fig. 7A). Consequently, it could also accurately predict the tuning of these products to SOA in individual neurons (Fig. 7B) (Miller et al., 2015; Rowland et al., 2007a).
Model accuracy was, of course, best when all three parameters were selected optimally for each sample (mean sum-of-squares error ± s.d., SOS=173 ± 64, see Methods). However, τ could be fixed to its population-optimal value (8 ms) and there was only a minor (8.0%) increase in mean error (SOS=187 ± 72) across the population. But even when all parameters were fixed to their population-optimal values (τ = 8 ms, σ = 1.5, h = 0.0016) the CTM model was still significantly more accurate (SOS=253 ± 108) than the additive model (SOS=268 ± 119). Even when all parameters were fixed, the CTM model outperformed the additive model 75% of the time in the region around the ETOC and 70% of the time overall (Supplementary Fig. 1).

Fig. 7: The CTM is highly accurate in predicting the higher-order features of multisensory integration. A: As a consequence of having accurately predicted the moment-by-moment multisensory response, the CTM also accurately predicted the total response magnitude, and thus such traditionally-quantified metrics of multisensory integration as ME (left) and AI (right). Each point in these plots represents a different neuron, with the line of unity indicating an exact match between the CTM and the empirical data. B: The model also accurately predicted the temporal tuning of the multisensory products in each neuron at each SOA tested, as illustrated by these four exemplars with different SOA preferences.
**Using the Model to Predict New Relationships:** The optimal model parameters for a given multisensory response were not arbitrary. A given multisensory response was much more accurately predicted by parameters optimized for different SOA tests from the same neuron (mean SOS=225 ± 127) than another randomly-selected neuron in the dataset (SOS=324± 116), which generated predictions even worse than the additive model. In fact, the fitted values provide new insight into how the SC multisensory response is calibrated.

One long-standing puzzle of interest is in how so-called "non-integrating" SC neurons are responsive to multiple sensory modalities but fail to produce enhanced responses to cross-modal cues presented in spatiotemporal concordance within their respective receptive fields (cf. Stein and Meredith, 1993; exemplars in Fig. 8A). How is it possible to receive two mostly-excitative and independent inputs, but not produce a more robust response than if only one were present? One theory is that, for these neurons (at the time of testing), the visual and auditory input traces were mismatched in some way as to preclude facilitatory interactions. Another possibility is that these neurons lack some critical internal multisensory machinery. Interestingly, the CTM predicts the multisensory responses of integrating (n=172) and non-integrating (n=86) samples with roughly equal accuracy (SOS=168 ± 66 vs. SOS=182 ± 60; t-test, p=0.12). This was initially surprising, given that one of the premises of the model is the linear real-time synthesis of the cross-modal inputs. However, examination of the temporal profiles of enhancement in integrating and non-integrating samples (i.e., the difference between actual multisensory SDF and
additive SDF) revealed very similar patterns, except the non-integrating samples exhibited much larger suppressions later in the response (Fig. 8B). In line with this, non-integrating samples were found to be optimally-fit by larger values of the parameter $h$ (Fig. 8C) (Mann-Whitney U, $p<0.001$). This suggests that interneuronal variability in the amount of inhibition (or sensitivity to it) within the SC may be principally responsible for variation in the products of multisensory integration. Whether the same results would be obtained for non-integrating SC neurons in different populations in which they constitute the majority of samples (e.g., neonatal animals, those raised in the darkness or in omnidirectional noise, or those with compromised cortical inputs: cf. (Wallace and Stein, 2000; Xu et al., 2014; Yu et al., 2010)) is currently unknown.

**Fig. 8:** The CTM explains variations in enhancement magnitude among neurons. **A.** Four atypical exemplar neurons are shown: the top two show subadditive multisensory enhancement, the bottom two show no multisensory enhancement. Nevertheless, all have temporal profiles similar to the typical integrating neurons with a clearly defined IRE. Conventions are the same as in Fig. 5. **B:** These similarities are also apparent at the level of the population of integrating or non-integrating neurons. The most apparent difference between integrating and non-integrating samples appears to be the stronger delayed inhibition in the latter. **C:** The magnitude of the CTM inhibitory scaling parameter ($h$, Y-axis) was highly predictive of the neuron’s integration capabilities as measured by the additivity index.
(the higher $h$ factor the lower the additivity index. Also shown is the best fit line (dashed) to the relationship).

**Supplementary Fig 1: Parameter sensitivity.**

**A.** Fixing various numbers of free parameters (3, 2, 1, 0) shows that even with no free parameters the CTM model is better at predicting the multisensory response than an additive model, and this is true regardless of SOA. 

**B.** SOS error is similarly better than for the additive model, regardless of the number of free parameters. All bars differ significantly (t-test, $p<0.001$). 

**C.** Fixing all parameters based on the best parameters found using the same neuron but at a different SOA (second bar) showed to be far better than simply using the best parameters for the overall population (third bar) or choosing the parameters from a random unit/SOA (fifth bar). Using the tau and noise parameters from the same unit/other SOA and allowing only the inhibition scale to vary freely further reduced the error. All bars differ significantly (t-test, $p<0.001$).
Discussion

In 1952, Hodgkin and Huxley derived a nonlinear continuous-time model of action potential generation from careful study of the statistical principles of this phenomenon in the squid giant axon (Hodgkin and Huxley, 1952). This provided a framework for studying and predicting neuronal behavior. It was followed by a number of other influential works in which the basic biophysical properties of individual neurons, inferred from empirical study, were then tested for their completeness via forward simulation (cf. Churchland and Sejnowski, 1992; Dayan and Abbott, 2001; Gerstner and Kistler, 2002; Koch, 1999; Koch and Segev, 2003; Rieke, 1997). The present study continued that spirit with an objective to predict the products of multisensory integration. Two basic principles of the individual SC neuron’s multisensory transform proved to be essential for predicting these products: (1) cross-modal inputs (visual and auditory) are continuously synthesized in real-time as they converge on their common target neuron, and (2) later portions of the multisensory response are reduced by a delayed inhibition. These principles governing the underlying biological process were exposed via statistical analysis of neural responses to the presentation of visual and auditory cues individually and in combination. To ensure these two principles were sufficient to accurately and completely describe the multisensory transform, they were embedded within a simple neural network model (i.e., the CTM) that was used to predict the actual magnitude and timing of a neuron’s moment-by-moment multisensory (visual-auditory) response based only on its two unisensory responses.
The predictions of the CTM proved to be highly accurate, producing the typical response profile in which the transform yields superadditive products during the early phase of unisensory input convergence, and then gradually transitions to additive products thereafter. Indeed, most of the predictions of the model were practically equivalent to the empirical responses. The model was consistently accurate for neurons with differently-shaped response profiles, as well as for responses obtained from different SOAs from the same neuron. The few predictive errors that were observed were small and unbiased. On this basis we conclude that the model captures the most reliable aspects of the multisensory transform; that is, the identified principles that it embodies constitute a sufficient description of the underlying biological process.

As noted above, the current approach derives the model structure directly from elemental neural response properties. By doing so, it avoids the need to speculate about which of many possibilities are actually at work. It also goes beyond reproducing general relationships between the averages of unisensory and multisensory responses across large populations of neurons, to reproducing the moment-by-moment yield of the multisensory transform in each individual neuron. This, of course, also allows for more global predictions of net multisensory magnitude of a response (i.e., the proportional enhancement of the multisensory response relative to the unisensory responses), and the sensitivity of this enhancement to other factors, such as SOA. In short, the model can be used to predict how a neuron will respond to any combination of visual and auditory cues, at
any SOA and (presumably) under any environmental condition, based only on its unisensory responses. It can also be used as a tool to parameterize the transform implemented by each neuron and compare it across different neurons, cohorts, manipulations.

As a specific demonstration of how the model's predictive power can be used as a tool in data analysis, it was used here to understand how so-called "non-integrating" neurons could be responsive to multiple sensory modalities yet fail to produce a net enhanced response to their combination. The responses of these neurons were predicted with the same accuracy as were those of integrating neurons. The only difference was higher values for h, the parameter controlling the strength of the delayed inhibitory dynamic. This result suggested that these neurons have the same basic capabilities as their counterparts but, for unknown reasons, are more strongly affected by delayed inhibition.

The delayed inhibitory component is an abstract entity here and does not have a 1:1 relationship with any single input, but it is likely to be calibrated by preceding portions of the multisensory response, as is commonly the case elsewhere in the brain. In many cases (e.g., olfactory cortex and visual cortex) principal neurons are reciprocally connected with local inhibitory circuits that enforce a homeostatic "reset" to baseline activity levels following sensory processing events (Francis et al., 1994; Koch and Davis, 1994; Ren et al., 2007). Such a network in the SC could involve AES and basal ganglia (Jiang et al., 2011), but this remains to be determined.
The presence of a superadditive computation near the onset of the multisensory response deserves special attention here. It is a highly characteristic feature of the multisensory response, being seen in nearly all neurons under all conditions (Fig. 4). It is also not, as has been considered (Fetsch et al., 2013; Holmes and Spence, 2005), a phenomenon linked to a threshold device. Rather, this initial response enhancement (IRE, see Rowland et al., 2007) is not strictly limited to the time at which the visual and auditory inputs first converge on their common target neuron and initiate activity, but continues for some time (e.g., 40 ms) thereafter, influencing the most robust portion of the multisensory response. Because in most cases of common origin from events near the subject, visual and auditory cues are likely to have a relatively limited range of SOAs (e.g., 0-100 ms visual-before-auditory), the IRE in most neurons under natural circumstances would be expected to range from approximately 30-140 ms after visual stimulus onset. Given that SC-mediated behavioral responses (e.g., saccades) are often initiated 80-180 ms after stimulus onset (Fischer and Boch, 1983; Sparks et al., 2000), the IRE may be the response phase most important for SC-mediated behavioral decisions (e.g., see also Rowland et al., 2007). The additive/subadditive computations that take place later in the response as the result of inhibition may have little to do with these immediate decisions and reactions, but may support later-occurring processing aims, or a resetting mechanism without informative value.
In sum, the present observations reveal the underlying mechanism by which neurons integrate information across the senses on a moment-by-moment basis, and found that the biophysical mechanisms underlying this process are likely to be much simpler than previously suggested (Anastasio and Patton, 2003; Cuppini et al., 2010; Rowland et al., 2007b). The model explains not just the general relationship between canonical multisensory and unisensory responses, but observed variation in the actual outcomes of multisensory integration over time and across neurons. Thus, the model can be a useful tool moving forward by quantitatively parameterizing the multisensory transform implemented. Here we demonstrated how this parameterization can be used to resolve a long-standing puzzle in the field: understanding the computations engaged by "non-integrating" multisensory SC neurons. It seems reasonable to expect that this approach will be effective with multisensory neurons in other areas of the brain, and with other sensory modality combinations, but this remains to be determined.
References


CHAPTER THREE

IMMATURE CORTICOTECTAL INFLUENCES INITIATE THE DEVELOPMENT
OF MULTISENSORY INTEGRATION IN THE MIDBRAIN

Ryan L. Miller¹, Benjamin A. Rowland¹, and Barry E. Stein¹

¹ Department of Neurobiology and Anatomy
Wake Forest School of Medicine
Medical Center Blvd.
Winston-Salem, NC, 27157, USA

The following manuscript is in preparation. Ryan L. Miller collected and analyzed
the data, and prepared the manuscript. Drs. Barry E. Stein and Benjamin A.
Rowland assisted in planning the experiments, analyzing the data and preparing
the manuscript.
Introduction

Neurons in the cat superior colliculus (SC) develop the capability to integrate cues across sensory modalities (e.g., visual-auditory) in early postnatal life. The developmental chronology of this process is puzzling, because this midbrain integrative process is dependent on inputs from cortex (especially from the anterior ectosylvian sulcus, AES), but it is axiomatic that development of cortex is delayed relative to that of the midbrain, (see Jiang et al., 2001; Wallace and Stein, 2000 for discussion). One possibility is that this axiom is too broad and, in this case, the relevant corticotectal neurons in AES that are critical for this midbrain process (i.e., unisensory AES neurons) reach functional maturity sooner than expected. To examine this question, neurons were simultaneously recorded from the SC and AES in groups of animals at different ages (4-22 weeks). Relatively few studies have examined multisensory integration in the developing cat, and none have compared the development of these brain regions in the same animals. The results showed that the development of SC multisensory integration begins at a time at which AES unisensory neurons are surprisingly immature. Even when SC multisensory integration is effectively adult-like, the visual and auditory latencies of AES neurons are effectively twice what they ultimately will be, their response thresholds are high, and their receptive fields are large and diffuse. Apparently, SC multisensory integration can reach an adult-like state of maturity even while relying on immature, yet functional, corticotectal inputs.
Methods

Surgical and experimental procedures were nearly identical to those described in preceding Chapters. In SC, either a single varnish-insulated tungsten electrode (HJ Winston, 1-3 MΩ impedance) or an 8-channel multi-electrode (Thomas Recording, 0.5 – 3 MΩ, 125 μM spacing) was used to record the activity of individual neurons. In AES, a single tungsten electrode was used. The youngest animals were used acutely, while older animals were used for several experiments (no more than two per week).

Stimuli consisted of 100 ms white LED flashes and 100 ms broadband noise bursts placed in a grid on a plane approximately 60 ms in front of the animals head. Visual and auditory receptive fields were initially mapped in order to find the approximate locations yielding the strongest responses. Next, dynamic ranges were tested using a range of 20 log-spaced values from 0.004 - 6.5 Lux for the visual and 10-60 dB SPL for auditory. Finally, integration was tested by presenting interleaved visual, auditory, and cross-modal cues. The visual and auditory stimulus intensities were chosen as the weakest intensity which produced a significant response in the preceding dynamic range test. Stimulus onset asynchrony (i.e., the time between onset of the visual and auditory components) was 25 ms, visual before auditory. This was found in Chapter 2 to be the optimal SOA for producing multisensory integration.
In total, 2396 SC neurons and 1467 AES were recorded from 36 animals between 31 days and 171 days of age.
Results

Superior colliculus (SC) neurons in the youngest animals had highly variable responses, as shown by the exemplar in Fig. 2 (top left). These neurons were also incapable of integrating their inputs to produce a multisensory response greater than the best response to the individual components (i.e., they lacked "multisensory integration" capabilities). At this age, unisensory AES neurons were also immature, with highly variable responses and high spontaneous rates of activity (Fig. 2, top right). By approximately 14 weeks, SC multisensory neurons became capable of multisensory integration (Fig. 2, bottom left) and AES unisensory neurons exhibited more robust and reliable responses (Fig. 2, bottom right).

Figure 1: A. Schematic showing principal inputs to the SC. Intermediate and deep-layer SC neurons receive "ascending" visual and auditory input from sources other than association cortex, as well as "descending" unisensory input from unisensory neurons in association cortex, including AES. B. Unisensory inputs derived from visual and auditory subregions of AES converging onto multisensory SC neuron (Fuentes-Santamaria et al., 2009).
**Figure 2: Exemplars of neural development with age. Top:** Two immature neurons recorded from an 8-week old animal. The SC neuron (left panel) responded to both visual and auditory cues, but didn’t respond any better to their combination. The responses were also highly variable as can be seen from the size of the standard error bars and the deviations in the spike density functions (middle panel). The unisensory AES neuron (right) showed a similarly variable pattern of spontaneous activity and a weak visual response (blue). **Bottom:** By 14 weeks the neurons in both structures appeared more mature. Spontaneous activity was reduced and more consistent, responses were more robust and had less variability, and the SC neuron demonstrated multisensory integration as the multisensory response was significantly stronger than the best unisensory response (left panel).

Multisensory neurons developed earlier in SC than in AES (Fig. 3A), although they both reached their adult-like integrating capabilities at a similar age (Fig. 3B). Surprisingly, the effectiveness of multisensory integration was observed to be at
adult-like levels virtually as soon as neurons with integration capabilities first appeared in both SC and AES (Fig. 3C).

Figure 3: Multisensory neurons appear earlier in SC than in AES. A. SC multisensory neurons reached their adult-like incidence more rapidly in SC than in AES. B. Although multisensory integration capabilities appeared in SC earlier than in AES, they reach their adult-like incidence simultaneously. C. SC and AES neurons both exhibited adult-like levels of multisensory enhancement as soon as this capability appeared.

There were interesting changes observed in the unisensory properties of these neurons. Response latencies were previously observed to shorten dramatically with age (Stein et al., 1973b). Here it was found that temporal pattern of the response also changes dramatically, and that this was a significant contributing factor to these earlier observations (Fig. 4). Each neuron represented in Fig. 4 was tested with a range of visual and auditory stimuli at 20 different intensities, and responses were averaged across the population. The temporal pattern of these responses was highly complex in young animals, consisting of both an “on” and “off” portion. The strength of the "on" response was extremely weak in the youngest animals, virtually
indistinguishable from spontaneous activity at the level of the individual neuron. Consequently, the response latencies of individual neurons may have appeared to be much longer in previous studies, because only the later ("off") portion of the response was visible. Thus, while the present observations regarding latency changes are in agreement with those of previous studies, but the magnitude of this latency decrease was less than previously estimated because much of the developmental change was in the complexity and variability of the response (Fig. 5).

![Figure 4: Unisensory response latencies decrease in both SC and AES.](image)

A heat-map revealed the dual visual responses in the population of SC and AES neurons to a flashed spot of light. Note that the latencies of both components shortened with age, but the most outstanding change was in the appearance of a reliable first response.
Figure 5: The populations of SC and AES visual and auditory latencies are plotted. Note that both decrease with age.

One of the most dramatic changes in development was a marked increase in response magnitude. Visual and auditory responses in both SC and AES had peak firing rates that increased from less than 5 impulses/s to over 20 impulses/s over the span of a few weeks (Fig. 6).

Figure 6: SC and AES unisensory responses become more robust. Peak firing rate increased several fold over a 24 week period for both the visual (left) and auditory (right) responses in both SC and AES.

The strongest predictor of the development of multisensory integration capabilities was the reliability of the responses. Reliability was assessed by calculating Z-scores (mean firing rate divided by standard deviation of firing rate) for each moment of the
unisensory spike density functions. The peak Z-score for the two unisensory responses was averaged for each neuron as a metric of the average, peak reliability of a neuron’s unisensory responses. Not only did response reliability increase with age in both structures, but integrating neurons had more reliable responses than non-integrating neurons at every age (Fig. 7).

Figure 7: SC and AES unisensory responses become more reliable. **Left:** Unisensory response reliability increased dramatically between 8 and 14 weeks in integrating SC multisensory neurons (closed circles). In comparison, the non-integrating neurons (open circles) only modestly increased their response reliability. Note that at 12-14 weeks, when response magnitude was still comparatively weak in AES neurons (Fig. 6), reliability was already high in SC integrating neurons. **Right:** The same developmental process was evident in AES neurons, albeit was delayed and less pronounced.
Discussion

The primary goal here was to compare the development of the multisensory and unisensory response properties of neurons in the superior colliculus (SC) and the anterior ectosylvian sulcus (AES). It is axiomatic that the development of the AES, as a region of higher-order (i.e., association) cortex, would be delayed relative to that of the midbrain SC. Since SC neurons depend on AES inputs for their multisensory integration capabilities, this developmental discrepancy provided a confusing scenario.

Indeed, the development differences that were expected regarding the maturation of multisensory integration capabilities were observed. This process did appear in SC neurons far earlier than in AES neurons. However, the AES is believed to be more “plastic” and more sensitive to experience than the SC. This seemed to be the case, as once multisensory integration capabilities appeared in some AES neurons, the population matured at a faster rate than it did in the SC. Thus, both areas reached their adult-like status at the same time.

Surprisingly, however, it appeared that by most measures the unisensory response properties of both areas developed in near parallel. Nevertheless, multisensory integration capability, a process usually considered to be highly complex, was apparent in SC neurons at the same time that their unisensory properties were still immature, as were their critical unisensory inputs from AES.
Presumably, then SC neurons do not require that their unisensory component responses to be mature, or their converging unisensory inputs from AES to be mature in order to develop their “complex” multisensory integration capabilities. This is not consistent with the intuition that maturation of neuronal response properties proceeds from simple to complex. The present findings reveal that a more nuanced consideration of sensory development may be warranted.

Finally, it appeared that the reliability of SC unisensory responses (which presumably partially reflects the reliability of its unisensory AES inputs) is the *sine qua non* for the development of SC multisensory integration capabilities.
References


SUMMARY & CONCLUSIONS

While multisensory integration has been studied at the behavioral level for over a century, it is only in the last few decades that the neural correlates of this behavioral effect have come under focused research (Stein and Meredith, 1993). Stein et al. (1973) were among the first to identify and examine the properties of multisensory neurons in the superior colliculus (SC) of cats. The SC was shown to have individual neurons which responded to visual, auditory, and somatosensory stimuli. Shortly thereafter, investigators began to examine how signals conveyed by the different senses (e.g., visual-auditory or visual-infrared) might be combined (Meredith and Stein, 1983; Newman and Hartline, 1981). To describe the multisensory products of this synthesis, terms like multisensory “enhancement” (a multisensory response magnitude that is stronger than the strongest unisensory response) and multisensory “depression” (multisensory response magnitude weaker than the strongest unisensory response) came into use. This is also when multisensory responses first began to be quantified relative to the strongest unisensory response, known today as the measure “Multisensory Enhancement”, or ME. Since that time, multisensory integration has been identified at many levels of the neuraxis (including the inferior colliculus (Bergan and Knudsen, 2009), putamen (Gentile et al., 2011), primary visual cortex (Shams and Kim, 2010), primary auditory cortex (Bizley et al., 2007; Lakatos et al., 2007)) across a wide range of species (human (Murray et al., 2005; Ross et al., 2007; Stein et al., 1996), monkey (Gu et al., 2008; Lakatos et
al., 2007; Wallace et al., 1996), ferret (Bizley et al., 2007), mouse (Olcese et al., 2013), hamster (Meredith and Stein, 1983), guinea pig (King and Palmer, 1985), fish (Pluta and Kawasaki, 2008), snake (Newman and Hartline, 1981), and barn owl (Zahar et al., 2009) and is thought to be a fundamental feature of nearly all living organisms (Stein and Meredith, 1993). This process of multisensory integration, particularly in superior colliculus (SC), is closely tied to the behavioral manifestations of multisensory integration – orientations of the eyes and head towards a stimulus are faster and more accurate (Alais and Burr, 2004; Angelaki et al., 2009; Battaglia et al., 2003; van Beers et al., 1999; Ernst and Banks, 2002; Gepshtein and Banks, 2003; Hollensteiner et al., 2015; Stein et al., 1989, 1996). This benefit depends on multisensory SC activity, as has demonstrated by lesions that selectively destroy multisensory SC neurons (Burnett et al., 2004, 2007).

Through this research, several principles have emerged which seem to hold across species and level of study. These principles were originally established in cat and hamster single unit superior colliculus (SC) recordings (Meredith and Stein, 1983, 1986; Meredith et al., 1987), then extended to monkey (Wallace and Stein, 1996; Wallace et al., 1996) and barn owl (Zahar et al., 2009) and have since been found to apply more generally even at the level of human behavior (Frens et al., 1995; Harrington and Peck, 1998; Hughes et al., 1998; Stein et al., 1996; Stevenson et al., 2012, 2014). The first principle, known as the spatial principle, is that cross-modal stimuli originating from areas of space near each other are likely to lead to
multisensory enhancement, while spatially disparate components often lead to multisensory depression (Meredith and Stein, 1983). The second principle, the temporal principle, which describes that cross-modal stimulus components which are temporally concordant are likely to be integrated while temporally discordant stimuli will not lead to integration (Meredith et al., 1987). The third principle, inverse effectiveness, notes that weakly effective cross-modal stimulus components are likely to be robustly integrated while robust stimuli are typically weakly integrated (Meredith and Stein, 1986).

Chapter 1 adds to this list a new principle, that the relative efficacy of the two cues being synthesized (i.e., their "balance") is a powerful predictor of the magnitude of the multisensory response: cues which are "balanced" in efficacy produce the largest multisensory product. Though this principle is derived from evaluating a single neurons responses and whether they are balanced, this is consistent with findings at the human behavioral level that the multisensory benefit (measured in speeded reaction times) is greatest when the performance in response to the modality-specific components are balanced in efficacy (Otto et al., 2013). Although the single-neuron approach and behavioral approach necessarily differ in that the

It may seem obvious that cross-modal cues balanced in efficacy would yield the strongest multisensory enhancements because if the cues were balanced in efficacy, and one was made weaker, there would be less overall drive provided to the neuron (and so a smaller enhancement would be expected). To correct for this,
multisensory response magnitudes were compared with the sum of unisensory response instead of being compared with the best unisensory response. This analysis showed that balanced samples still produced significantly greater multisensory enhancement than imbalanced samples even after correcting for this. This suggests that the amount of cumulative drive to the neuron cannot account for the balance effect, and that it may be an effect fundamental to the nature of multisensory integration.

That balanced stimuli should be preferentially integrated may be considered intuitive under the assumption that there is a correlation between the strength of modality-specific components in natural experience. For example, a predator nearby may present as a robust visual and auditory (and even somatosensory) stimulus while that same predator at a greater distance will tend to present as a less robust visual and auditory stimulus. Both sound and light intensity tend to fall as the square of the distance from the source (i.e., inverse-square law), so these two quantities will tend to be correlated for a given cross-modal stimulus. Similar to how the spatial and temporal rules help the brain decide which modality-specific components go together, the balance rule could serve as an additional indicator of which cues go together and which do not.

In addition to predicting the level of multisensory enhancement, the relative imbalance between the unisensory responses predicts the way temporal cue disparities will impact the magnitude of multisensory enhancement. With balanced
responses, the multisensory neuron will typically respond to temporally-disparate cues in one direction (e.g., auditory before visual) similarly to the way it responds to temporally-disparate cues in the opposite direction (e.g., visual before auditory).

When an imbalance between the unisensory responses exists, however, an asymmetry is seen in the way the neuron responds to temporally-disparate cues. If, for example, the visual response is stronger, enhancement is more robust when the visual response begins first. The general rule is “stronger first” to produce the strongest multisensory enhancement. The potential ecological source of such a state of imbalance and the behavioral relevance of this rule is further discussed below.

It’s important to note here that the general interaction between stimulus-onset asynchrony and the magnitude of enhancement has received some attention at the level of single-neuron electrophysiology (Meredith et al., 1987), population (Bushara et al., 2001), and behavior (Colonius and Diederich, 2004; Fujisaki et al., 2004). This work established the way in which the brain responds to temporally-discordant cues, finding that neurons demonstrate the greatest enhancement when the temporal disparity between cues is low, and the level of enhancement decreases as one or the other stimuli is advanced in time relative to the other (i.e., the “temporal tuning curve” peaks near simultaneity and falls off on both sides more or less symmetrically). Though some of the published data exhibited asymmetries in the way temporal disparity was handled by the brain (e.g., advancing a visual stimulus in time relative to an auditory results in a larger reduction in enhancement compared
to advancing the auditory relative to the visual) (Meredith et al., 1987), no explanation for this observation was provided. Here it was shown that the degree and direction of asymmetry is both (a) predictable based on the imbalance of the component responses and (b) malleable through adjustment of the stimulus (and thus response) efficacies. The fact that the temporal “tuning” of a multisensory neuron is not fixed but is dependent on the relative response imbalance is particularly important for application at the level of behavior because it suggests that the same findings should apply there. Because changing the stimuli changes the tuning of each component (i.e., multisensory neuron) of the population, and because these neurons mediate the relevant behavior (Burnett et al., 2004, 2007), we can predict that changing cue efficacy will change the shape of behaviorally-observed temporal tuning curves in future psychophysical studies.

This “stronger-first” rule is potentially useful for an animal due to two separate mechanisms. First, stronger responses tend to be shorter latency as can be seen in Chapter 3, Figure 4. This means that, all else equal, “stronger first” is the natural tendency of the system. If you start with two balanced responses which begin at the same time, increasing the intensity of one of the stimuli will tend to cause that response to be stronger and earlier and thus a stronger first arrangement. Second, it’s been found that repeatedly providing a certain cross-modal stimulus with a temporal disparity tends to lead to the earlier response increasing in strength (Yu et al., 2009). This, again, would lead to a stronger-first arrangement, simply as a consequence of the animal experiencing the statistics of its environment. Thus, as a
consequence of how the rest of the nervous system operates, multisensory neurons will tend to receive the stronger input first and will therefore get maximum effectiveness from the integration of those cues.

In Chapter 2, additional empirical observations are made about the nature of multisensory integration. As suggested in Chapter 1, the timing of unisensory component responses seemed to be an important determinant of multisensory responsiveness. Timing was given special focus in Chapter 2. Instead of evaluating responses in the typical fashion (i.e., number of evoked impulses per trial or average firing rate within a window), responses were evaluated in a continuous time fashion as spike density functions, a measure of the instantaneous firing rate of a neuron. From this analysis, there are three primary empirical findings: (1) the multisensory response is well-correlated, moment-by-moment, with the sum of the unisensory responses, (2) multisensory integration begins immediately at the earliest time in which both inputs are available to the multisensory neuron, and (3) multisensory integration is a highly dynamic process whereby the level of integration is consistently large at the beginning of the multisensory response period and much reduced thereafter.

In order to assess the completeness of these empirical findings in describing multisensory integration, they need to be placed into the formalized descriptive framework of a model. The multisensory models that have been offered in the past have been comparatively abstract. The most similar of these models operate at the
level of neurons, but produce responses that are in units of impulses per trial and are only constrained by population level empirical findings. One example involves a two-stage neural network model of the SC and a combination of primary and modulatory inputs following Hebbian learning rules (Anastasio and Patton, 2003). This model replicates broad aspects of multisensory responsiveness such as inverse effectiveness. In the same vein, Rowland et al. (2007b) provides a speculative model that reproduces major findings of SC multisensory integration, but does so in an abstract manner. Similarly, a model by Cuppini et al. (2010) focuses on a particular anatomical framework incorporating what is known or suspected about the projections from AES to SC and embeds that within a multi-layer neural network to again mimic broad aspects of multisensory integration in the SC. A fourth example uses a multi-layer neural network exhibiting divisive normalization to predict both neural activity (in monkey dorsal medial superior temporal cortex) and behavior (Ohshiro et al., 2011). Again, this model successfully recreates the general findings of multisensory research (i.e., inverse effectiveness and spatial principle) but has no representation of time and treats multisensory integration as a phenomenon which occurs trial-by-trial rather than moment-by-moment.

None of these models attempt to predict neural activity in a continuous time fashion, nor do they work at the level of a single neuron, nor do they typically test themselves against actual single-unit data. Instead, they use a high-level approach to determine whether multisensory integration tends to follow the prescriptions of an abstract framework. However, due to the loose constraints which burden such an abstract
framework, it is not difficult to mold them to fit the generally observed findings. Here we sought to take a different, descriptive approach and set a more ambitious goal: to predict the response of an actual, individual neuron in continuous time.

The empirical findings mentioned above were implemented in a formalized descriptive framework, referred to as the continuous-time multisensory (CTM) model, which represents a multisensory neuron receiving continuous visual and auditory unisensory input and, through a spiking neuron model, producing a multisensory response. The spiking neuron model (leaky integrate-and-fire model) requires input current to produce a spiking response, and those input currents were inferred using the observed spiking outputs in response to unisensory stimuli. Simply summing these inferred unisensory input currents and passing that summed current through the spiking neuron model produces a multisensory prediction that is quite similar to the actual multisensory response in the earlier portion, but consistently overshoots the multisensory response in later portions. This could be expected based on the empirical finding (#3 above) that multisensory integration exhibits strong enhancement at the beginning and then shifts to much weaker enhancement later and thus does not appear to result from a single, constant transform. This was implemented by adding an inhibitory input which is shaped according to the unisensory response shapes. This proved highly effective in enabling prediction of the entire multisensory spike density function using only the unisensory responses to form that prediction.
Because the CTM model operates at the level of a single spiking neuron and describes the empirically defined features consistently seen in multisensory neurons, it can naturally account for a wide variety of other observations such as the temporal principle, inverse effectiveness, balance, and the interaction between timing and balance discussed in Chapter 1. It also doesn’t just accurately predict the behavior of canonical integrating multisensory neurons, but also idiosyncratic responses that are particularly weak or strong or slow or fast or unusually shaped. Because we're considering the responses on a moment-by-moment basis, and because we're modeling the fundamental nature of multisensory integration as it happens in a single neuron, idiosyncratic neurons are as well-fit as canonical neurons. All neurons appear to be performing the same operation and are therefore equally predictable.

Further, because the neuron itself was modeled in a biologically-realistic way and was shown to accurately predict neurotypic integration, it can also be used to diagnose abnormalities. It is possible that certain manipulations (e.g., psychoactive pharmaceuticals) could dramatically impact multisensory integration in a way that fundamentally changes how it operates. In such a circumstance, the model would be expected to no longer be able to match the neurophysiology and could thus be useful in recognizing normal vs. abnormal operation.

A less dramatic alternative is that, in response to a manipulation, the model is still able to predict the integrated response of a neuron while also recognizing that
something has changed about that integrative operation. This was the outcome that was found when examining non-integrating neurons. The majority of neurons in the deep layers of the SC will typically demonstrate multisensory integration, assuming the principles mentioned above (spatial, temporal, inverse effectiveness) have been followed. However, some neurons (33% in the sample studied in Chapter 2) still don’t integrate their inputs. In past studies of integration, these non-integrating neurons were usually ignored (however, see Jiang et al., 2001) under the assumption that they were part of some other circuit and perhaps engaged in some other operation. Here we were able to show that there isn’t anything particularly unusual about these non-integrating neurons. Not only was the model still able to accurately model their activity, but it was able to tell us why they were different: they have a stronger inhibitory influence than do integrating neurons. For normally-integrating neurons, the inhibitory component of the model drove the multisensory response down from super-additive at the beginning of the multisensory response to additive in later portions. For non-integrating neurons, the multisensory response still began super-additive (this is a fundamental feature of the model and of actual neurons) but was typically driven sub-additive in later portions.

So, a sizeable minority of SC neurons which were previously thought to be doing some other task may now be thought of as fundamentally no different from the integrating neurons. Of course, it’s still possible that the non-integrating population is indeed part of a different circuit and this is why they demonstrate stronger inhibition,
but it no longer should be the default assumption that they are separate and distinct from integrating neurons. From the evidence provided here, they can be assumed to be not categorically different.

The most speculative component of this model is the additional inhibitory influence which is supposed to exist in multisensory conditions. This does not suggest that there is no inhibition in unisensory conditions, rather in the model this inhibitory input in unisensory conditions is already combined with the excitatory drive which we observe through recording the neural activity. It was found that combining the presumptive aggregated unisensory inputs consistently over-predicted the activity in the late portion of the multisensory response. An additional inhibitory effect was needed to account for the results. But it also makes sense that the SC would benefit from a mechanism that helps to “push the brakes” shortly after a multisensory input that drives the neuron to robustly respond, thereby helping to make sensory activity punctate and of higher temporal fidelity than would happen if multisensory responses were allowed to proceed for longer periods.

There are multiple ways in which this modeled inhibitory input could correspond to biological reality. One approach would be to suppose a global inhibitory influence which could represent an input from substantia nigra pars reticulata (SNr) (Hikosaka, 1989; Liu and Basso, 2008). While SNr clearly provides inhibitory input to the SC, it does so for unisensory stimulation and so would already be included in the measurement of unisensory responses. More importantly, the analyses provided
here indicate that the timing of the apparent inhibitory influence is closely related to the timing of the activity of the SC neuron under study, while a global inhibitory influence would be expected to be consistent across the population regardless of the particular timing of excitatory input of the neuron under study. In fact, such a modeling approach was attempted (data not shown) and it was entirely unsuccessful for the reasons stated above: the inhibition appears related to the specific shape and timing of the neuron under study and so the inhibitory influence needs to vary accordingly. Nevertheless, SNr inhibitory input should be explored in future studies to determine the ways in which it may specifically influence SC multisensory integration.

Instead of modeling global inhibition, the approach taken here was to use a given neuron’s unisensory responses to define the shape and timing of the inhibitory influence. This could be anatomically implemented in a number of ways. One possibility is that unisensory projections to the SC synapse both on the principal neuron as well as a nearby unisensory interneuron which subsequently projects to the principal neuron under study. If these inhibitory inputs undergo dendritic integration, thereby producing a multisensory enhancement-like effect of the inhibitory input within a dendritic compartment of the principal neuron, the effect would be consistent with what we see empirically and with how the model functions. Another possible architecture would include inhibitory interneurons which themselves integrate unisensory inputs and subsequently provide an “enhanced” suppression to the principal neuron under multisensory conditions. Anatomical work
supports this general architecture (AES projecting onto inhibitory interneurons and principal neurons) but isn’t exhaustive enough to differentiate between the two possibilities outlined above (Fuentes-Santamaria et al., 2008). A third possibility is that the principal neuron is self-inhibitory by projecting onto an inhibitory interneuron which feeds back onto itself. All of these possibilities rely on local inhibitory interneurons and could thus be tested by blocking (pharmacologically or optogenetically) their effect. The CTM model provides testable predictions for the impact of such a manipulation on a neuron-by-neuron basis. For example, optogenetically silencing GABA-ergic neurons near the principal neuron under study should lead to a multisensory response predicted by the model with the additional inhibitory factor removed.

Given the simplicity of this model (particularly compared with extant models of multisensory integration), it may seem that multisensory integration is an inevitable outcome of multisensory convergence. This, however, is not the case. There are a number of conditions under which multisensory convergence doesn’t lead to multisensory integration but the most relevant here is during development. For neonatal animals (including humans), multisensory convergence can develop quite early (Stein et al., 1973) but experience with that animal’s environment is required before multisensory integration can be “learned”. This experience is normally provided in the first few months in cat (Wallace and Stein, 1997; Wallace et al., 2006) and that development can be obviated by preventing or disrupting cross-modal experience through raising an animal in darkness (Yu et al., 2010) or in a
high-intensity sound environment (Xu et al., 2014). Multisensory development can also be delayed by long-term deactivation of association cortex (AES; anterior ectosylvian sulcus) during the 4-12 week postnatal age at which integration normally develops in SC (Rowland et al., 2014). Providing animals with only controlled cross-modal stimuli during development leads to SC integration that reflects the statistics of that controlled cross-modal stimuli (Wallace and Stein, 2007; Xu et al., 2012) and the same is true even when that experience is provided under anesthesia as adults (Xu et al., 2012, 2014, 2015; Yu et al., 2010). Even as an animal matures and demonstrates normal multisensory integration, it can be easily disrupted by reversibly deactivating AES (Alvarado et al., 2007, 2009; Jiang et al., 2001; Wallace and Stein, 1994, 2000). This demonstrates that the projection from AES to SC is instrumental in the development and expression of multisensory integration in SC.

This dependence of SC on AES contributes to an apparent paradox when considering the relative rates of development of these two structures. Axiomatically, midbrain develops before cortex (Yakovlev and Lecours, 1966) and this seemed to be true in this circuit as well. Multisensory integration capability in SC has been found to begin around 5 weeks post-natal and reach adult-like levels around 12 weeks (Wallace and Stein, 1997). However, the same capabilities develop in AES much later – not beginning until around 12 weeks and reaching adult-like levels after 20 weeks. Of particular relevance to this circuit are the unisensory neurons, because these are the neurons which are known to project to SC (Fuentes-Santamaria et al., 2009; Meredith and Clemo, 1989) and be critical for multisensory
integration. Surprisingly, these unisensory AES neurons which project to, and enable integration in, SC seemed to be absent or unresponsive until an age at which SC has reached adult-like levels of multisensory integration. Visual responsiveness was not found until 12 weeks of age in AES (Wallace et al., 2006). So, paradoxically, it seems that adolescent SC must rely on the input from unisensory AES neurons which cannot be provided because these neurons don't yet respond to sensory input.

One possible explanation for this paradox is that the data and conclusions from multiple previous studies cannot be directly combined in this case even though they bear on the same question. To obviate the potential confounds inherent in combining studies conducted by different investigators using different equipment and recording from different animals, SC and AES were studied in the same animals by the same investigator using the same equipment. By doing this, the paradox was resolved by finding that (a) unisensory neurons in AES are actually present at the earliest ages in which SC integration is found, (b) in many ways, responsivity in AES isn't actually delayed relative to that in SC, and (c) full maturity of AES neurons is evidently not required for integration in SC. The one factor that does seem to be required for multisensory integration is that the unisensory drive be reliable (i.e., consistent across multiple stimulus presentations).

Some contradictions between the findings here and previous work should also be discussed. The most important difference is that unisensory drive is present in AES
much earlier than previously thought. This may result from the search procedure used to identify cortical neurons. Wallace et al. (2006), from which the estimates were originally drawn, appear to have only recorded from neurons which demonstrated an overt response which could be manually detected (i.e., seen on an oscilloscope or heard from the audio-amplified neurophysiological signal). In this case, it would be quite easy to miss the evoked neural activity because at young ages it is particularly noisy and sensory responses are weak and unreliable (Chapter 3). The search strategy employed in the present study was to simply identify a neuron with sufficient isolation, and then test the neuron with a systematic visual and auditory receptive field mapping protocol rather than the more rapid, but less sensitive procedures used previously. This allows for a much greater likelihood of detecting sensory responses than before. A second reason may be the greater number of animals used here (31 vs. 16) which makes detection of rare neurons more likely. A third possible explanation is that the animals in the previous study at ages less than 12 weeks (i.e., the ages when visual responses were not found) were studied immediately after surgical preparation when cortex may have still been traumatized. Indeed, in the present study, a single animal was studied in a similar way but the data were not used because cortical activity was nearly impossible to find even after ~12 hours post-surgery. All animals used in this study were allowed to recover post-surgery back in the home pen for at least 12 hours before being anesthetized for the recording session and that seemed to obviate any difficulty locating cortical neurons.
A second difference between previous observations and the data presented here is that the development of multisensory integration capabilities is not as different between SC and AES as was previously found. It was previously found (Wallace et al., 2006) that cortex develops the ability to integrate primarily in the 12-20+ week age range and that SC develops the same capability primarily between 5 and 12 weeks of age. Here we find that both SC and AES are largely transitioning from the naïve state (i.e., demonstrating little to no multisensory integration capabilities) to the fully developed state (demonstrating adult-like multisensory integration capabilities) between 10 and 15 weeks of age. This new finding is based on a far larger number of neurons as well as utilization of within-animal controls afforded by studying both structures in the same animals, in addition to the same investigator, equipment, and analytical methods.

A finding consistent with extant literature is that AES (and SC) is indeed immature when SC (and AES) is developing integration capabilities. Though it seemed counter-intuitive that something thought to be a higher-order function like multisensory integration would be enabled by immature inputs, this now appears (a) to nevertheless be true and (b) be less counter-intuitive in light of the findings of Chapter 2 here, where the mechanistic interactions necessary for generating multisensory integration is revealed to be much simpler than previously thought.

These new findings elucidate the process of how multisensory integration develops in the SC by obviating the need to explain how SC integration can depend on AES
neurons which don’t exist because those AES neurons do, in fact, exist. This will assist the field in moving forward with greater confidence as to how multisensory integration develops. The observation that SC and AES multisensory integration capabilities develop in tandem (despite the earlier appearance of multisensory convergence in SC) shows that two axioms of the brain are supported: (1) midbrain develops before cortex, and (2) cortex demonstrates greater plasticity than subcortex. In this case, the circuitry is seemingly intact earlier in midbrain, but the cortical neurons more quickly learn to synergistically combine the cues than does midbrain.

It is interesting to consider this work in light of a Bayesian framework of plasticity and development which ties together all the findings in this series of studies. Such a framework has been proposed to explain how multisensory integration capabilities can develop in the SC (Cuppini et al., 2012; Deneve and Pouget, 2004; Rowland et al., 2007a). These models differ in their details, but generally describe how the synaptic weights of (particularly AES) projections to SC are modified through experience and ultimately allow for multisensory integration. For our purposes, this might look like synaptic weights increasing during early development to account for the increase in reliability and robustness of SC activation with development (Chapter 3). These weights are continuously adjusted in accordance with the statistics of the animal’s environment such as the relative strength and timing of modality-specific components typically encountered (Chapter 1). Once these weights are at least
somewhat established, multisensory integration occurs according to some very simple rules (Chapter 2).

Taken together, these findings present a new way of thinking about both the development and mechanistic basis of multisensory integration. Besides an intact circuit, a new prerequisite of multisensory integration capability is that unisensory inputs are reliable. Once these prerequisites are achieved, the multisensory response can be wholly understood simply by considering the timing and magnitude of the component unisensory responses and modeling their convergence via a simple computational framework. The simplicity of this approach will provide an elegant method for the future assessment of multisensory integration across animal model, brain structure, and mental state.
References


somatosensory multisensory interactions in low-level sensory cortices are not

information in bimodal neurons in the rattlesnake optic tectum. Science 213,
789–791.


responses to a single moving object. J. Exp. Biol. 211, 2919–2930.

you see what I am saying? Exploring visual enhancement of speech
comprehension in noisy environments. Cereb. Cortex N. Y. N 1991 17,
1147–1153.

multisensory spatial localization with the physiological properties of the superior
colliculus. Exp. Brain Res. Exp. Hirnforsch. Expérimentation Cérébrale 180,
153–161.

mechanisms underlying multisensory integration in the superior colliculus.
Perception 36, 1431–1443.

Neurosci. 34, 7198–7202.


Stein, B.E., and Meredith, M.A. (1993). The merging of the senses (Cambridge,
Mass: MIT Press).


Ryan Lloyd Miller
ryamille@wakehealth.edu

**Education**

2008-Present  PhD Program, Neurobiology and Anatomy, Wake Forest School of Medicine
Advisors: Barry E. Stein and Benjamin A. Rowland

2005-2008  Bachelor of Science, Psychology, North Dakota State University

2003-2005  Associate of Science, Computer Science, Bismarck State College

**Academic and Professional Honors**

2008-Present  Member, Society for Neuroscience

2011-2013  T32 NS 073553-01 Training Grant Title: Training Program in Multisensory Processes

**Awards**

2016  Norman M. Sulkin Scholarship Award in Neuroanatomy

2013  Fine Science Tools travel award

**Research Experience**

2009-Present  Dissertation work in cat electrophysiology laboratory of Dr. Barry E. Stein: Temporal and developmental aspects of multisensory binding

2008-2009  Graduate work in monkey electrophysiology laboratory of Dr. Ramnarayan Ramachandran: recording auditory responses of awake monkey cochlear nucleus and inferior colliculus.

2006-2008  Undergraduate work in human EEG laboratory of Dr. Wolfgang A. Teder-Sälejärvi: Effects of attention on multisensory integration

2006-2008  Undergraduate work in human psychophysics lab of Dr. Mark J. Brady: Identifying image edge sources for the purpose of computer vision

2006-2007  Undergraduate work in human psychophysics lab of Dr. Linda K. Langley: Age-related changes in visual attention
Publications and Presentations


Miller RL, Stein BE, and Rowland BA. Multisensory integration uses a real-time unisensory-multisensory transform. Submitted.


