Multisensory integration produces an initial response enhancement

Benjamin A. Rowland and Barry E. Stein

INTRODUCTION

The brain has evolved the ability to integrate information across the senses in order to improve the detection and disambiguation of biologically significant events. This multisensory synthesis of information leads to faster (and more accurate) behavioral responses, yet the underlying neural mechanisms by which these responses are speeded are as yet unclear. The aim of these experiments was to evaluate the temporal properties of multisensory enhancement in the physiological responses of neurons in the superior colliculus (SC). Of specific interest was the temporal evolution of their responses to individual modality-specific stimuli as well as to cross-modal combinations of these stimuli. The results demonstrate that cross-modal stimuli typically elicit faster, more robust, and more reliable physiological responses than do their modality-specific component stimuli. Response measures sensitive to the time domain showed that these multisensory responses were enhanced from their very onset, and that the acceleration of the enhancement was greatest within the first 40ms (or 50% of the response). The latter half of the multisensory response was typically only as robust and informative as predicted by a linear combination of the unisensory component responses. These results may reveal some of the key physiological changes underlying many of the SC-mediated behavioral benefits of multisensory integration.

Keywords: multisensory, superior colliculus, physiology, cross-modal, information, latency

MATERIALS AND METHODS

Surgical preparation

Two adult cats were implanted with stainless-steel recording chambers using aseptic surgical techniques in accordance with the Guide for the Care and Use of Laboratory Animals (National Institutes of Health publication 86–23) and an approved Institutional Animal Care and Use

© 2007 Rowland and Stein. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.
Committee protocol. Each animal was anesthetized with ketamine hydrochloride (20 mg/kg, i.m.) and maintained with halothane (1.5–4%). The recording chamber was attached to the skull over a craniotomy giving access to the SC via the overlying cortex. Post-surgical analgesics (butorphanol tartrate; 0.1–0.4 mg/kg for 6 hours) were administered as needed and antibiotics (cephazolin sodium; 25 mg/kg) were administered b.i.d. for 7 days.

Recording sessions began 7–10 days after surgery. No wounds or pressure points were induced during recording. The animal was prepared for recording with a mixture of ketamine hydrochloride (20 mg/kg, i.m.) and acepromazine maleate (0.4 mg/kg, i.m.), intubated, and paralyzed with pancuronium bromide (0.3 mg/kg), and maintained with an infusion of ketamine (10–15 mg·kg⁻¹·h⁻¹, i.v.), pancuronium (0.1–0.2 mg·kg⁻¹·h⁻¹, i.v.), and 5% dextrose Ringer’s (1 mg·kg⁻¹·h⁻¹, i.v.). Respiratory rate and volume were controlled so that end-tidal CO₂ was ~4.0%. At the end of each experiment anesthetics and paralytics were terminated, and the animal was returned to its home cage after it recovered normal respiration and was ambulatory.

Only SC neurons responsive to both visual and auditory stimuli were studied and some of the characteristics of this dataset were previously described (see Rowland et al., 2007). Visual stimuli were stimuli of light-emitting diodes (LEDs), auditory stimuli were bursts of bandpass-filtered noise delivered by small speakers close to each LED. Stimuli were controlled using a Spike II (Cambridge Electronics Design, Cambridge, UK) analog-to-digital converter connected to a PC. Every isolated neuron was tested with brief (typically 50–100 ms) modality-specific (visual, auditory) and cross-modal (visual-auditory) stimuli at a given location. To ensure a broad sample of responses, each modality-specific stimulus was adjusted to three different intensities (selected online) to span its dynamic range. An analog-to-digital converter connected to a PC. Every isolated neuron was analyzed using a time-insensitive and time-sensitive response measures. Response measures (i.e., latency) and offsets were determined by the three-step geometric method (Figure 1) which performs very well over a large range of response parameters (Rowland et al., 2007). First, the mean cumulative impulse count over time \( C(T) \) was calculated. Second, the function \( C(T)/T \) was calculated, its maximum \( E \) after the stimulus onset \( S \) identified, and a “null hypothesis” line \( H_0 \) drawn between \( C(S−(E−S)) \) and \( E \). In the absence of a response, \( C(T) \) would follow \( H_0 \). The response onset \( R \) was identified as the time that \( C(T) \) maximally diverged from \( H_0 \), measured by the length of a plumb line perpendicular to \( H_0 \) that intersected \( C(T) \). The response offset was identified in a similar manner using a null hypothesis line drawn from \( C(R) \) to \( C(E+(E−R)) \). Significant responses were identified using a t-test between impulse counts in the putative response and spontaneous windows, using a corrected alpha value to maintain a 5% error rate (Rowland et al., 2007).

Response magnitude was calculated by averaging the number of impulses between \( R \) and \( E \) minus the expected number of spontaneous impulses adjusted for size of the response window. Maximum mutual information (MMI) was calculated using a naïve (bias uncorrected) estimator of mutual information (Nelken et al., 2005). Mutual information is a measure of the reliability of the sensory response, but not necessarily its magnitude (though these frequently covary). Briefly, the impulse raster was transformed to a spike-density function (SDF) through convolution with a square-wave (i.e., “boxcar”) kernel with width \( W \). The resulting SDF was sampled every 1 ms on every trial to populate a 2-D joint probability matrix relating observed SDF values (e.g., 2 impulses/bin) to whether they occurred inside the response window or outside of it. Next, the marginal probability distributions were calculated by summing across the columns and summing across the rows of this matrix. The marginal distributions were then multiplied (cross-product) to form the probability matrix expected from random association (not shown in Figure 1). The original joint probability matrix was then divided (dot quotient) by the random association matrix. The logarithm (base 2) of this matrix was then multiplied (dot product) by the original joint probability matrix. The sum of the resulting matrix (across columns and rows) determined the mutual information extractable from the SDF regarding the response window (between 0 and 1 bit) using the kernel with width \( W \). Multiple values of \( W \) were tested for each response (2–200 ms); MMI was the largest value produced by any value of \( W \).

The cumulative stimulus-evoked impulse count \( qsum \) was the mean number of stimulus-evoked impulses observed on or before each moment in time. It was calculated by correcting \( C(T) \) for spontaneous activity (subtracting time multiplied by the spontaneous rate). The \( qsum \) approximates a straight line when the firing rate is constant. Its slope corresponds to the elevation of the firing rate over spontaneous activity. Event estimates were used as a measure of the instantaneous “report” obtainable from the neuron’s response at each moment in time derived using an “optimal” decoder. This measure was derived from the SDF computed by convolving the impulse raster with the square-wave function producing the MMI. Each observed value in the SDF was replaced with the probability that that value was observed in the response window (versus being observed in the spontaneous window), calculated by transforming the joint probability matrix described above to a conditional probability matrix by dividing by the marginal distribution of SDF values. The result was averaged across trials, and corrected for spontaneous activity levels by subtracting the average estimate prior to the stimulus onset.

Data analysis
The data were analyzed in several steps. First, the unisensory responses (latencies, magnitudes, MMI, peak event estimates) were quantified and cross-correlated. Next, multisensory responses were compared to the best (i.e., fastest or largest) of the associated unisensory responses using each of these measures. The magnitude of the multisensory enhancement (MSI) for each measure was calculated as a percent change: \( MSI = 100 \times (M − U)/U \), where \( M \) represents the multisensory response value and \( U \) represents the best unisensory value (Meredith and Stein, 1983). Cross-correlations between MSI values for the different measures were computed. These analyses thus quantified the incidence, magnitude, and relationships between multisensory enhancements measured over the entire response window (i.e., using time-insensitive measures).
The next part of the analysis quantified the temporal profiles of the multisensory enhancements over the population of responses using the qsum and event estimate measures. The focus here was on the average trend in the amount of enhancement over time; that is, whether response enhancements increased, stayed constant, or decreased throughout the response. All data samples were synchronized to the onset of the multisensory response. Enhancement was quantified as the difference between the multisensory response and its associated best unisensory response, which was determined by selecting the greatest difference between the multisensory response and its associated best unisensory component response) to visualize the general trends. The derivative (slope) of the average qsum at every moment in time indicates the amount of enhancement at that moment in time, while the derivative of the estimate indicates the rate at which the amount of enhancement is changing. These derivatives were estimated from 10th-order polynomials fit to the average qsum and estimate functions. The derivative functions were expressed as a percentage of their maximum value. To determine how representative the resulting functions were of the average trend in the amount of enhancement over time; that is, whether the derivatives calculated in this manner did not show the same trend as derivatives computed on the average qsum and estimate functions, then the averaged functions were unlikely to be representative of individual data samples (this did not happen here).

The above analysis profiled the existence and rate of enhancement at each moment in time after the multisensory response onset. These procedures were repeated after all data samples were normalized by the multisensory response duration, that is, all responses were “stretched” or “shrunk” in time to fit in the same window of 100 time units. This second analysis quantified the existence and rate of enhancement relative to the response duration.

Differences in the time-sensitive response measures were also assessed using a “time-to-criterion” assessment, which identified if and when each multisensory and unisensory response crossed particular threshold values (1–20 impulses for cumulative impulse count, 0–1 for event estimates). Quantification involved calculating the incidence of multisensory responses crossing thresholds earlier (or later) than either of their component unisensory response, and the timing difference between the crossings (assuming at least one unisensory response crossed the threshold).

Finally, the magnitude of the multisensory event estimate at each moment in time was compared to the predicted sum of the unisensory event estimates, computing by adding the two and subtracting their product. The multisensory response exceeds this unisensory sum when it
consistent with other observations (Rowland et al., 2007) that responses magnitude of the multisensory enhancements measured in this way were event estimate are measures of multisensory integration that are not time-
Changes in response latency, total impulse count, MMI, and the peak event are shorter or longer, but when unisensory latencies were comparatively long, multisensory responses were much more likely to be shorter. However, the magnitude of the latency shift did not appear to monotonically increase with the unisensory latency. Latencies were on average 23% shorter (median = 47%), MMI was usually 64% larger (median = 27%), and peak event estimates were on average 56% (median = 24%) larger in the multisensory versus the best unisensory component response. The distributions of enhancements in each response measure were biased to smaller values, but all exhibited long “tails” extending far in the enhanced direction which corresponded to the incidence of data samples with weak or modest multisensory responses.

As expected, there were significant correlations between the amounts of multisensory enhancement calculated using these different measures (0.47–0.6), albeit (as expected) weaker than the unisensory correlations (see Figure 2). In the majority of cases (61%) the responses had a positive MSI for all three measures, while 11% had a negative MSI in all three.

The temporal profile of multisensory enhancement
The continuous qsum and event estimate measures provide methods to assess the timing of multisensory enhancements; specifically, how they change over time. Because the qsum accumulates over time differences between qsums permit greater sensitivity to subtle or slowly evolving differences, while comparisons between event estimates allow direct assessments of the differences in the responses at each moment in time. The multisensory and best unisensory qsums and event estimates at each moment in time were averaged across the data samples after synchronizing them by the multisensory response onset (Figure 4, top). Because the data were not normalized by the robustness of the unisensory response, the average enhancements appeared to be modest, although in reality they represented large proportional enhancements (Rowland et al., 2007). Normalizing the data by the unisensory response magnitude did not affect the trends.

The average difference between the multisensory and best unisensory qsums (Δqsum) revealed a function that abruptly increased at the response onset and afterwards approached a plateau. This was also reflected in the decreasing slope (Figure 4, middle left) of the mean Δqsum which was halved within the first 40 ms of the response. This indicated that after 40 ms the average difference between the multisensory and best unisensory qsums was still growing but not as quickly as earlier (the magnitude of the enhancement was not as great). The slope reached approximately zero within 200 ms. These qsum observations are paralleled by the trends evident in the average event estimate difference (Δestimate) between the multisensory and best

Figure 2. Unisensory response properties of multisensory SC neurons. Shown are the probability distributions (PDFs) and cumulative (probability) distribution functions (CDFs) of the response latency, magnitude, MMI, and peak event estimates for unisensory visual (top, N = 55) and auditory (bottom, N = 90) responses. Dotted lines identify the mean value for each parameter. Shown on the right are the cross-correlations between different parameters (L, latency; I, impulses (magnitude); MMI, maximum mutual information; P, peak event estimate), indicating strong positive correlations between all of the measures except response latency, which showed a weak, but consistent negative correlation (i.e., shorter latency) with the other measures.
unisensory responses (Figure 4, middle right). The \( \hat{D} \) estimate function indicates the amount of enhancement present at each moment in time. Its slope indicates the rate of enhancement. In parallel to the result obtained from the \( D \) qsum, the amount of enhancement indicated by the \( \hat{D} \) estimate peaked at around 40 ms after the response onset (slope of the line approached zero). There is a good deal of similarity between the slopes calculated after the responses are averaged and the individual data samples contained in the average.

When the data were normalized by the multisensory response duration, the same trends were apparent but were less dramatic (Figure 4, bottom). The slope of the \( \Delta qsum \) was halved within the first 50% of the multisensory response duration, approximately at the same time that the slope of the estimate approached zero. These results are consistent with the hypothesis that multisensory integration is characterized by an IRE (Rowland et al., 2007); that is, enhancements are present and the rate of response magnitude enhancement is greatest at the beginning of the response.

**Multisensory enhancement speeds responses**

Single neurons within a network are thought to act as logic gates, generating responses proportional to their inputs once they have exceeded a threshold value. Within this framework, it is reasonable to assume that neurons receiving input from the SC initiate their responses when inputs cross certain threshold criteria. A “time-to-criterion” analysis calculated the earliest time that each response crossed threshold criteria (Figure 5) and revealed that multisensory qsums and event estimates were roughly twice as likely to cross most threshold values than their unisensory component responses, and that there was a very low incidence (<5%) of multisensory responses that did not meet a criterion met by either of its unisensory responses (Figure 5, middle row, see asterisk). Not only were multisensory responses more likely to reach threshold criteria, they almost always reached them earlier: multisensory responses reached criteria earlier than either unisensory response in 76–95% of the samples in each criterion set (average difference $\approx 10$ ms, no consistent trend across criteria). Multisensory enhancements in this regard were slightly more dramatic when the response was measured by event estimate versus qsum. Thus, multisensory integration had a more substantial impact when the time domain was considered, even greater than is evident from measures averaged across the entire response duration (e.g., see Figure 3).
prediction for the multisensory response that is exceeded when the multisensory response reflects a non-linear, superadditive interaction. Thus, the “impact” of integration is greatest when the multisensory response exceeds this sum. To evaluate this phenomenon, event estimates were sampled with 1 ms resolution and normalized by the multisensory response duration (see above). Multisensory event estimates during the response were then compared to the sum of the unisensory estimates, minus their product, revealing a clear trend over time.

Figure 4. The temporal profile of multisensory enhancement. Multisensory and best (i.e., largest) unisensory responses are quantified as qsums (left) or event estimates (right), shifted to synchronize with the multisensory response onset, then averaged across data samples. Average multisensory qsums and event estimates reached higher values than the corresponding best unisensory responses (top). Subtracting the multisensory and unisensory responses and averaging across samples yielded mean difference functions (middle, black lines). The mean difference functions were fit with 10th-order polynomials, illustrated with dashed black lines but virtually invisible given the closeness of the fit. The derivatives of these functions are illustrated by dotted lines and scaled proportional to their maximum value (read off right axis). Darker lines indicate the slope of the function fit after averaging across samples. Lighter lines indicate the slope fit to the individual data samples and then averaged. Slopes show a general decreasing trend. Normalizing these functions for different values of response duration (bottom) revealed the same overall trend: enhancements were high at the beginning of the response, and decreased rapidly within 40 ms or 50% of the total response duration. However, enhancement continued throughout the response.

Figure 5. Multisensory responses reach higher levels earlier. Left: this analysis determined if and when responses quantified as qsums (top) or event estimates (bottom) reached threshold criteria levels. Middle and Right: multisensory responses crossed criteria levels more often (Middle) and earlier (Right) than did unisensory responses, and there were virtually no cases (see asterisk and arrow) where unisensory responses reached a criterion not also reached by the multisensory response. See text for more details.
% Multisensory Response Duration

1-25%  26-50%  51-75%  76-100%

Figure 6. Initial response enhancement. The multisensory event estimate was compared to the sum of the two unisensory estimates (minus their product) in different stages of the multisensory response (240 multisensory responses sampled with 1 ms resolution). In the first quarter, multisensory event estimates were usually (87% of the time) greater than the sum of the unisensory estimates. This elevation was maintained in the second quarter, where multisensory event estimates exceeded the unisensory sum 83% of the time. The final two quarters showed increasing parity between the multisensory and summed unisensory event estimates (3rd = 75%, 4th = 65%). The figure on the right provides the incidence of multisensory event estimates greater than the unisensory sum over the duration of the multisensory response (1–100% in 1% steps). Initially more than 85% of multisensory estimates exceeded the sum, but at the end of the response, less than 65% did so.

DISCUSSION

Prior research has established that multisensory enhancements are proportionally largest at the beginning of the response, creating what has been referred to as an initial response enhancement, or IRE (see Rowland et al., 2007). A major contributing component of that result was that multisensory responses typically exhibited shorter latencies; so that the initial responses were compared to non-existent or minimal (just beginning) unisensory responses. The present results showed that the IRE is not restricted to this circumstance. Indeed, the acceleration in the magnitude of the IRE continues even when the unisensory response is present. Overall, its acceleration is greatest during the first 40 ms (or 50% of the total multisensory response duration).

The implication of these results is that multisensory integration takes place immediately so that the streams of concordant information that are derived from cross-modal sources are being integrated as soon as they are available to the neuron. This functional model is consistent with shortening of response latencies to cross-modal stimuli (Figure 3), presumably due to the summation and integration of subthreshold inputs at the SC neuron where the computation is being performed (Rowland et al., 2007). This observation has yet to be reconciled with evidence that the SC obtains cross-modal information directly, but its multisensory response enhancement requires inputs that are also relayed through association cortex (Jiang et al., 2001; Jiang and Stein, 2003; Wallace et al., 1993). We believe that this likely reflects the long period during which subthreshold inputs from all sources are being integrated.

These issues speak of the importance of the temporal domain in the analysis of multisensory processes. The SC is believed to be engaged in the detection and localization of stimuli, and subsequently in the evocation of orienting responses (Sprague and Meikle, 1965). Ideally, responses should be made as soon as possible. Consequently, the amount of information present at the beginning of the response and the rate with which it is communicated are likely to be critically important in determining the success or failure of a behavioral product. This is likely achieved by shortening response latencies as shown here as well as decreasing the interval between the sensory and premotor responses (see also (Bell et al., 2006; Whitchurch and Takahashi, 2006)).

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGMENTS

We thank Nancy London for technical help and Stephan Questy and Terrence Stanford for their help with initial dataset. This research was supported by NIH grants NS36916 and EY016716.

REFERENCES

Bell, A. H., Meredith, M. A., Van Opstal, A. J., and Munoz, D. P. (2006). Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. Exp. Brain Res. 174, 53–59.

Binnis, K. E., and Salt, T. E. (1996). Importance of NMDA receptors for multimodal integration in the deep layers of the cat superior colliculus. J. Neurophysiol. 75, 920–930.

Frens, M. A., Van Opstal, A. J., and Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Percept. Psychophys. 57, 802–816.

Geelen, S. C., Schmidt, R. A., and van den Heuvel, P. J. (1983). On the nature of intersensory facilitation of reaction time. Percept. Psychophys. 34, 161–168.

Goldring, J. E., Dorris, M. C., Cornell, B. D., Ballantyne, P. A., and Munoz, D. P. (1996). Combined eye-head gaze shifts to visual and auditory targets in humans. Exp. Brain Res. 111, 68–78.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., and Hendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. J. Exp. Psychol. Hum. Percept. Perform. 20, 131–153.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., and Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. J. Neurophysiol. 85, 506–522.

Jiang, W., Jiang, H., and Stein, B. E. (2002). Two corticocortical areas facilitate multisensory orientation behavior. J. Cogn. Neurosci. 14, 1240–1255.

Jiang, W., and Stein, B. E. (2003). Cortex controls multisensory depression in superior colliculus. J. Neurophysiol. 90, 2123–2135.

Meredith, M. A., and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. Science 221, 389–391.
Nelken, I., Chechik, G., Mrsic-Flogel, T. D., King, A. J., and Schnupp, J. W. (2005). Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. *J. Comput. Neurosci.* 19, 199–221.

Perrault, T. J., Jr., Vaughan, J. W., Stein, B. E., and Wallace, M. T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J. Neurophysiol.* 93, 2375–2396.

Perrott, D. R., Saberi, K., Brown, K., and Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Percept. Psychophys.* 48, 214–226.

Pouget, A., Denève, S., and Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nat. Rev. Neurosci.* 3, 741–747.

Rowland, B. A., Quesy, S., Stanford, T. R., and Stein, B. E. (2007). Multisensory integration shortens physiological response latencies. *J. Neurosci.* 27, 5879–5884.

Sprague, J. M., and Meikle, T. H. (1965). The role of the superior colliculus in visually guided behavior. *Exp. Neurol.* 11, 115–146.

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

Stein, B. E., Wallace, M. W., Stanford, T. R., and Jiang, W. (2002). Cortex governs multisensory integration in the midbrain. *Neuroscientist* 8, 306–314.

Wallace, M. T., Meredith, M. A., and Stein, B. E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J. Neurophysiol.* 69, 1767–1809.

Whitchurch, E. A., and Takahashi, T. T. (2006). Combined auditory and visual stimuli facilitate head saccades in the barn owl (Tyto alba). *J. Neurophysiol.* 96, 730–745.

Wilkinson, L. K., Meredith, M. A., and Stein, B. E. (1996). The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp. Brain Res.* 112, 1–10.