DEVELOPMENT OF MULTISENSORY INTEGRATION: TRANSFORMING SENSORY INPUT INTO MOTOR OUTPUT

Barry E. Stein,* Mark T. Wallace, and Terrence R. Stanford
Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, North Carolina

By pooling and integrating signals from different sensory channels, specialized populations of "multisensory" neurons not only help to maximize the brain's ability to detect and identify external events, but also help to initiate reactions to them. Although multisensory neurons are found in many areas of the brain, those in the midbrain (i.e., superior colliculus, SC) have been studied most extensively and have served as a model for understanding some of the neural operating principles of multisensory integration, as well as the impact of these processes on overt attentive and orientation behaviors. However, this capability is not hard-wired at birth. Very young SC neurons are responsive only to unimodal inputs; it is not until many days later that some of them begin to respond to inputs from more than a single sensory modality, and even then they are not yet capable of integrating these inputs to produce the synthesized multisensory signals that characterize the normal adult. The most significant occurrence to precede this maturational change is the appearance of influences from association regions of the neocortex. These influences appear abruptly on any given individual neuron, but because different neurons are targeted at different times during development, it takes many weeks before the mature complement of such neurons is achieved. It is likely that the maturational timing of the interplay between the cortex and SC determines not only the kinds of multisensory information that can be integrated in SC neurons, but the kinds of multisensory behaviors that SC neurons are able to mediate at different stages of development.*

© 1999 Wiley-Liss, Inc.

Key Words: multisensory; cross-modal; superior colliculus; development; vision; somesthesia; audition

CROSS-MODALITY PERCEPTION AND BEHAVIOR

That having multiple sensory systems increases one's ability to detect external events seems self-evident; with more systems operational, more cues become available for this purpose. Not so obvious, but also of considerable importance in this regard, is the fact that when cues from different sensory modalities are present simultaneously, they can produce activity in the brain that exceeds the level that would be predicted by summing their individual influences [see Stein and Meredith, 1993]. This increases the salience of the signal in the brain, and consequently increases the likelihood of detecting the occurrence and initiating a response to it.

In behavioral experiments it has been shown that the probability of detecting and reacting to a visual and an auditory stimulus presented together is often greater than the combined probabilities of responses to the visual and auditory stimuli individually [Stein et al., 1989]. Moreover, in addition to the enhanced detection of an event composed of multiple sensory cues, there is an enhancement in the speed with which one reacts to that stimulus. In humans and nonhuman primates, where eye movements are frequently used as a measure of attention and orientation, presenting auditory and visual cues together can reduce the time taken to elicit an eye movement to below the minimum for either cue alone [Frens et al., 1995; Goldring et al., 1996; Hughes et al., 1994; Lee et al., 1991; Perrott et al., 1990; Zahn et al., 1978].

These synergistic processes are evoked most often when the cross-modality stimuli are closely associated with one another, such as when they are derived from the same event and therefore linked in space and time. However, there are many circumstances in which the presence of cues from multiple sensory modalities (or even the same modality) can interfere with one another, reduce detection probabilities, and/or give rise to perceptual illusions. This is especially true when the stimuli are spatially or temporally disparate but not so disparate that it becomes obvious that they are derived from different events. Indeed, some of the cross-modality perceptual illusions produced in these situations are so compelling that even substantial experience with a particular illusion has little effect on its potency. The profound nature of cross-modality effects has fascinated observers since the time of Aristotle, and questions such as why do we mislocate sounds in the presence of nearby visual cues (i.e., the "ventriloquism effect")? why when seeing the lips form one sound while hearing another sound are we led to perceive neither one nor the other, but a synthesis of the two (the "McGurk effect")? See M.C. McGurk and M. C. D.onald, 1976)? and why do some people invariably experience sensations in two modalities when presented with a single modality-specific stimulus (i.e., "synesthesia")?, are just a few examples of the issues with which current researchers are trying to come to grips.

Grant sponsor: NIH; Grant number: EY 06562; Grant number: NS22543.
*Correspondence to: Barry E. Stein, Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, NC 27157–1010.

© 1999 Wiley-Liss, Inc.
THE SUPERIOR COLLICULUS (SC) NEURON AS A MODEL FOR UNDERSTANDING THE NEURAL BASES OF MULTISENSORY INTEGRATION

Yet, despite our familiarity with the many ways our senses can interact to enhance or disrupt performance and/or perception, we have remained rather ignorant about the neural processes by which such interaction takes place, as well as when and how such processes develop. The lack of data was due, in part, to the already formidable task of understanding the fundamental operational (and developmental) neural properties of each individual sensory modality, the inherent difficulties involved in assessing neural interactions among modalities, problems in linking neural phenomena to specific perceptual and/or behavioral events, as well as difficulties in placing these observations into a meaningful developmental context.

In recent years, however, neurons in the midbrain of the cat, located in a structure called the superior colliculus (SC), have been used as a model system with which to explore each of these issues. The choice of this model system was based, in large part, on the fact that there are many multisensory neurons in the cat SC, and there was already a wealth of research detailing the importance of this structure in specific behaviors—specifically, orientation movements of the eyes, ears and head toward an object of interest [Casagrande et al., 1972; Schneider, 1969; Sparks, 1986; Sprague and Meikle, 1965; Stein and Meredith, 1991; Wurtz and Goldberg, 1972a,b]. Furthermore, many of the sensory and motor processes in the SC develop gradually during postnatal life [see Stein, 1984]. Consequently, multisensory integration in SC neurons can be assessed from neural, behavioral, and developmental perspectives, and unless specified otherwise, the data described below refer to neurons in the cat SC.

Although the anatomical and physiological features of the SC are quite complex, conceptually its primary functional organization is comparatively simple: incoming sensory information initiates activity in its neurons, and this activity is transformed into a motor signal that is sent to the brainstem and spinal cord to produce a coordinated movement of the eyes, ears, head, and body toward the source of stimulation (Fig. 1).

In the normal adult, neurons in the SC respond in large measure to sensory cues from the opposite (i.e., contralateral) side (e.g., a sensory cue to the left activates neurons in the right SC). In addition, many neurons respond prior to, as well as during, orientation movements (e.g., eye movements) directed toward contralat-
eral stimuli. At birth, few neurons respond to sensory stimuli, and those that do are responsive only to tactile stimuli [Stein et al., 1973a]. While it is not clear what specific contributions the tactile-responsive neurons in the SC make to the neonate’s sense of touch, tactile cues from the mouth region are critical for the kitten to locate its mother’s nipple and get into its mouth [Larson and Stein, 1984]. Yet, despite their early activity, these SC sensory neurons are distinctly immature. They have weaker responses, fatigue much more readily, and have far larger receptive fields (i.e., that region of sensory space in which a stimulus can activate the neuron) than their adult counterparts [Stein et al., 1973a; Wallace and Stein, 1996].

**THE MATURATION OF RECEPTIVE FIELDS AND SENSORY MAPS IN THE SC**

Because all responsive neurons in the newborn SC are somatosensory, the structure is considered purely unimodal (i.e., responsive to a single sensory modality) at this time. It is not until about 5 days postnatal (dpn) that some neurons become responsive to auditory cues, but they too are unimodal. Visual responses are the last to develop, and this does not occur until the third postnatal week. Although unimodal neurons predominate early in development, by the end of the second postnatal week neurons appear that respond to more than a single sensory modality. The development of such multisensory neurons follows a maturational sequence similar to unimodal neurons somatosensory-auditory neurons appear first, at about 12 dpn, and visual multisensory neurons appear last, close to the end of the third postnatal week. The incidence of multisensory neurons with all the various sensory convergence patterns increases gradually thereafter, until the adultlike complement is reached at about 12 postnatal weeks (see Fig. 2) [Wallace and Stein, 1996].

The receptive fields of all neonatal neurons are extremely large. Visual receptive fields encompass most of the contralateral visual field, and somatosensory receptive fields extend across much of the contralateral body. Neonatal auditory receptive fields are categorized as “omni-directional” because the neurons respond to stimuli anywhere in space. But, as postnatal maturation proceeds, there is a progressive shrinking of receptive fields in each of the modalities so that they become increasingly adultlike as development progresses [Wallace and Stein, 1996, 1997a,b; Wallace et al., 1993b]. When examined under “standard” experimental conditions (with all peripheral sensory organs aligned and facing forward), adult receptive fields exhibit a well-ordered topographic, or maplike, arrangement. Furthermore, it is evident that there is an impressive register among these maps. They are organized so that visual and auditory receptive fields of neurons in the rostral end of the SC are located in frontal or central space, and somatosensory receptive fields are located on the face or forelimb. As progressively more caudal SC locations are sampled, visual and auditory receptive fields move more peripheral in contralateral space, and somatosensory receptive fields move further back on the contralateral body. Similarly, along the medial-lateral axis of the SC there is a progression of receptive fields from superior (i.e., upper) to inferior (i.e., lower) sensory space. Consequently, a visual stimulus in the upper quadrant of central visual space will activate neurons medial and rostral in the structure. The development of these maps and their interrelationships is particularly evident in multisensory neurons (Fig. 3). For example, visual-auditory neurons, which initially have very large receptive fields, develop restricted fields that overlap one another in extrapersonal space.

Although the maturational time-frame is somewhat different, similar trends in the contraction of receptive fields with age have been noted in the SC of the monkey [Wallace et al., 1995], ferret [King et al., 1996], owl [Brainard and Knudsen, 1995] and mouse [Benedetti, 1991]. Presumably, a similar developmental trend in humans helps explain why the spatial concordance of visual and auditory stimuli becomes increasingly important to infants as they age [Morrison, 1994].

**THE MATURATION OF MOTOR MAPS IN THE SC**

The sensory maps are in register not only with one another, but with the structure’s motor maps for eye, ear and head movements [see Mclwain, 1991; Robinson, 1972; Roucoux and Crommelink, 1976; Schaefer, 1970; Sparks, 1986; Stein and Crammond, 1981]. Thus activation of sites progressively more caudal in the SC produces progressively larger eye, ear, and head movements toward presumptive targets in the periphery of the contralateral space (Fig. 4). Similarly, activation of the medial SC evokes upward orientation movements and activation of the lateral SC evokes downward movements [see Stein and Meredith, 1993].

Although far less is known about the maturation of the motor or sensorimotor aspects of the SC than is known about its sensory development, it has been postulated that the development of the motor organization precedes its sensory development [Sparks and Nelson, 1987]. Consistent with this idea is the demonstration that at least some eye movements can be evoked via direct electrical stimulation of the SC as early as 2 days postnatal [Stein et al., 1980]. This is long before its neurons are visually responsive (although tactile-evoked responses can be evoked in the late-term fetus see Stein et al. [1973a]) and long before visually guided behaviors are apparent [Fox, 1978; Nor-
Although the thresholds for evoking neonatal eye movements are high when compared with the adult, and only few sites in the SC are effective, those that are effective produce coordinated movements of both eyes in the appropriate direction. Thus, activation of sites in the left SC moves the two eyes rightward, and stimulation of similar sites in the right SC produces mirror-image leftward eye movements, just as in the adult [Stein et al., 1976, 1980]. These observations suggest that the premotor connections underlying visuomotor behaviors are being established even before natural visual inputs have access to this circuitry.

That some of the SC’s motor circuitry is already organized before visual information-processing takes place is consistent with theories suggesting that self-generated eye movements play an extremely important role in facilitating learning to make sense of visual inputs [Hain et al., 1979]. The simple passive movement of images across an animal’s visual field is a poor substitute for the more active participation reflected by self-generated eye movements, and animals that can move their heads but not their eyes do not develop a normal visual system [Garraghty et al., 1982]. Given that all of the sensory modalities represented in the SC have access to its motor outputs [see Stein and Meredith, 1993], it may be possible for early somatosensory and auditory responsive neurons to play a role in initiating the eye movements that could then set the stage for later visuomotor associations. It is difficult to evaluate the likelihood of this postulate at present. For, although there are some reports that in newborn infants and congenitally blind infants auditory stimuli can initiate appropriate eye movements [see Butterworth, 1981], other observers indicate that humans who have had no visual experience have little, if any, control over their eye movements [see Leigh and Zee, 1980].

**COMMON COORDINATES FOR SENSORY REPRESENTATIONS**

The preceding sections alluded to the development of topographic register among sensory maps as well as between the sensory and motor maps. Establishing and maintaining these relationships makes intuitive sense in light of the sensorimotor role of the SC, because whether a given location in space is the source of a visual, auditory, or somatosensory stimulus, or some combination of these stimuli, an orienting movement toward that location requires activation of the same region of the SC motor map. This condition is met by having a common map for the different senses. As discussed earlier, the cross-modality receptive field register is particularly obvious in multisensory neurons, and the importance of this in terms of integrating information from different modalities will be discussed in detail below.

Although it is simple conceptually, the establishment and maintenance of receptive field alignment across modalities is not a trivial task. This is because the coordinate systems of visual, auditory, and tactile space use distinctly different referents: visual axes are centered on the retina; auditory axes are aligned with respect to the head (and/or ears); and the location of a tactile stimulus is referred to the body surface. When all of the peripheral sensory organs are aligned (as when an animal has its eyes, ears and body directed forward), the axes of visual, auditory, and tactile “space” co-register, and it is in this circumstance that the receptive field alignment among multisensory neurons has been examined most often. But since the eyes, ears, head, and body can move independently (to a degree that is species dependent), there is no unique relationship between a coordinate in one reference frame and that in another. In other words, if receptive field register is to be maintained in the behaving animal, modality-specific signals cannot be encoded in their original reference frames (e.g. retina-centered, head-centered, body-centered).

A good example of this problem is when an animal moves its eyes independent of moving its ears and body, because this changes the relationship between the visual axis and those of the auditory and somatosensory domains. What happens in this circumstance underscores the dynamic nature of sensory maps. Specifically, compensatory mechanisms are initiated in an attempt to maintain receptive field alignment. Evidence that many SC auditory and somatosensory receptive fields shift as a function of eye position to maintain alignment with the visual receptive fields has been obtained in both cats.
and monkeys [Groh and Sparks, 1996; Hartline et al., 1995; Jay and Sparks, 1987a,b; Peck et al., 1995]. These observations suggest that multisensory representations are of a higher order, being the product of neural computations that transform signals derived from different modalities into a common coordinate frame. One possibility, as discussed by Sparks and colleagues [e.g., Sparks and Nelson, 1987], is that SC sensory maps are the end-product of coordinate transformations designed to represent modality-specific information in a common "motor" frame of reference. In such a scheme, the site of sensory-contingent activity in the SC is more closely linked to the movement required to orient to the stimulus than with the actual location of the stimulus itself. The consequence of this is a modality-independent code whereby stimulus location with respect to the current gaze position, not stimulus modality, is the major determinant of the spatial distribution of activity across the multisensory population.

**THE PRINCIPLES OF CROSS-MODALITY ("MULTISENSORY") INTEGRATION IN THE ADULT SC**

Establishing and maintaining overlapping representations in the brain is not simply a passive process. Rather, throughout development, active mechanisms operate to ensure that the maps are properly aligned with one another [e.g., see Benedetti, 1995; Groh and Sparks, 1996; Hartline et al., 1995; Jay and Sparks, 1984; King et al., 1988; King and Carlile, 1993; Knudsen and Brainard, 1991; Peck et al., 1995; Withington-Wray et al., 1990a,b], resulting in a common coordinate frame. This, in turn, ensures that, independent of modality, a stimulus originating from a given position in sensory space will produce activity in the same position in the SC and lead to similar motor output signals. It should also be noted that there is substantial value in the fact that, for any given mature multisensory SC neuron, there is a spatial alignment among its different receptive fields. As will be seen below, this provides a mechanism whereby SC neurons can respond differentially to a multisensory stimulus combination where each stimulus originates from the same location (presumably from the same event) versus those that originate from different locations (presumably from different, and, perhaps, competing events). These different responses, in turn, play an important role in enhancing (or degrading) the likelihood that an overt response will be generated [Stein et al., 1989; Wilkinson et al., 1996a]. But this requires some explanation and an understanding of the kinds of multisensory interactions that can take place in mature SC neurons.

**MULTISENSORY ENHANCEMENT**

As noted earlier, cues (e.g., visual and auditory) derived from the same event are likely to originate from the same location in space and, thus, are likely to activate the same multisensory neurons (because of the spatial register among the different receptive fields of each neuron). In this situation adult SC neurons transform the information derived from the two stimuli into an integrated product. This is manifested as a response that is substantially greater than the response to the most effective single-
modality stimulus [Meredith and Stein, 1986a; Wallace et al., 1996]. At times this enhancement can be sufficiently large that it far exceeds the response predicted by summing the two unimodal responses (Fig. 5).

In general, multisensory response enhancement is greatest when the individual stimuli are least effective and, thus, are difficult to perceive and/or ambiguous. This makes intuitive sense as the organism derives little benefit from enhancing the effects of cues that are already potent. Thus, the combination of poorly effective cross-modality cues is the most likely to lead to the largest enhancements, whereas the combination of individual stimuli of greater effectiveness is most likely to result in relatively smaller enhancements. This principle, known as "inverse effectiveness" [Meredith and Stein, 1986], is illustrated in Figure 6. As long as the combined response is statistically greater than the best unimodal response, the product is operationally defined as multisensory enhancement [Meredith and Stein, 1983]. This is more important than it might first seem, for as will be discussed below, the combination of two cross-modality stimuli need not lead to a response that is any greater than the most effective one of them. This is in fact, most often the case before multisensory neurons have developed the capability to synthesize cross-modality cues, and in conditions in which this capability is experimentally compromised.

**MULTISENSORY DEPRESSION**

If, in the example given earlier, the auditory stimulus is not presented at the same location as the visual, but is positioned outside its receptive field (and is, therefore, spatially discordant with the visual stimulus), response enhancement does not occur. In fact, under such circumstances there is a good probability that the exact opposite effect—response depression—will occur (Fig. 5). In such circumstances even a robust visual response can be rendered ineffective by presenting a spatially disparate auditory...
Fig. 8. Behavioral responses to multisensory stimuli are enhanced (or degraded) by spatial coincidence (or disparity) in much the same way as are the responses of SC neurons. Cats were trained to fixate on a central LED and to make an orientation movement to a target LED for food reward. **Top:** In spatial coincidence trials, visual and auditory stimuli were presented either alone or in combination (simultaneously) at the same location. Note that the stimulus combination resulted in enhanced behavioral performance at each of the seven locations tested. **Bottom:** In spatial disparity trials, the visual stimulus was presented alone or in combination with an auditory stimulus located 60 degrees disparate. For LED locations left or right of the midline, the auditory stimulus was displaced medially (i.e., visual at 60 degrees left, auditory at 0 degrees; visual at 60 degrees right, auditory at 60 degrees left). When the LED was on the midline (i.e., 0 degrees), the auditory stimulus was presented 60 degrees left and right. Note that responses to the visual stimulus at all locations tested were depressed when combined with a spatially disparate auditory stimulus. Adapted from Stein et al. [1989]. Reprinted with permission from the MIT Press.
stimulus [Kadunce et al., 1997; Meredith and Stein, 1996; Wallace et al., 1996]. Multisensory enhancement and multisensory depression can be evoked in many of the same SC neurons. Response enhancement is the result of the synergistic interaction of two excitatory inputs coming from stimuli appearing within their respective receptive fields. But, because some receptive fields have inhibitory or suppressive regions outside their borders, response depression is a consequence of the antagonism between the inhibitory input produced by the stimulus outside its receptive field and the excitatory input produced by the stimulus inside its field [Kadunce et al., 1997].

Not only is relative stimulus position a key feature in producing integrated multisensory responses, but so is the relative timing of the stimuli. When there is a substantial delay in when the two stimuli occur, they are treated as independent events, regardless of their spatial relationship. Nevertheless, this delay can be rather large and a multisensory interaction will still result [Meredith et al., 1987; Wallace and Stein, 1997a]. Such a temporal “window,” which can be hundreds of milliseconds long, makes it possible for two sensory modalities to interact despite very different latencies (i.e., the time it takes for the information to get to the SC) (Fig. 7). For example, the latencies for auditory and somatosensory responses in multisensory SC neurons are approximately 10–25 msec, while the range for visual responses is 40–120 msec. Clearly, if a neuron depended on an exact match between the latencies of two different inputs, the vast majority of visual-auditory and visual-somatosensory events would result in very little integration. This is certainly not what happens, in large part because a single visual, auditory or somatosensory input has a long-lasting effect on a multisensory SC neuron. Consequently, the long temporal window for multisensory integration compensates for the inherent latency differences among sensory modalities.

In addition, the differences in visual and auditory latencies are minimized by events that occur at a distance from the subject. The speed with which visual (i.e., light) and auditory (i.e., sound) signals travel in air is very different. For most biological purposes the speed of light can be considered infinite, so that the arrival of visual information at the eye is instantaneous. Thus, stimulus distance has little impact on the comparatively long visual response latencies. On the other hand, sound travels relatively slowly (approximately 345 m/s), so that the farther the stimulus is from the subject, the more the auditory information arriving at the SC is delayed. The result is that visual and auditory stimuli from events occurring at some distance from the subject may produce inputs that arrive at the SC neuron far closer in time than might have been expected based on their absolute latencies.

Furthermore, most SC neurons appear to be “tuned” to a specific cross-modality temporal disparity. This means that their peak multisensory responses are obtained at specific stimulus onset asynchronies, and that progressively weaker multisensory responses are evoked at longer or shorter temporal disparities. In the case of visual-auditory neurons this suggests that some of them may be most sensitive to events that take place at different distances from the subject. It
would be interesting to determine whether (and how) this information would be useful for SC information-processing and whether these cross-modality temporal disparity-sensitive neurons are arranged in any maplike representation in the SC.

**SC-MEDIATED MULTISENSORY BEHAVIORS**

The majority of SC multisensory neurons are output neurons that project into its major descending pathway, the predorsal bundle [Meredith et al., 1992; Wallace et al., 1993a]. This is the major route through which the SC contacts the contralateral brainstem and spinal cord and thereby controls orientation movements of the eyes, ears, mouth, and head [see Stein and Meredith, 1991]. Based on this, it seemed reasonable to predict that multisensory stimulus configurations that produce enhanced physiological responses would be more salient and, therefore, would facilitate SC-mediated attentive and orientation behaviors, whereas multisensory stimulus combinations that reduce physiological responses would be less salient, and would degrade SC-mediated behaviors. This prediction was tested in the following behavioral experiments.

Cats were acclimated to a hooplike apparatus in which lights (LEDs) and speakers were positioned at regular intervals (Fig. 8). They were then trained to look directly ahead, but as soon as a light was briefly illuminated, they were to orient toward its location and immediately approach that location in order to obtain a food reward. Some of the animals were also trained to approach the location from which a very brief, low-intensity sound originated. Since the modality-specific stimuli were brief and of low intensity, the animals had some difficulty detecting and locating them. Sometimes they failed to respond, and other times they moved to the wrong location. However, when the visual and auditory stimuli were presented simultaneously and at the same location, the animal's performance improved markedly (Fig. 8). Correct responding was enhanced far more than was expected based on statistical predictions. In other experiments, correct responding was also significantly enhanced in animals that were trained to ignore the auditory stimulus (it was never paired with reinforcement), and in animals that had no experience with the auditory stimulus.

On the other hand, in trials in which the auditory stimulus was moved 60 degrees medial to the visual stimulus, the probability of correctly responding now dropped well below that evoked by the visual stimulus alone. Once again, this result was independent of the specific training history and was obtained regardless of an animal's prior experience, or lack of experience, with the auditory stimulus. These behavioral results closely parallel those obtained in physiological studies, thereby confirming the prediction that there is a positive relationship between stimulus effectiveness at the level of the single SC neuron and at the level of overt attentive/orientation behavior [see Stein et al., 1989].

**THE MATURATION OF MULTISENSORY INTEGRATION**

Given the discussion above, it might seem reasonable to expect that when a developing SC neuron is first able to respond to inputs from different modalities, it would de facto have the ability to integrate these inputs as does a mature SC neuron. But this is not the case. As it turns out, the youngest multisensory neurons typically respond no better to the combination of two different sensory cues than to the most effective of them. Neither response enhancement nor response depression is seen. The absence of adultlike multisensory integration is apparent in newborn cat SC neurons of every modality convergence type [Wallace and Stein, 1997a], and similar findings have been obtained in newborn monkeys [see Wallace et al., 1995]. The conspicuous absence of multisensory integration in a visual-somatosensory neuron of a 20-day-old cat is shown in Figure 9.

At the end of the fourth postnatal week of feline development, an abrupt change occurs in some multisensory neurons, and they can now integrate cross-modality cues very much like the adult. Thus, response enhancement can be obtained with spatially coincident stimuli, response depression can be obtained with spatially disparate stimuli, and, just as in the adult, the greatest response enhancements are obtained with combinations of the weakest modality-
specific stimuli. Even the magnitudes of the response enhancement and response depression appear adultlike, and there is little evidence that these change systematically as a function of age (Fig. 10). What does change with age is the incidence of these "integrating" neurons (Fig. 10) and the temporal window during which multisensory integration can take place. Thus, there is a progressive increase in the average level of multisensory enhancement across the population of multisensory SC neurons as the animal matures (Fig. 10) and a progressive decrease in the requirement that cross-modality stimuli be synchronous. This last point deserves some further explanation.

As noted earlier, the broad temporal window of multisensory integration in mature neurons allows for a good deal of flexibility in the timing of cross-modality stimuli. In contrast, neonatal neurons exhibit little of this flexibility. They require that two stimuli from different modalities occur very close to simultaneity, and even a timing difference as small as 50 msec can eliminate an interaction (Fig. 11). This fits with the idea that, during the first few postnatal weeks, the only cues of importance for the neonate come from space immediately on and around the animal and will impinge on its respective sensory organs nearly simultaneously. With increasing age, the importance of more distant cues increases, the limited temporal window expands, and neurons begin to show preferences for specific stimulus onset asynchronies (Fig. 11). This fits with the idea that, during the first few postnatal weeks, the only cues of importance for the neonate come from space immediately on and around the animal and will impinge on its respective sensory organs nearly simultaneously. With increasing age, the importance of more distant cues increases, the limited temporal window expands, and neurons begin to show preferences for specific stimulus onset asynchronies (Fig. 11).

**MULTISENSORY INTEGRATION IN THE SC DEPENDS ON INFLUENCES FROM THE NEOCORTEX**

The observation that developing neurons that can already respond to cross-modality cues still require a good deal of maturation before they can integrate them as does the adult seems rather curious at first. But a wide variety of possible developmental events might account for this lag between the appearance of multisensory neurons and the appearance of mature multisensory integration. Among these are the (nonmutually exclusive) possibilities that some inherent, yet critical, changes in the membrane properties of the neuron must take place before adultlike integration is possible, or that some late-developing afferent influences are essential. Although the first possibility cannot be ruled out, it is only the second possibility for which there is empirical support. This evidence comes from experiments in which inputs to the SC that arise from an area of the neocortex (the anterior ectosylvian sulcus, AES) were reversibly silenced. The results were quite dramatic, and during the deactivation of the AES many multisensory SC neurons lost their integrative characteristics and began to respond no better to the combination of two cross-modality cues than to the most effective modality-specific cue.

This critical area of neocortex sends a robust projection to the SC [see Stein et al., 1983]. It is composed of three unimodal domains: somatosensory (SIV [Clemo and Stein, 1982]), visual (AEV [M. I. Uccke et al., 1982; Olson and Graybiel, 1987]), and auditory (Field AES, or "FAES" [Clarey and Irvine, 1990]). Multisensory SC neurons receive convergent inputs from unimodal neurons in these different AES subregions (Fig. 12) [see Wallace et al., 1993a], as well as from unimodal ascending pathways either originating from the sensory organ itself or relayed from other regions in the brainstem [see Edwards et al., 1979]. Thus, a visual-auditory SC neuron may receive ascending visual inputs from the retina and auditory inputs from the inferior colliculus, as well as descending inputs from the AEV and FAES subdivisions of the AES. The receptive fields of ascending and descending projections also are in fairly good register, and the SC receptive field reflects their combined territories (Fig. 12). From a functional perspective, however, the inputs from the neocortex are quite special, and eliminating their influences produces remarkably specific effects on the great majority of multisensory SC neurons: their capacity to integrate cross-modality is severely compromised, while their unimodal responses may be only modestly affected (Fig. 13) [Wallace and Stein, 1994].

That these AES corticotectal influences would have so specific a function is quite surprising. Yet, it seems particularly fitting that an area that, in classical terminology, had been considered an "association cortex" does indeed underlie associative functions. In this case its
associative role is a cross-modal one, and it uses the circuitry of a midbrain structure to effect it. This provides the AES with a means not only of controlling the integration of visual, auditory, and somatosensory information, but also of controlling the attentive and orientation functions the SC mediates. This became particularly evident from a parallel study using behavioral techniques. The paradigm was the same as that described earlier and illustrated in Figure 8.

The only difference was that there were cannulae chronically implanted in the brains of these animals so that drugs could be injected into specific regions. When an anesthetic agent, lidocaine, was injected into the AES, the response enhancement induced by spatially coincident visual and auditory stimuli was lost. In contrast, when lidocaine was injected into the SC, the response enhancement was preserved.

THE MATURATION OF MULTISENSORY INTEGRATION IN THE SC LIKELY REFLECTS THE ONTOGENY OF CORTICOTECTAL INFLUENCES FROM ANTERIOR ECTOSylvIAN CORTEX

Presumably, then, the lack of multisensory integration in neonatal SC neurons is due to the immaturity of inputs arising from the AES. The abrupt onset of multisensory integration that was noted in developing SC neurons might then reflect the rapid onset of a functional circuit between the AES and SC, much like the opening of a “gate” (see Wall and Stein, 1996). In fact, the rapid onset of cortical influences on SC neurons has also been noted in other contexts involving the complex properties of visual neurons (e.g., Stein and Gallagher, 1981). Preliminary support for this notion comes from observations that as soon as a neonatal multisensory SC neuron can engage in multisensory integration, the influence of AES on that neuron can be detected.

That multisensory neurons and multisensory integration develop very gradually in postnatal life is in contrast to expectations based on the idea of a “primitive unity” among the senses at birth (see Bower, 1977; Maurer and Maurer, 1988), which holds that the different senses are closely allied with one another at birth and that there is a very active transference of information and activity from one to another. Rather, these data on SC maturation are consistent with a sequential appearance of function among sensory modalities, with somatosensation appearing first, audition second and vision third (see Stein et al., 1973a,b; Wall and Stein, 1997a). This, in turn, is in keeping with the idea that for a short time during ontogeny there is a focus on developing the fundamental organization and information-processing capabilities of one sensory modality before repeating this process for another and before dealing with the most complicated task of integrating their different signals (see Turkewitz and K enny [1985] for a discussion of the possible adaptive
CONCLUSION

William James [1890] made a lasting impression on his readers when he described our initial impression of the world as a “blooming, buzzing confusion” of sensory stimuli. His point was that we require a good deal of experience with our senses before we can sort among external stimuli, relate them to one another, and relate them to specific consequences. These are necessary steps in learning about our world and understanding the consequences of our own actions. We know now that the very experience of using the senses has long-lasting effects on their neural organization and functional capabilities.

One of the effects that is of particular importance in the current context is the aligning of sensory receptive fields in multisensory SC neurons. The importance of this alignment for multisensory integration was stressed heavily in this discussion. Presumably, however, experience also can play a role in calibrating the impact of some sensory cues on multisensory neurons and even on some of the operating principles governing multisensory integration. While we have very little information about these issues, we do know that within the broad range of experiences that characterize normal development, at least some of the same principles that govern multisensory integration in SC neurons also govern multisensory integration in neocortical neurons [see Jiang et al., 1994; Ramachandran et al., 1993; Wallace et al., 1994]. Furthermore, recent evoked potential and imaging studies illustrate that multisensory processes characterize a number of neocortical areas [Barth et al., 1995; Paulesu et al., 1995; Sams et al., 1991]. The impact of multisensory integration at very different levels of the neuraxis is likely to have consequences on very different functions. This is, in fact, a most parsimonious way to coordinate the midbrain processes involved in overt behavior and the cortical processes involved in perception. The final result, of course, is the subjective impression of a seamless melding of behavior and perception.

ACKNOWLEDGMENTS

We thank Nancy London for her excellent assistance.

REFERENCES


