



Visual and Auditory Integration

JENNIFER M. GROH and URI WERNER-REISS

Dartmouth College

- I. Overview
- II. The Reference Frame Problem
- III. Brain Areas Involved in Saccades to Visual and Auditory Stimuli
- IV. Algorithms for Coordinate Transformations
- V. Parallels with Other Multisensory Motor Systems
- VI. Conclusion

GLOSSARY

binding The mutual association of all the component features of an individual object or event. Within the visual domain, the features that must be associated include shape, color, and location. When objects are detected by more than one sense, binding refers to the association of all the different sensory cues generated by the same object or event.

saccade A rapid eye movement that brings the fovea to bear on a stimulus of interest. Saccades can be guided by visual, auditory, tactile, or remembered stimuli.

receptive field The range of locations from which a stimulus must originate in order to modulate the activity of a neuron.

movement field The range of eye movements in space that are associated with a modulation in the activity of a neuron. A movement field is essentially a receptive field for motor-related activity.

frame of reference The particular set of axes used to describe a location in space.

spatial registry Perfect alignment between the locations of the unimodal receptive fields in a multimodal neuron; also, alignment between the receptive fields of neighboring neurons responsive to different modalities.

spatial correspondence A systematic relationship falling short of registry between the unimodal receptive field locations in individual multimodal neurons. Correspondence can also occur between the receptive fields of neighboring unimodal neurons that respond to different modalities.

linear function A function $f(x) = mx$, where m is a constant. Linear functions have the property of superimposibility: $f(a + b) = f(a) + f(b)$

microstimulation A technique for activating a small population of neurons (*in vivo*) by passing electrical pulses into the neural tissue through an electrode.

Our experience of the world as a unitary place requires that our brains meld together the information gleaned from our different sense organs. We can then act on this information, regardless of the source. The multisensory neurons in the brain may contribute to this process. In this article we review the properties of multisensory neurons in areas of the brain that participate in one kind of multisensory behavior: saccadic eye movements.

I. OVERVIEW

Knowledge of the world comes to us through our different senses. Our eyes monitor light patterns, our ears detect air pressure changes, our skin reports the pressure of objects impinging on it, and our noses and mouths measure the chemical content of the sensory scene. Despite the variety of routes that sensory information takes to reach our brains, we tend to experience the world as one place. Furthermore, we are capable of using different types of sensory information in equivalent ways, regardless of which sensory system was initially responsible for detecting the stimulus. For example, object shape can be detected either visually or through our sense of touch. Saccadic eye movements can be directed to the locations of visual, auditory, or tactile stimuli. When discrepancies occur between different sensory cues, one sense can dominate an-

other. When visual and auditory stimuli are mismatched in space, vision sometimes “captures” the location of the sound, a phenomenon known as the ventriloquism effect.

How is the mental experience of a combined multisensory event constructed from a collection of neurons sensitive to the individual components? Called the binding problem, this class of issues has received considerable attention within a single sensory domain. Vision scientists wonder how our percept of objects is synthesized from the activity of separate populations of neurons encoding individual visual features such as color, contours, and motion. For multisensory binding, the information to be bound together comes from many different sensory systems. Many neurons in the brain respond to more than one sensory modality; such neurons serve as a potential neural substrate for this binding. Multisensory neurons are actually common, having been identified in the parietal, temporal, and frontal lobes as well as in various brain stem areas. In this article, we discuss primarily visual and auditory integration. We take as our model system the areas of the brain that are responsible for guiding saccadic eye movements to visual and auditory stimuli. We review the response properties of neurons in these areas as well as those in selected regions of the brain that provide input to these areas. All the experiments we describe were conducted in either cats or monkeys unless otherwise noted. We seek to answer the following questions:

1. How do the responses of multimodal cells respond to each of the modalities alone? Are the receptive fields in the same frame of reference? Are the visual and auditory receptive fields of individual neurons in spatial register with one another?

2. How do they respond to combinations of stimuli? Is the response to a combined stimulus the linear combination of the responses to the components, for example?

In many instances, the complete answers to these questions are not known. Nevertheless, they will serve as a useful framework for exploring the current state of multisensory research.

II. THE REFERENCE FRAME PROBLEM

How does the brain determine that two pieces of sensory information arise from a common source? Visual and auditory stimuli are more likely to be the

result of the same event if they arise from the same location, so determining whether two stimuli of different modalities are coming from the same place is an important task for the brain. However, this task is complicated by the fact that the visual and auditory systems employ different reference frames for their initial encoding of stimulus location. Visual stimuli are initially represented in retinal coordinates. Light is reflected from objects in the world, enters the eye, and activates a particular location on the retina. The site of retinal activation informs the brain concerning the location of the visual stimulus with respect to the direction of gaze. When the eye moves, the site of activation on the retina shifts.

The locations of auditory stimuli are computed based on interaural timing and level differences as well as spectral cues, yielding a head- and ear-centered frame of reference. If the source of a sound is located to the right, the sound will arrive in the right ear first, and it will be louder in the right ear. These interaural timing and level differences can be used to compute the location of the sound with respect to the orientation of the head and ears. Obviously, the position of the eyes would be irrelevant.

In species whose eyes have significant mobility (humans, monkeys, cats, but not barn owls, for example), these different reference frames pose a problem. Moving the eyes without moving the head will affect the location of retinal activation without influencing the auditory cues generated by a given object. On the other hand, moving the head while keeping the eyes pointed in a stable direction in space will alter the interaural acoustic cues without affecting the site of retinal activation. The solution to this problem appears to be a transformation of auditory signals into an eye-centered frame of reference so that comparison with eye-centered visual information can take place more easily.

III. BRAIN AREAS INVOLVED IN SACCADDES TO VISUAL AND AUDITORY STIMULI

Saccades are rapid eye movements that direct the fovea to the positions of stimuli of interest. Saccades may be elicited by visual, auditory, or tactile stimuli. Our discussion focuses on some of the motor areas that have been implicated in the control of saccades to these visual and nonvisual stimuli as well as one of the areas that provides auditory input to these areas. The superior colliculus (SC), frontal eye fields (FEF), and

lateral intraparietal areas (LIP) are the major motor or motor-related areas that we discuss. One way that auditory signals reach this pathway is through an auditory cortical area in the anterior ectosylvian sulcus (AES, cat; the location is unknown in primate). Other sensory areas such as the inferior colliculus undoubtedly also contribute, but little is known about the role they play in multisensory processing so they will not be discussed.

A. Superior Colliculus

1. Oculomotor Function

The SC is a laminated structure located on the roof of the midbrain. The superficial layers receive a direct projection from the retina, but the intermediate and deep layers are primarily motor in function. Microstimulation in these layers elicits saccadic eye movements with a very short latency. The direction and amplitude of the evoked movement depend on the location of stimulation in the SC. In primates, the left SC encodes conjugate eye movements toward the right and vice versa. Movement size increases as the stimulating electrode is moved from rostral to caudal positions. Movements with an upward component are represented medially, whereas downward movements are represented laterally. Recording studies reveal that a prominent feature of the discharge characteristics of SC neurons is the vigorous burst of activity that they generate just prior to saccades. This activity is selective for the direction and amplitude of the saccade. The most vigorous activity occurs when the saccade will be directed to the center of the cell's so-called movement field. Movement fields are topographically organized and correspond to the direction and amplitude of the saccade evoked by stimulation at that site. Lesions of the SC, when combined with lesions of the FEF, abolish saccades completely. Taken together, these features place the SC convincingly within the saccade pathway.

2. Sensory Properties

The SC is not just a motor structure, however. Neurons in the intermediate and deep layers of the SC discharge in response to sensory events in three different sensory modalities: vision, hearing, and touch. These responses are sensory in the sense that they occur time locked to the onset of the sensory stimulus and they do not appear to actually trigger an eye movement. In other words, the neural responses

occur even if the animal is trained to delay its saccade to the location of the stimulus until some kind of "go" cue is received. Although the responses can be dissociated from saccades in this fashion, the responses are considerably more robust if the animal subsequently makes a saccade to the stimulus than if the animal is asked to ignore the stimulus. Thus, the true character of the discharge is intermediate between purely sensory and purely motor. Indeed, individual neurons can show both sensory-related and motor-related properties. A schematic of an idealized SC neuron's response is shown in Fig. 1. When an animal is performing a delayed-saccade task, neurons in the SC may show a brief burst when the sensory stimulus is first turned on, followed by some level of sustained activity, and culminating in a second burst beginning just prior to the saccade. Actual individual neurons do not necessarily show activity during all three of these epochs.

Many different but overlapping schemes have been used to classify different cell types found in the intermediate and deep SC. Neurons that exhibit a burst of activity at approximately the time of the saccade are usually referred to as saccade-related burst neurons. Neurons that have both visual and motor-related activity are frequently referred to as visuomotor cells. Neurons that have both sustained and motor-related activity have also been called prelude-bursters. Finally, neurons showing any type of activity prior to

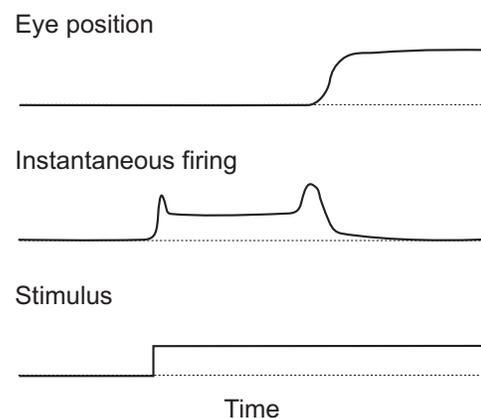


Figure 1 Response profile of an idealized SC or FEF neuron. The top trace illustrates the movement of the eyes, and the bottom trace shows the onset of a stimulus. (Middle) The discharge of the neuron is correlated with both the sensory stimulus and the movement. A burst of activity occurs at target onset and is sustained until another burst occurs at movement onset.

the saccade-related burst have been referred to as “buildup” neurons in many studies. However, this term is not particularly apt because the activity has not been shown to build up—the discharge is typically better characterized as being sustained at a steady level until the saccade-related burst occurs. All of these terms describe overlapping populations of SC neurons.

Early studies of the multisensory properties of these neurons were conducted primarily in anesthetized, paralyzed animals such as cats, rodents, and barn owls. Under these conditions, any motor-related properties of the cells cannot be determined. Instead, cells were classified based on their responses to visual, auditory, or tactile stimuli. Unimodal, bimodal, and trimodal cells have all been identified. Circumstantial evidence suggests that the multimodal cells studied in anesthetized animals might be part of the population of cells that show motor-related activity in awake animals: Multimodal neurons can usually be activated antidromically by microstimulation in the efferent output pathway from the SC. This suggests that they are a source of output from the SC, a role that is likely to involve saccade-related activity.

3. Receptive Field Registry and/or Correspondence

When sensory responses occur, the existence and location of the receptive field(s) become an important issue. Before delving further into this topic, it is important to define our terminology. At issue is the question of the relative locations of visual, auditory, and tactile receptive fields of individual SC neurons. We define the term “spatial registry” to mean that an individual SC neuron responds to stimuli of these different modalities at the same location in space. For example, does the cell respond to the sight, sound, and touch of a mosquito on the back of the hand? If so, then the visual, auditory, and tactile receptive fields may be said to be in register. In contrast, we define “spatial correspondence” to mean that the unimodal receptive fields of multimodal neurons bear a systematic relationship to one another but are not necessarily in truly the same location in space. For example, if most neurons with tactile receptive fields on the hindquarters (outside the field of view) had visual receptive field in the lower eccentric region of the visual scene, the tactile and visual receptive fields would clearly be correlated in a nonrandom way, but because they are not actually in the same position in space they would not be in true register.

In anesthetized cats, visual and auditory receptive fields are usually in spatial register. Alex Meredith and Barry Stein compared the visual and auditory receptive field locations in multimodal neurons and reported that the area of spatial overlap averaged 86% of the smaller receptive field. Somatosensory receptive fields show a spatial correspondence with visual or auditory receptive fields: Tactile receptive fields on the face tend to be correlated with central visual receptive fields, whereas tactile receptive fields on the flank or hindquarters tend to cooccur with more peripheral visual receptive fields (Fig. 2). These representations are

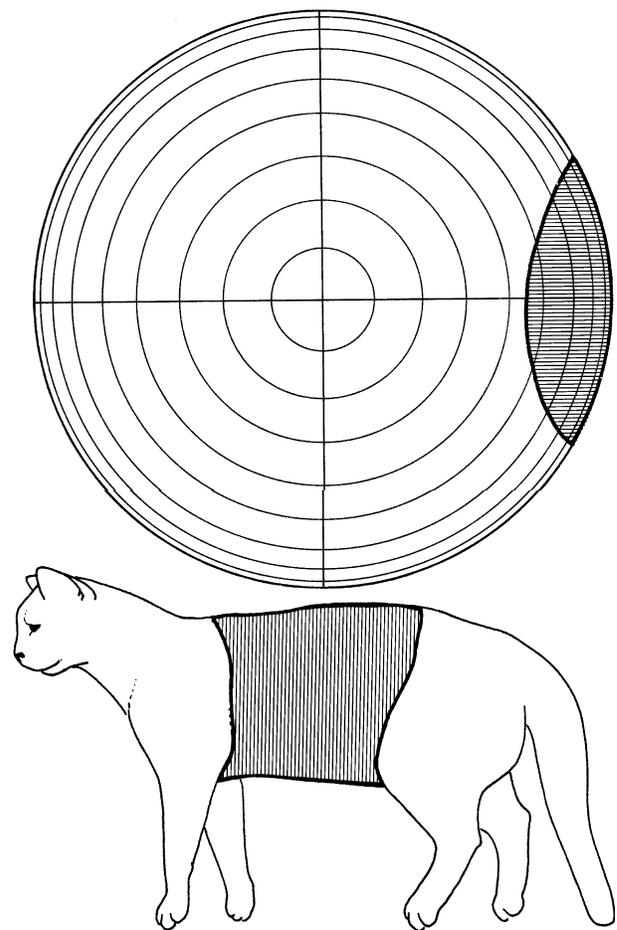


Figure 2 Spatial correspondence between the visual and tactile receptive fields in cat SC. The shaded regions indicate the receptive field locations. The visual (top) and tactile (bottom) receptive fields are closely related to one another but not in register because the body surface containing the tactile receptive field is actually out of the animal's field of view [adapted with permission from Meredith, M. A., and Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons, *J. Neurophysiol.* **75**, 1843–1857].

topographically organized, creating maps of the visual, auditory, and tactual scenes. In the rostral part, the auditory and visual receptive fields are centrally located, and the somatosensory receptive fields are on the face or forelimb. Neurons located more caudally have receptive fields located more peripherally in the contralateral space, with tactile receptive fields on the back, hindlegs, or tail. The sensory maps also generally correspond with the motor map for eye movements.

4. Frames of Reference

A priori, one might have expected that registry between receptive fields would be quite poor, due to the fact that the different sensory systems employ different frames of reference. Thus, a visual receptive field anchored to the retina and an auditory receptive field anchored to the position of the head and ears would only match up if the eyes happened to be in the one and only position that would result in alignment. The convergence of visual, auditory, and tactile information onto the SC therefore prompted investigations into the frame of reference of these signals.

Martha Jay and David Sparks were the first to test the influence of eye position on collicular auditory receptive fields. They trained monkeys to make saccades to the locations of auditory stimuli from different initial eye positions. The animals' heads were held fixed so that eye movements were dissociated from head movements. Jay and Sparks found that many of the auditory receptive fields shifted when the eyes moved (Fig. 3), a result that has been confirmed in subsequent studies involving awake cats. Although in principle this should permit perfect registry between visual and auditory receptive fields to be maintained when the eyes move, the alignment of visual and auditory receptive fields was not tested in this experiment. Paradoxically, their findings may actually cast doubt on the possibility of perfect alignment: The receptive fields did not shift as much as they should have to maintain a perfect eye-centered representation of the location of sounds. Instead, the receptive fields shifted an average of only about half the amount of the eye movement, which should disrupt the registry with visual receptive fields. Thus, the alignment of visual and auditory receptive fields in the SC of the awake primate is deserving of further investigation to resolve this apparent conundrum.

An investigation into the frame of reference of tactile signals in the SC was conducted by Jennifer Groh and David Sparks. Monkeys were trained to

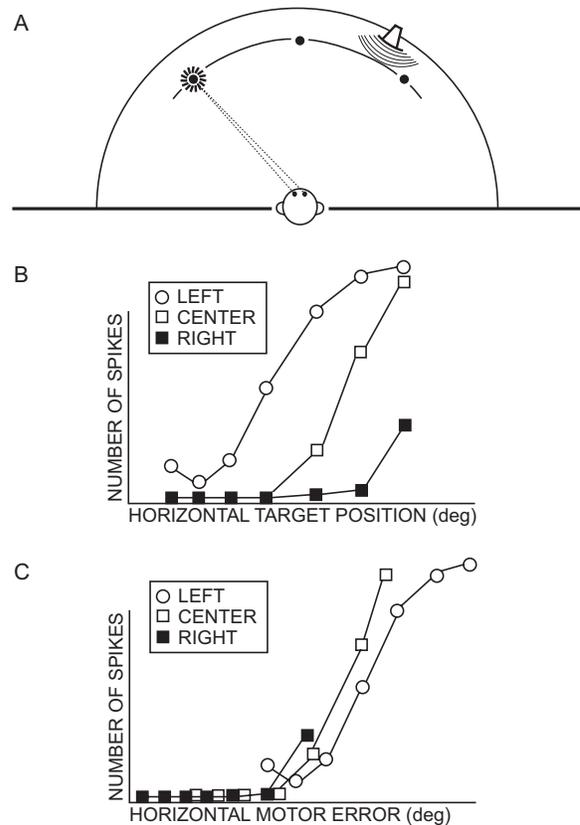


Figure 3 Effect of eye position on auditory responses of a neuron in the SC. (A) Experimental design. Monkeys fixated one of three LEDs while sounds were presented from a range of locations using a speaker mounted on a rotating hoop. Neural responses were better correlated with the location of the sound with respect to the eyes, or motor error (C), than with the location of the sound in space (B) [adapted with permission from Jay, M. F., and Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* 309, 345–347].

make saccades from a variety of different initial eye positions to the locations of vibrotactile stimuli delivered to the palms of the hands. The locations of the somatosensory stimuli on the body surface were held constant, as was limb position. Under these conditions, the positions of the eyes influenced the responses of many SC neurons to tactile stimuli. The effect of eye position was suggestive of an eye-centered frame of reference: Cells tended to respond best to the tactile stimulus when the initial position of the eyes brought the location of the tactile stimulus into the visual receptive field or movement field of the neuron under study.

Although the frame of reference of visual and saccade-related activity in the SC is indeed eye

centered, it is more properly described as a motor error or updating eye-centered frame of reference than a strictly retinal frame of reference. This was demonstrated in 1980 by Larry Mays and David Sparks. They recorded from neurons in the SC that were visually responsive in an ordinary visual saccade task. They then asked monkeys to make saccades to a sequence of two briefly flashed visual stimuli. Both stimuli were illuminated and then extinguished before the first saccade began (Fig. 4). Thus, the direction and amplitude of the saccade to the second target were dissociated from the original retinal locus of that target. Monkeys can perform this task quite well.

If the saccade vector between the first and second targets corresponded to the location of the receptive field, then many cells discharged between the first and second saccades, even though the retinal locus that

would normally be associated with that saccade vector had never been activated. Thus, although these cells respond to visual stimuli, in fact they can also represent the remembered locations of visual stimuli in a frame of reference anchored to the eyes—a frame of reference that is updated when the eyes moved. These characteristics led them to be called quasi-visual cells.

5. Multisensory Integration

If a cell responds to both visual and auditory stimuli separately, what happens when the two stimuli are presented together? One might expect individual neurons to treat stimulus energy as equivalent, regardless of the source of the energy. Thus, the response to a joint stimulus might be approximately equal to the sum of the responses to the individual components. In

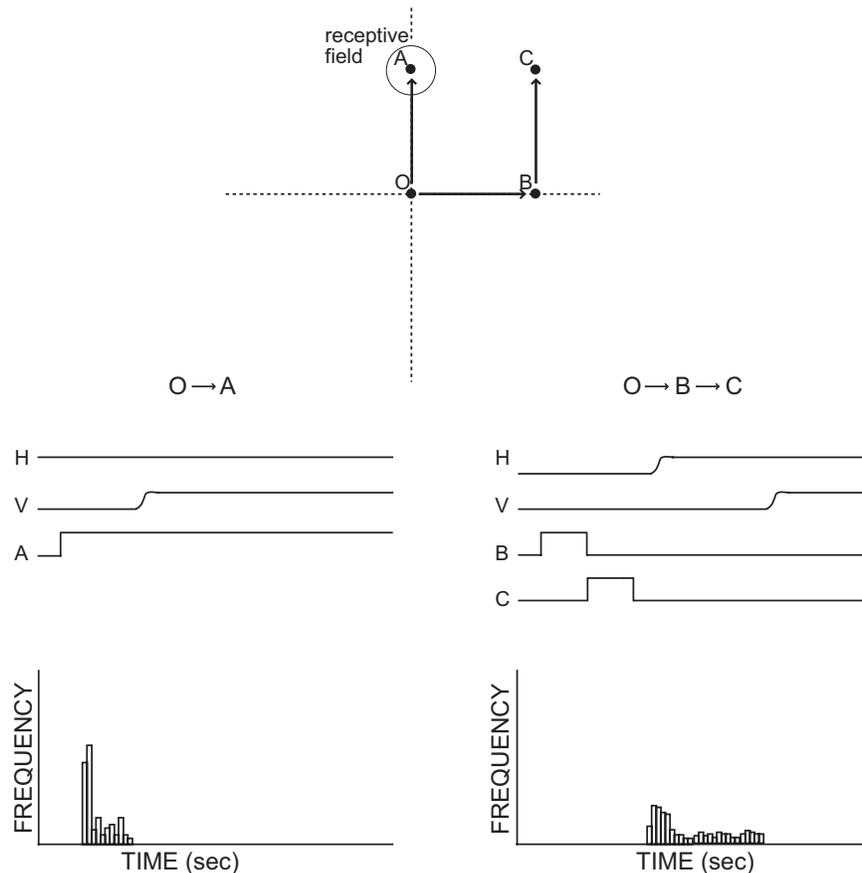


Figure 4 Idealized response of a quasi-visual cell in the SC. (Top) The locations of receptive fields and various visual stimuli. (Bottom left) The response of the cell when the visual stimulus is presented in the receptive field and the monkey makes a saccade directly to it. (Bottom right) The response of the cell when no stimulus is presented in the retinal location corresponding to the receptive field. Rather, a sequence of two stimuli is presented. After the animal makes a saccade to the first, the remembered location of the second lies in the new position of the receptive field and the cell becomes active [adapted with permission from Mays, L. E., and Sparks, D. L. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *J. Neurophysiol.* **43**, 207–232.

other words, SC neurons might behave linearly. Violations of linearity should occur for weak stimuli near the threshold of the neuron, where the combined response should exceed the sum of the two responses, and again for strong stimuli, when responses to combined stimuli might saturate at a level below the sum of the unimodal responses.

An alternative means of characterizing multisensory interactions has been employed by Meredith, Stein, and colleagues. These researchers emphasized the potential benefit for stimulus detection that may be provided by combined sensory stimuli. They therefore compared the joint response to the larger of the two unimodal responses using the following formula:

% Interaction =

$$\frac{(\text{joint response} - \text{maximal unimodal response}) \cdot 100}{\text{maximal unimodal response}}$$

Response enhancement is defined as a joint response exceeding the maximal unimodal response, and depression is defined as a joint response that is less than the maximal unimodal response. If the joint response equals the maximal unimodal response, this is defined as “no interaction.” Comparing this classification scheme with linearity, any interaction that exceeds 100% must be supralinear: If the joint response is more than twice as large as the largest unimodal response, it must also be larger than the sum of the two unimodal responses. Similarly, any interaction that is less than 0% must be sublinear: If the joint response is less than the maximal unimodal response, it must also be less than the sum of the two unimodal responses (assuming the weaker response is not inhibitory). However, between these two extremes a comparison between the percentage interaction and linearity cannot be made without knowledge of the individual component responses.

When Meredith, Stein, *et al.* characterized the multimodal responses in terms of percentage interaction, they found evidence of all three types of effect: enhancement, depression, and no interaction. The type of interactions depended on three factors: the effectiveness of the component stimuli at driving the cell, the spatial locations of the component stimuli, and the timing between the component stimuli. Response enhancement tended to be largest when the component stimuli evoked only modest responses on their own. Enhancement tended to occur when both stimuli were in the center of their respective receptive fields, whereas depression was more likely to occur when one of the two stimuli was in the suppressive area of its receptive

field. The degree of enhancement or depression was largest when the stimuli were presented within the same time frame (with less than 100 msec separating the two). Specifically, stimulus-onset asynchronies small enough that the responses to the individual components would overlap in time tended to produce maximal enhancement. Enhancement decreased with greater stimulus-onset asynchrony, and when asynchrony was increased even further depression often occurred.

Both methods of characterizing multisensory interactions imply certain assumptions. As mentioned previously, comparing the joint responses to the sum of the individual responses suggests a null hypothesis that stimulus energy is stimulus energy, regardless of which sensory system detected the event. Two stimuli have more energy than one, so the response to two should be close to the sums of the individual responses (i.e., linear). In contrast, the percentage interaction (enhancement/depression) analysis emphasizes the relationship between the combined response and the stronger unimodal response. When the two are equal, this is defined as no interaction. However, failure of the joint response to exceed the maximal unimodal response might actually indicate that the stronger unimodal stimulus suppresses any response to the weaker unimodal stimulus. This would hardly be an inconsequential interaction between the sensory signals. Large enhancements (those that exceed linearity) clearly indicate synergistic facilitation between sensory inputs, but anything less than a large enhancement may indicate equally interesting competition between the different types of inputs to individual neurons. Further work is needed using both types of analysis to clarify these issues.

B. Frontal Eye Field

The FEF is a cortical region that is involved in controlling saccadic eye movements. The FEF is similar to the SC in many ways. In awake monkeys, many FEF neurons discharge in conjunction with saccades. Cells have movement fields and discharge most vigorously if the saccade will be directed to the center of the movement field. The idealized FEF neuron has an activity profile with sensory and motor components similar to the idealized SC neuron shown in Fig. 1. Microstimulation in FEF evokes saccades. As mentioned earlier, combined lesions of FEF and SC completely abolish saccades.

Like the SC, the FEF has been implicated in the control of saccades to sounds as well as visual stimuli.

Gary Russo and Charles Bruce investigated the motor-related activity of neurons in the FEF to determine if they were active for saccades to auditory stimuli as well as visual stimuli. Monkeys were trained to make saccades to the locations of both lights and sounds. They found that all the cells in their sample responded prior to saccades to both kinds of targets. Furthermore, the movement fields for both were anchored to the initial position of the eyes. These results, were similar to those of Jay and Sparks for primate SC.

Neurons in FEF also respond to sensory events. Many respond to visual stimuli and a few respond to sounds. Neurons that respond to sounds can be driven by a variety of different types of sound stimuli, and receptive fields tend to lie on the contralateral side. About half of auditory neurons also respond to visual stimuli. Acoustically responsive cells tend to be found primarily at locations in FEF that represent large amplitude movements. Beyond this, little is known. The frame of reference of these sensory responses has not been examined, nor have the interactions between visual and auditory stimuli been characterized. Do bimodal cells have receptive fields that are in register with one another? Do they stay in register when the eyes move? Do the cells show response enhancement/depression to combined stimuli? How do the responses to combined modality stimuli compare with the linear sum of the unimodal responses? Further experiments are needed to answer these questions.

C. Lateral Intraparietal Cortex

Posterior parietal cortex is a region of the brain that has been implicated in spatial attention and eye movements. In humans, lesions of the parietal lobe cause patients to ignore sensory stimuli located in the contralateral half of space, a phenomenon known as neglect. What is the frame of reference of this neglect? One intriguing experiment that addressed this question was conducted by Edoardo Bisiach and Claudio Luzzatti in 1978. They asked patients with parietal lobe lesions to imagine themselves facing the cathedral at the Piazza del Duomo in Milan and report all the landmarks they could recall. From this imaginary vantage point, the patients named primarily landmarks on the side of the piazza ipsilateral to the side of their brain lesion. Then, they were asked to imagine themselves standing on the cathedral steps, facing in the opposite direction. From this opposite perspective, the deficit was reversed. They could now recall landmarks on the opposite side of the piazza, but they failed

to name landmarks on the first side—landmarks that they had been able to recall from the first imaginary vantage point.

In monkeys, the lateral intraparietal region of posterior parietal cortex, or area LIP, is believed to play a role in guiding saccades to the locations of visual, auditory, and remembered stimuli. Various groups have debated whether the function of LIP is better characterized as attentional, overseeing the sensory processing of stimuli at a particular location, or “intentional,” preparing the motor circuitry to make a saccadic eye movement to a stimulus at that location. Clearly, these two concepts are closely related, perhaps so much so that no experiment could adequately distinguish between them. We therefore remain agnostic on this point and will merely review the relevant literature concerning the representations of visual, auditory, and saccade-related information in LIP.

Neurons in LIP resemble those in both SC and FEF in having both sensory and motor-related activity. For many neurons in LIP, the visual responses are not strictly dependent on the location of the stimulus on the retina. The position of the eye is also important. Richard Andersen, Vernon Mountcastle, and colleagues found that visual stimuli at a given retinal location evoked different responses depending on where the animal’s eyes were pointing (Fig. 5). The effects of eye position were dubbed “gain fields” because the level of the visual response was affected by orbital position. Although the term gain normally implies a multiplicative interaction, whether this eye position effect is truly multiplicative has not been studied. Nevertheless, the location of the receptive field stays anchored to the retina when the eyes move, but the responsiveness of the cell varies. Thus, the frame of reference of visual activity is not strictly retinal but reflects the combination of retinal and eye position signals.

Larry Snyder, Peter Brotchie, Richard Andersen, and colleagues further investigated these gain fields to determine whether the position of the eyes with respect to the head, body, or world was the important factor. They varied the positions of the eyes with respect to the head, the position of the head with respect to the body, and the position of the body with respect to the world while keeping the visual stimulus at a fixed position on the retina. These manipulations revealed that the magnitude of the response to a given retinal stimulus depended on the position of the eyes with respect to the body. In LIP, it mattered little whether gaze shifts with respect to the body were accomplished largely by eye-

VISUAL AND AUDITORY INTEGRATION

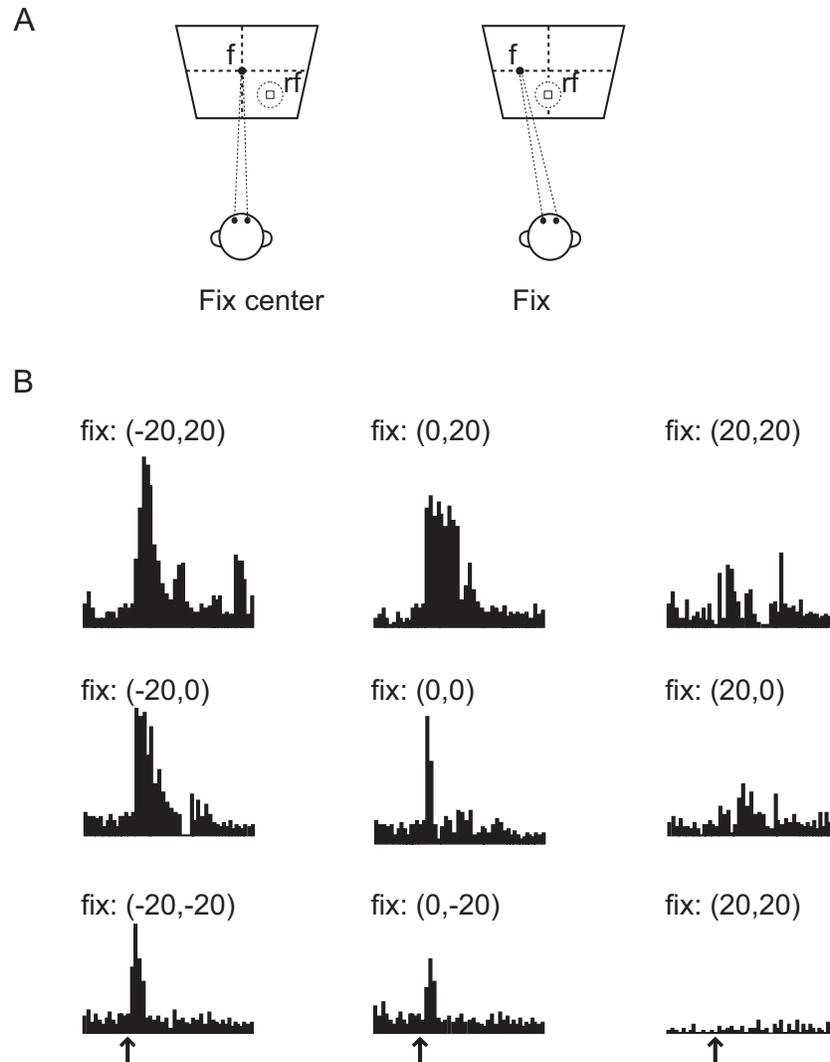


Figure 5 Effects of eye position on visual responses in posterior parietal cortex. (A) Schematic of experimental design. The responses of neurons to visual stimuli at a particular retinal location were examined for a range of orbital positions. (B) Responses varied with eye position, although retinal location was held constant [adapted with permission from Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science* **230**, 456–458].

in-head rotations or head-on-body rotations. (In nearby area 7a, however, more cells were affected by the position of the eyes with respect to the world than by the position of the eyes with respect to the body.)

What about auditory signals in LIP? Pietro Mazzoni, Richard Andersen, and colleagues examined how LIP represents sound location. They presented sounds from two speakers, one located contralaterally and the other in the ipsilateral field. They found that approximately half of neurons in their population responded to sounds at one or both of these locations. Most responded more vigorously from one speaker than

from the other. Most neurons also responded to visual stimuli and tended to prefer visual stimuli located near the preferred speaker. However, receptive fields were not mapped in this study, so it is not known how extensive the visual and auditory receptive fields are, nor how closely they correspond with one another.

The frame of reference of auditory signals was examined by Brigitte Stricanne, Richard Andersen, and Pietro Mazzoni. These researchers trained monkeys to make saccades to the locations of five speakers separated 12 degrees apart along the horizontal meridian. Three fixation positions 12 degrees apart

were used. An analysis of variance confirmed that the location of the sound was an important determinant of neural activity. The location of the sound with respect to the head, the location of the sound with respect to the eyes, or both of these factors were predictive of how cells would respond. However, if receptive fields are large relative to the separation between fixation positions, it would be difficult to determine from this analysis whether the activity is better correlated with head vs eye-centered frames of reference because the head- and eye-centered locations are correlated with one another. Nevertheless, this study does suggest that eye position influences the responses of at least a subset of auditory neurons in LIP.

To summarize, LIP contains neurons that respond to auditory stimuli, but many important questions remain unanswered. What is the relationship between visual and auditory receptive fields? Do they overlap? Are they in register? What is the frame of reference? Furthermore, nothing is known about how the cells respond to combined visual and auditory stimuli. Do the cells sum their inputs linearly? Do they show the kind of response enhancement/depression that is found in the SC? These issues await further inquiry.

D. Anterior Ectosylvian Sulcus

In cat, a major source of descending input to the SC derives from the AES. This descending input is of particular interest because this cortical region (whose primate homolog is unknown) contains subdivisions for different sensory modalities. Visual responses are present in area AEV, somatosensory activity occurs in area SIV (the so-called fourth somatosensory cortex, pronounced “S4”), and auditory signals are found in field AES. At the periphery of these unimodal subdivisions are multimodality neurons that do not project directly to the SC. Overall, approximately one-third of the cells are auditory, one-fourth somatosensory, and one-fifth visual. The remaining cells are bimodal or, rarely, trimodal, with visual–auditory cells being most common in this group.

Lee Wilkinson and colleagues investigated the functional role of AES in integrating multisensory information by deactivating it and testing the performance of cats on a task involving multisensory cues. The cats were trained to orient toward a visual target. When the visual target was dim enough to be near threshold for detection, orientation performance could be improved by pairing the visual stimulus with a sound at the same location. Pharmacological deac-

tivation of the AES prevented this facilitation by the paired auditory cue. Orientation to a visual stimulus alone was not affected by the deactivation. Inactivation of primary auditory cortex had no such effect. These results suggest that area AES plays an important role in facilitating orienting movements based on auditory information.

Inactivation of AES has also been shown to affect the responses of SC neurons in an interesting way. When AES is cooled, the responses of SC neurons to unimodal stimuli are unaffected. However, the responses to combined stimuli cease to show the same type of interaction (enhancement or depression) that they showed prior to the inactivation. This result is particularly surprising given that the projection from AES to SC derives from the unimodal cells, not the multimodal cells in that region.

Although they do not project to the SC, the multimodal cells in AES do share many of the properties of multimodal cells in SC. Receptive field locations for the different sensory modalities show a correspondence in space. Visual–somatosensory cells with tactile receptive fields on the forelimbs tend to have visual receptive fields in the lower quadrant of space and so forth. This correspondence is surprising because the unimodal subdivisions of the AES region do not seem to have a topographic organization. As in the SC, multimodal cells show response enhancement when the relevant stimuli are presented simultaneously in their respective receptive fields. How their responses to combined stimuli compare with the linear sum of the responses to the individual components has not been tested.

Our knowledge of area AES is currently quite preliminary. A number of questions remain to be answered. How do these cells respond in the awake, behaving animal? What is the frame of reference of these signals? Are the receptive fields of multimodal cells in true spatial register, and how is the registry influenced by movement of the eyes, ears, head, and body? What is the primate homolog? Experiments in awake monkeys are needed to address these issues.

IV. ALGORITHMS FOR COORDINATE TRANSFORMATIONS

Despite the fact that many of these areas show signs that auditory signals have been translated into an eye-centered reference frame, little is known about how the brain does this. Two models for how this may be accomplished have been proposed by Groh and Sparks.

A. Computation

Transforming signals from one frame of reference into another is mathematically straightforward. If the location of an auditory target with respect to the head (A_h) is known, and the position of the eyes with respect to the head (E_h) is also known, then subtracting the eye position signal E_h from the head-centered target location signal A_h will yield a signal of the location of the sound with respect to the eyes (A_e) (Fig. 6A). The first model, the vector subtraction model (Fig. 6B), implements an algorithm resembling this mathematical calculation using neurons as the computational elements. The second model, the dendrite model (Fig. 6C), uses a different algorithm: a neural analog of a multidimensional lookup table, which is implemented through local circuitry at the dendrites of the units in the eye-centered auditory map.

B. Input and Output Signals

As far as possible, the component signals are provided by units that resemble known populations of neurons. Eye position signals are carried by neurons from a variety of brain areas, such as the brain stem gaze centers, nucleus prepositus hypoglossi, the flocculus of the cerebellum, and posterior parietal cortex. Each of these areas contains neurons whose firing rate is typically linearly related to eye position. This information is most likely derived from a copy of the motor command to move the eyes, although proprioceptive feedback from the extraocular muscle spindles is also possible. No evidence of a place code of eye position, with neurons having response fields for eye position, has been reported. Thus, the eye position units in the model signal the position of the eyes in the orbits in a linear or rate-coded fashion.

For lack of any evidence to the contrary, the representation of the inputs to the model is assumed to be an auditory map of space in a head-centered frame of reference. Units in this map have receptive fields defined with respect to the head. However, convincing evidence for an auditory map of space in a head-centered frame of reference has not been reported—the frame of reference has either not been studied (classical auditory areas such as IC and auditory cortex) or the results have indicated a possible eye-centered frame of reference (SC, FEF, and LIP). Thus, how the inputs to this transformation are actually encoded is an open question.

The output of both of the models is an eye-centered auditory map resembling the primate SC. These units have receptive fields in an eye-centered frame of reference.

C. Neural Architecture

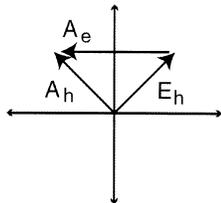
The circuitry of the vector subtraction model is illustrated in Fig. 6B. Neurons in the head-centered auditory map project to a set of intermediate units with a synaptic weighting that is proportional to their receptive field location. Thus, the level of excitation received by these intermediate units will be a linear rate code of the location of the sound with respect to the head. These units also receive an inhibitory synapse from a unit encoding eye position using a linear rate code. This inhibitory connection is the mechanism by which the eye position signal is subtracted from the signal of head-centered target location. The resultant signal encodes the location of the auditory target with respect to the eyes, again as a linear rate code.

The remainder of the circuitry in the model is devoted to converting this rate code into a place code in which neurons have receptive fields in an eye-centered frame of reference, like the map of oculocentric auditory space in the SC. This is accomplished using graded thresholds and inhibitory interneurons.

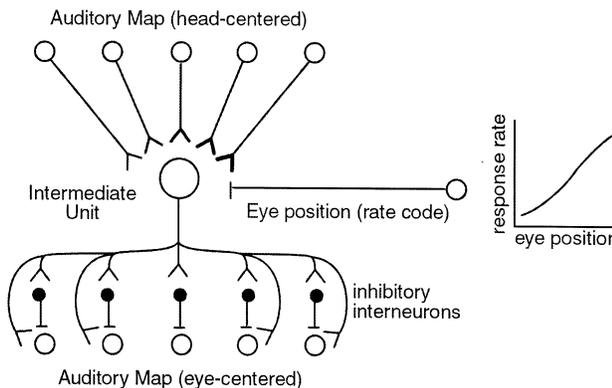
The simplicity of this mechanism is not without cost: Because of the rate coding at the intermediate stage of processing, the model functions properly for single auditory targets only. In order for an animal to make accurate saccades to the location of one auditory target among distracters, a target selection mechanism must choose the target before the coordinate transformation is executed. If two targets were present, and both were selected as saccade targets, the model as originally proposed would yield the vector sum of two auditory targets as an output. A small modification to this model would allow it to compute the vector average instead of the sum. Vector averaging is a more realistic response to multiple inputs and has been found in a number of other behavioral and physiological experiments. We therefore consider vector averaging of two auditory targets to be more likely than vector summation. Nevertheless, the vector subtraction model is simply not capable of preserving both target locations independently.

The alternative model, the dendrite model (Fig. 6C), can handle multiple targets because there is no rate-coding bottleneck at an intermediate stage of processing. Instead, every unit in the input map projects

A Computation



B Vector Subtraction Model



C Dendrite Model

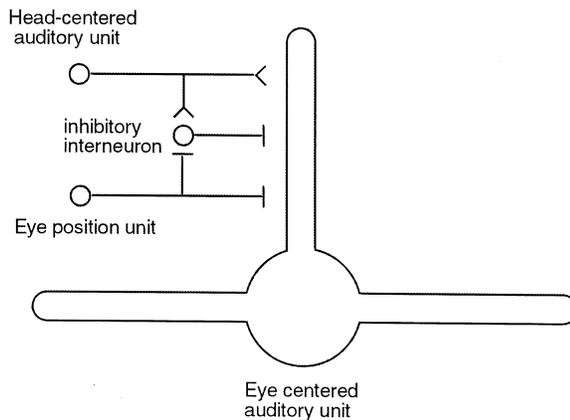


Figure 6 Computation of the eye-centered location of a sound. (A) If the auditory stimulus is located above and to the left of the head (A_h), and the eyes are directed upward and rightward in the orbit (E_h) as shown here, then the auditory stimulus is located directly to the left of the eyes (A_e). (B) The vector subtraction model. The head-centered auditory map projects to the intermediate unit with graded synaptic weights. The eye position signal is subtracted at an inhibitory synapse. The resulting rate code for the target's location with respect to the eyes is converted into a place code for auditory space in an eye-centered frame of reference through graded thresholds and inhibitory interneurons. (C) The dendrite model. Each dendrite receives input from eye position units and one head-centered auditory unit. Thresholds and synaptic weights are balanced so that the cell body receives net excitation if an auditory stimulus is present and the eyes are within a certain range of positions. Each dendrite performs this analysis independently for a particular location in head-centered space. *In toto* the dendrites of an individual neuron create an eye-centered receptive field [adapted with permission from Groh, J. M., and Sparks, D. L. (1992). Two models for transforming auditory signals from head-centered to eye-centered coordinates. *Biol. Cybernetics* 67, 291–302].

directly onto separate individual dendrites of every unit in the output map. The eye position units also project directly onto each dendrite. There are also interneurons associated with each dendrite. The local circuitry at each dendrite is set up to activate the unit if (i) there is a sound in the receptive field of the head-centered unit that projects to that dendrite and (ii) the position of the eyes places that sound within the eye-centered receptive field of that unit. Other sounds that are not in the eye-centered receptive field will not activate that particular unit, but they will activate other units in the eye-centered map via the projections to individual dendrites on those units. Because these connections are all independent, the model can accurately transform the locations of an unlimited number of targets.

These are two possible mechanisms whereby auditory signals can be translated from head- to eye-centered coordinates. Other algorithms are undoubtedly also possible. Furthermore, these algorithms are not limited to auditory coordinate transformations nor to translating information from head coordinates into eye coordinates. Translation of visual information from retinal to updated eye-centered coordinates or head-centered coordinates could be accomplished using similar mechanisms.

V. PARALLELS WITH OTHER MULTISENSORY MOTOR SYSTEMS

Thus far, we have concentrated primarily on brain areas implicated in multisensory processing and the control of saccades to sensory stimuli. However, recent work in premotor cortex has revealed that many of the same principles apply for the neural control of skeletal movements guided by sensory stimuli of different modalities. Neurons in premotor cortex can respond to visual or somatosensory stimuli or both. Auditory responses also exist but have been less extensively explored.

Michael Graziano, Charles Gross, and colleagues explored the issues of spatial registry/correspondence and the frames of reference of these signals. They found that bimodal visual/somatosensory neurons show an interesting spatial correspondence, with visual receptive fields tending to occupy a circumscribed three-dimensional volume of space in the immediate vicinity of the region of the body surface that contains the tactile receptive field. Furthermore, the frame of reference appears to be body-part centered. Neurons with tactile receptive fields on the

limb and visual receptive fields in nearby space employ a limb-centered frame of reference. The visual receptive fields move in space when the limb moves. If the eyes move but the limb does not, the visual receptive field remains in its original location in space, despite the fact that a different region of the retina would not be activated by stimuli at that location. Thus, the retinal location of a stimulus is irrelevant to these neurons; only the location of the stimulus with respect to the limb is important. Similarly, when the visual receptive fields are located near the face or head, they move when the head moves, even if gaze position in space is constant. In short, like the areas of the oculomotor pathway, the skeletal motor pathway appears to employ frames of reference that are tailored for guiding the movements of individual body parts, regardless of the modality of the stimulus that evokes the movement.

How these visual signals might be translated into a body-part-centered frame of reference is not known. Mechanisms similar to those outlined in the previous section for the translation of auditory signals from head- to eye-centered coordinates could apply in this case as well. For example, the translation of visual signals from eye- to head-centered coordinates is simply the inverse of the auditory transformation and could be accomplished using a retinotopic map as the input. The vector subtraction model would become the vector addition model, with the eye-in-head signal being added to the retinal vector to produce a signal of visual stimulus location with respect to the head. Similar modifications to the dendrite model could also be made. Translation of visual signals from a retinal- to a limb-centered frame of reference is obviously more complicated, but even this computation could be accomplished in an analogous fashion if the brain were to first compute the position of the limb with respect to the eyes. Replacing the eye-in-head signal with an eye-with-respect-to-limb signal in the models might allow the computation to proceed efficiently from that point, without requiring a series of intervening coordinate transformations. Of course, computing such a signal might not be easy, given how many joint angles would be involved.

VI. CONCLUSION

Our brains receive a constant barrage of sensory information from our different sense organs. Making sense of this onslaught requires determining which visual, auditory, somatosensory, olfactory, and gusta-

tory cues match up with one another and derive from a common event. The multisensory neurons in the brain are likely to play a critical role in this process. Our review of multisensory activity within the saccade pathway reveals evidence for common frames of reference and spatial registry and/or correspondence across sensory modalities. However, we still have much to learn about these cells. Detailed receptive field mapping and more extensive quantitative analysis of the responses to combined modality stimuli in awake animals are needed. Such studies will serve as the foundation on which a bridge between multisensory neural activity and behavior can ultimately be built.

See Also the Following Articles

Suggested Reading

- Andersen, R. A., and Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **3**, 532–548.
- Bisiach, E., and Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex* **14**, 129–133.
- Brotchie, P. R., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* **375**, 232–235.
- Graziano, M. S. A., and Gross, C. G. (1998). Spatial maps for the control of movement. *Current Opin. Neurobiol.* **8**, 195–201.
- Groh, J. M., and Sparks, D. L. (1992). Two models for transforming auditory signals from head-centered to eye-centered coordinates. *Biol. Cybern.* **67**, 291–302.
- Groh, J. M., and Sparks, D. L. (1996). Saccades to somatosensory targets. III. Eye-position-dependent somatosensory activity in primate superior colliculus. *J. Neurophysiol.* **75**, 439–453.
- Jay, M. F., and Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* **309**, 345–347.
- Mays, L. E., and Sparks, D. L. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *J. Neurophysiol.* **43**, 207–232.
- Mazzoni, P., Bracewell, R. M., Barash, S., and Andersen, R. A. (1996). Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *J. Neurophysiol.* **75**, 1233–1241.
- Meredith, M. A., and Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol.* **75**, 1843–1857.
- Russo, G. S., and Bruce, C. J. (1994). Frontal eye field activity preceding aurally guided saccades. *J. Neurophysiol.* **71**(3), 1250–1253.
- Snyder, L. H., Grieve, K. L., Brotchie, P., and Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature* **394**, 887–891.
- Stein, B. E., Wallace, M. T., and Stanford, T. R. (1999). Development of multisensory integration: Transforming sensory input into motor output. *Mental Retardation Dev. Disabilities Res. Rev.* **5**, 72–85.
- Stricanne, B., Andersen, R. A., and Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J. Neurophysiol.* **76**, 2071–2076.
- 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.