Three-dimensional structure-from-motion selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey

Kathleen C. Anderson[*] and **Ralph M. Siegel**
Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, New Jersey

Abstract

Human and non-human primates similarly are able to perceive three-dimensional structure from motion displays. Three-dimensional structure-from-motion (object-motion) displays were used to test the hypothesis that neurons in the anterior division of the superior temporal polysensory area (STPa) of monkeys can selectively respond to three-dimensional structure-from-motion. Monkeys performed a reaction time task that required the detection of a change in the fraction of structure in three-dimensional transparent sphere displays. Neurons were able to distinguish structured and unstructured three-dimensional optic flow. These cells could differentiate the change in structure-from-motion at stimulus presentation and when the animal was making a detection of the amount of structure in the display. Some of these neurons were also tuned for characteristics of the sphere stimuli. Also cells were tested with navigational motion and many were found to respond both to three-dimensional structure-from-motion as well as this spatial cue. These results suggest that STPa neurons represent specific aspects of three-dimensional surface structure and suggest that neurons within STPa contribute to the perception of three-dimensional structure-from-motion.

Keywords

structure-from-motion; optic flow; single-unit recording; monkey; visual pathways; temporal cortex

---

In 1909, Helmholtz demonstrated that *Homo sapien* can integrate motion information to create three-dimensional percepts (von Helmholtz, 1924). Psychophysicists defined the constructive nature of this task (Wallach et al., 1953) and computational studies have established many of the constraints underlying this perceptual ability (Hoffman and Bennett, 1986; Longuet-Higgins and Prazdny, 1980; Marr, 1982; Ullman, 1979). The ability to perceive structure-from-motion is remarkable as it involves the combination of information thought segregated into two separate visual streams (the ventral "what" and dorsal "where" pathways) (Mishkin et al., 1983).

A psychophysical task was developed to test whether monkeys had the same psychophysical ability to perceive structure-from-motion as humans. In that study, both species were tested with a structured, hollow three-dimensional sphere and a control "unstructured" stimulus (Siegel and Andersen, 1988), and indeed the two species were similar across a broad range of stimulus parameters indicating the monkey was a valid model for the human percept.

Address: correspondence to: Ralph M. Siegel, Ph.D. Center for Molecular and Behavioral, Neuroscience, Rutgers University, 197 University Avenue, Newark, NJ 07102, axon@cortex.rutgers.edu, voice: 973-353-1080 x3261, facsimile: 973-353-1272.
[*]**Present address**: Kathy C. Anderson, Ph.D. Cognitive Neuroscience Program, NIMH, 6001 Executive Boulevard, Room 7170, MSC 9637, Bethesda, Maryland 20892-9637

Neurons that are sensitive to the rotation of objects in depth have been reported in MT (Bradley et al., 1998), MST (Saito et al., 1986; Tanaka et al., 1986) and the inferior parietal cortex (Sakata et al., 1986; Sakata et al., 1994; Shikata et al., 1996). These neurons have been shown to be tuned to various aspects of three-dimensional structure-from-motion, and may form a first step in generating a representation of structure-from-motion; they have not been directly tested for their changes in the fraction-of-structure. There are a number of studies showing that inferior temporal neurons respond to shape (Gross et al., 1972; Kayaert et al., 2003), but these are not solely defined by motion. fMRI studies in human and monkey subjects have described areas that are selective for aspects of three-dimensional structure-from-motion (Orban et al., 1999; Sereno et al., 2002; Vanduffel et al., 2002); however these blood flow derivative studies cannot look at the temporal details of single neurons. Each of these areas may contain elements, or a complete representation, of three-dimensional structure-from-motion. Many of these converge upon to the temporal lobe.

A likely neural candidate in the temporal lobe based upon these results and connectional data is the superior temporal polysensory area (STP) (Anderson and Siegel, 1999) which lies in the upper bank and floor of the superior temporal sulcus. STPa is connected to both streams, in particular to regions rich in motion information (MST and 7a) and form information (TE) (Cusick et al., 1995). Indeed fMRI studies indicate areas that are either STP or near STP have blood flow that is dependent on the structure in a display. Selectivity for other complex global motion patterns (e.g., navigational optic flow) has been found in STPa using single unit methods (Anderson and Siegel, 1999; Bruce et al., 1981). STP has two portions; the anterior portion of the superior temporal polysensory area (STPa) receives input from both the dorsal (motion/spatial) and ventral (object) visual processing streams (Baizer et al., 1991; Boussaoud et al., 1990; Cusick *et al.*, 1995; Seltzer and Pandya, 1984). STPa neurons respond well to moving stimuli and are selective to different types of motion including biological motion (Bruce *et al.*, 1981; Oram et al., 1993; Wachsmuth et al., 1994).

Neurons were found in STPa that had the characteristics expected if these cells indeed represented the three-dimensional structure from motion. Indeed the change in neuronal activity of some of these neurons is directly correlated with the monkey's performance of the behavioral task. Portions of these results have appeared in abstract form (Anderson and Siegel, 1995, 1997).

METHODS

Neurons were recorded in STPa from three hemispheres of two male Rhesus monkeys that performed a reaction time task requiring the detection of three-dimensional structure-from-motion as compared to a control unstructured stimulus **(Methods)** (Ratzlaff and Siegel, 1990; Siegel and Andersen, 1988). These two stimuli were expressly constructed to have exactly the same local and global density of points, and the same distribution of local motion components. The only difference between the two displays was the fraction of structure (FOS) (Siegel and Andersen, 1988), which indicates the spatial organization of the motion components that define the three-dimensional shape (Longuet-Higgins and Prazdny, 1980). When the FOS was 1, all the motion trajectories were in the correct position, giving rise to a three-dimensional hollow sphere; a FOS=0 indicated that the motion trajectories were randomly shuffled yielding the "unstructured" display. Psychophysical studies have shown that monkeys and humans detect changes in these stimuli similarly (Siegel and Andersen, 1988) providing a foundation for exploring the properties of neurons in monkey in order to understand the neuronal basis of this perception in humans. Up to four different sphere displays were used to test each cell. The four spheres differed in diameter (10 and 20°) and axis of rotation (vertical and horizontal).

Behavior

Two male Rhesus monkeys (4–6 kg) were trained to perform a reaction time task while fixating a central 0.3° point as described elsewhere (Ratzlaff and Siegel, 1990). The monkey pulled a lever at the onset of the fixation point. Two seconds later, the visual stimulus came on. A change in the structure of the structure of the display occurred randomly between 3500 and 6000 msec after the fixation point onset. The monkey needed to release the key within a reaction time window of 150 to 800 msec for a juice reward. The monkey's head was fixed and one eye's position was monitored with an ISCAN infrared tracker to be within 1°. Saccades were not permitted. The monkeys correctly performed this task for 80–100% of the trials. Typically 8–12 trials were collected for each stimulus condition.

Visual stimuli

Spheres—The visual stimuli (Figure 1) all consisted of 128 points (0.1°) that had a limited lifetime of 532 msec (Morgan and Ward, 1980;Ratzlaff and Siegel, 1990;Siegel and Andersen, 1988). The spheres rotated at an angular velocity of 60°/sec around an axis that was in the plane of the display. Particular care was taken to ensure that the point density was kept constant to avoid density form cues for three-dimensional shape (Anderson and Siegel, 1999;Ratzlaff and Siegel, 1990). Receptive fields are typically over 40° in size (Anderson and Siegel, 1999) and include the fovea. Therefore, spheres were chosen to be well within the receptive field with a diameter of 10 or 20° of visual angle and centered on the fixation point (Figure 1a). An orthographic projection of the sphere resulted in fast motions in the centers and slower ones at the edges (Figure 1b). Rotation was along the vertical or horizontal axis. These manipulations of diameter and axis of rotation lead to the four different sphere *characteristics* used in this study. The displays were "unstructured" by randomly displacing entire motion trajectories in a square window whose width was equal to the diameter of the motion display (Figure 1c). In the unstructured display, the same distribution of motions was used, but the motion trajectories were randomly displaced. A completely structured display had a fraction-of-structure (FOS) of 1; an unstructured display had a FOS of 0.

Navigational optic flow—Neurons in STPa have been shown to be selective to navigational optic flow (Figure 1d-g). The question arises as to whether the STPa neurons selective to three-dimensional structure-from-motion also respond to the navigational optic flow. Thus cells were also studied with the navigational optic flow. These optic flow data are presented solely for comparison and are fully discussed in an earlier publication (Anderson and Siegel, 1999). The navigational displays had the same number of points, and the same point life. The points rotated at 60°/sec. Radial expansion and compression displays were constructed to have exactly the same speed distributions as the planar rotations by using the speed profiles from the planar rotation stimuli. The navigational stimuli were generally 40° in diameter.

Task difficulty—The stimuli were designed to have changes well above detectable thresholds. While the ease of the task ensured that the task difficulty was minimized across all stimulus conditions and that task difficulty could not account for differences in response, our experimental design precluded single trial error analysis as very few incorrect trials were observed.

Electrophysiology, anatomical methods, and statistical analysis

The monkeys were implanted with a cap made of Smith+Nephew Palacos orthopedic cement and Synthes screws with a pedestal to fix the head. A chamber was implanted so that penetrations could be made in the frontal plane. Single units were recorded using standard methods (Anderson and Siegel, 1999) with an interspike interval precision of 0.1 msec. Chamber placements and penetrations were guided using MRIs taken prior to the study and

confirmed using radiography of the electrodes *in situ.* Electrolytic lesions were made at the end of the recording and the recording sites were verified histologically (Anderson and Siegel, 1999).

Peri-stimulus time histograms were computed from the correct behavioral trials. Typically 8–12 trials were averaged and a 25 msec bin width was used. The responses of the neurons were quantified by measuring the firing rate for 500 msec before and after stimulus onset. When the change in firing rate at stimulus change was evaluated, the firing rate for 500 msec before the change was compared to the rate for the time following the change up to when the key was released, ~350 msec (Siegel and Read, 1997). The firing rate was expressed in Hz and the trial-by-trial data was subjected to a two-way analysis of variance as described elsewhere (Anderson and Siegel, 1999). A neuron was defined as *selective* if it significantly responded differentially to at least one of the stimuli within a given set of stimuli using a two-way analysis of variance ($P < 0.05$) (Anderson and Siegel, 1999; Siegel and Read, 1997). *Sensitive* cells had a significant effect of the stimulus onset ($P < 0.05$), but did not have a significant dependence on the type of stimulus.

The neurons described in this study were drawn from a population partly described previously (Anderson and Siegel, 1999). All procedures were approved by the Rutgers University Animal Institutional Review Board and were in accordance with the NIH Guidelines on the Care and Use of Animals in Research.

RESULTS

Of 464 visual neurons tested, 266 (57%) had a significant response to the onset of the structured sphere stimuli (Figure 2a), with 112 having responses that were selective for the size and/or axis of rotation. These significant responses could either be increases or decreases in firing rate. This initial onset response could be the result of the three-dimensional structure derived from motion or it could be due to the simplest qualities of the stimuli (e.g., the change in luminance at onset). The presentation of the unstructured control motion stimulus eliminates the latter explanation. The cell in Figure 2 did *not* have a significant response to the onset of unstructured motion display (Figure 2b) for which the motion components were exactly the same but the percept of a hollow sphere is lost; the only difference between the two displays was the spatial distribution of speeds which defined the sphere. The response to the structured motion and the dependence on the size of the display suggests the neuron was selective for some aspects of the three-dimensional structure of the display.

Response to sphere stimuli at stimulus change

A direct assessment of the neurons' response to changes in the FOS within each trial was obtained by evaluating the response at the time the stimulus changed from structured to unstructured motion or *vice versa* (Figure 3). At this time, not only is the animal attending to the stimulus, he is also in the process of detecting the change in the FOS of the stimulus. Cells were found that did not respond to the initial onset of the display, but did respond significantly when the FOS in the display changed (Figure 3a versus b). The increase in activity preceded key release (Figure 3c). This cell's response is reasonably difficult to explain on trivial grounds as its response cannot be attributed to simple characteristics of the stimuli such as directional selectivity, number of points, and change in luminance. The only characteristic of the display that changes is the spatial distribution of speeds across the display. When the stimulus is unstructured, there is a bounded random distribution; when the display is structured there are faster speeds in the center and slower ones at the edge.

Of 464 neurons tested with a reduction in the FOS (structured to unstructured), 70 (15%) responded significantly to the change. These responses could not be attributed to loss of

behavioral control as these changes preceded the animal's release of the key and juice reward (Figure 3c). Cells such as these responded to the *subtle* changes in the spatial distribution of speeds across the display. It is this precise event- the change in the structured motion trajectories across the display that define the difference between the structured hollow sphere and its unstructured control as shown by computational studies (Longuet-Higgins and Prazdny, 1980). Indeed, it is this same spatial ordering of motion speeds across the display that has been accepted as one definition of the psychophysical ability to perceive structure-from-motion (Logothetis et al., 1995; Siegel and Andersen, 1988; Vaina, 1994). Other neurons were found that responded to the transition from the unstructured control display to the structured hollow sphere (21% of 92 cells) with the majority of these having increases of firing rate.

Coincident with the change in fraction of structure, the monkey is generating motor planning signals for the ensuing key release. However motor planning alone cannot explain the change in neuronal firing rate, because cells were also selective to the stimulus characteristics as demonstrated by a dependence on the four types of different displays (two diameters by two axis of rotation). Of the 70 cells that showed a significant response to a decrease in structure-from-motion, 49% were selective for the stimulus characteristic of size or axis of rotation; of the 18 cells firing for an increase in fraction of structure, 53% showed selectivity for the stimulus characteristics of size and rotation axis. If the cells were only encoding the motor planning signals, then the responses would not depend on the stimulus characteristics. These data lead to the conclusion that the STPa neurons respond to the change in fraction-of-structure in a manner expected based upon psychophysical studies (Siegel and Andersen, 1988). Further, the temporal correlation between the neural activities associated with the transition in the fraction of structure that occurs prior to the behavioral response indicates that the neural activity of these cells can play a role in the perception of structure-from-motion.

One possible description of these cells is that they are only encoding transitions in the fraction-of-structure. However this does not appear to be the case for two reasons. First, many of these cells are tuned to the characteristics of the motion stimuli as shown by the significant effect of the size/orientation of the sphere stimuli. A second analysis can address this point directly. Eighty-one of the total cells tested were examined with both the "unstructured to structured" transition and the "structured to unstructured" transition using the analysis of variance described above. Of this group of neurons, 50 (62%) did not respond significantly to either transition, 14 (17%) responded significantly only to a decrease in the fraction-of-structure, and 13 responded significantly only to the increase in the fraction-of-structure. Only 4 neurons responded to both experimental runs indicating a general sensitivity to the transition in structure-from-motion in both directions; of these only 2 (2.5%) were similarly tuned to the size and direction characteristics of the optic flow. The gross majority of neurons (33% vs. 2.5%) were tuned to either an increase or decrease in the fraction-of-structure indicating that these neurons were not tuned to solely indicate transitions in structure-from-motion regardless of the underlying stimulus characteristics.

Response to sphere stimuli at onset

As an additional control for whether the cells could differentiate between structured and unstructured three-dimensional structure, comparisons were made across two experimental runs for individual cells. For example, the responses to structured motion onsets and unstructured motion onsets were compared. The hypothesis of a common component of the response attributable to the simple increase in luminance at onset could be tested by the analysis of variance. Similarly, if the comparison were made when the stimulus structure changed, the common premotor components would be discernible. As the reaction time task had a randomized variable delay to the change in fraction-of-structure, there were only two visual

events in the task across two experimental runs that could matched—stimulus onset and stimulus change.

The response to the onset of the structured motion displays was directly compared with the response to the onset of the "unstructured" motion control displays. The comparison was performed using a two-way ANOVA with one independent variable corresponding to the FOS of the display and the other to the characteristics of the sphere (two stimulus sizes by two axes of rotation). Thus the response of a cell with a significant effect of the FOS cannot be explained by simple effect of luminance or the presence of motion. This analysis was performed for the 90 cells that were tested with both structured and unstructured motion displays. (The two cells that were only tested with unstructured to structured motion were not included in this analysis; the 374 cells that were only tested with the structured to unstructured motion were also excluded.)

A comparison of the response at onset in Figure 2a and b indicates that this cell differentially responded to the onset of structured motion versus the onset of unstructured motion. Thirty-three cells of the 90 tested (37%) were able to distinguish the structured from the unstructured motion at stimulus onset (Figure 4a, $P<0.05$ USM vs. SM). These cells appear to be extracting the three-dimensional structure from the motion display.

One-quarter of these putative structure-from-motion cells were able to provide information about the characteristics of the stimuli in that the diameter and/or axis of rotation had a significant effect on firing rate (Figure 3a; $P< 0.05$ for USM vs. SM, tuned). The other 57 cells of the 90 tested (63%) were unable to distinguish unstructured from structured motion at onset ($P>0.05$ for USM vs. SM) at stimulus onset. The responses of this latter group of cells are most simply explained as a response to lower-order characteristics of the stimuli (e.g., increase in luminance or the presence of translation motion). Thus either the effect of the luminance or the translation motion may dominate the sensitivity to fraction-of-structure. Alternatively, there is no real selectivity for structure-for-motion initially.

The premotor signals do not appear to play a substantial role later in the trial, when the animal is detecting the change in the fraction-of-structure. At that time, 44 of the 90 cells (49%) had a significant response to the change in the structured and unstructured displays (Figure 4b, $P<0.05$ SM vs. USM). Half of these FOS selective cells were also encoding information about the characteristics of the sphere (Figure 3b, $P<0.05$ SM vs. USM, tuned).

Both of these population measures indicated an increase in the number of selective responses at the time the stimulus changed. A larger percentage of the 90 neurons tested were selective for the SM vs. USM comparison (44 vs. 33 cells; change vs. onset) and more were tuned to the stimulus characteristics at the change (22 vs. 9 cells). This change in selectivity as the task progressed that the attentional state of the animal could alter the structure-from-motion selectivity of the neurons as has been demonstrated for the inferior parietal lobule (Phinney and Siegel, 2000; Siegel and Read, 1997).

Thus we have shown that neurons respond selectivity to fraction-of-structure for a rotating sphere. This selectivity to the fraction-of-structure suggests that these cells are involved in the perceptual ability to extract three-dimensional structure-from-motion. An alternative explanation for the response of these neurons is that they simply respond to the gradient of speeds across the display. For example the gradient of speeds for the rotating sphere is faster speeds in the middle and very slow speeds at the edges. In contrast, the gradient of speeds for an unstructured sphere is a random gradient of speeds with at each position with the range of speeds limited to that contained within a structured motion display. However, it is precisely this gradient that defines the three-dimensional depth.

Dependence on size and axis of rotation—The effects of size and axis of rotation on the response of these putative three-dimensional structure-from-motion cells were evaluated with the Bonferroni *posthoc* tests ($P< 0.05$). Half of the putative three-dimensional structure-from-motion neurons had no selectivity to the characteristics of the sphere. In one sense these could be considered cells that solely represent "sphereness" from motion. The other half (22/46) of these putative three-dimensional cells showed responses that were modulated by characteristics of the sphere stimuli. Eleven neurons were modulated by the axis of rotation of the spheres (n=11), some by the size (n=5), while others were affected by both size and axis of rotation (n=6). These representation of both types of neurons within STPa suggests that there could be a hierarchical arrangement for the processing of this higher order motion components, where the size and orientation dependent cells are combined to form cells independent of these characteristics.

Comparison of response to spheres and navigational optic flow

STPa neurons are also known to be selective for optic flow patterns derived from egocentric motion (Anderson and Siegel, 1999). Ninety neurons were tested with both the onset of SM and USM to determine which responded exclusively to three-dimensional structure-from-motion or were more broadly tuned for navigational optic flow (Anderson and Siegel, 1999). Radial and rotating optic flow fields were used to test for navigation flow selectivity (Figure 5). This cell responded only to the three-dimensional stimulus and not to the optic flow stimulus. Cells were considered to respond to optic flow if they had a significant response relative to baseline at stimulus onset. Forty-three percent of the 90 cells tested were found to respond to optic flow alone and, as described earlier (Anderson and Siegel, 1999), showed a preponderance of selectivity for flow derived from forward egomotion (radial expansion). At stimulus onset, 37% of the 90 neurons were selective to the FOS in the three-dimensional motion displays. A substantial proportion (83%) of these three-dimensional selective cells was also selective to the onset of optic flow. Although there were a greater percentage of cells selective to the three-dimensional FOS when the stimulus changed (49%), the percentage of these cells that were also selective to the navigational optic flow at stimulus change remained about the same at 80%. Thus there seem to be two major classes of neurons within STPa— cells that respond to navigationally based optic flow exclusively and cells that combine this tuning with three-dimensional structure-from-motion selectivity. Neurons that respond exclusively to three-dimensional structure-from-motion are less common (under 10% of the cells). It is possible that these two populations could be correlated with two subregions of STPa is known from anatomical criteria (Cusick *et al.*, 1995), however it was not possible to determine if there was any spatial segregation of these neurons due to the long distance penetrations and chronic nature of these recordings.

DISCUSSION

A definition of structure-from-motion

Structured and unstructured motions displays were compared as the criteria to define whether neurons were tuned to represent three-dimensional structure-from-motion. This criteria is based upon the acceptance of the psychophysical approach of comparing structured and unstructured motion as a measure of perception of structure-from-motion (Siegel and Andersen, 1988, 1990). Not only has this approach been used to examine the responses of parietal neurons (Siegel and Read, 1997), it has been the foundation of fMRI studies in monkeys and humans examining of the expression of structure-from-motion selectivity across multiple visual areas (Sereno *et al.*, 2002). The general idea is quite similar to the approach of using varying coherence in translation stimuli to alter the monkey's perception of motion (Britten et al., 1993); a controlled change in the stimulus is selected to directly test an area's involvement in the perceptual process. The other guidance in the selection of these criteria is the clarification

and definition by computational neuroscience as to what components of structure-from-motion are crucial to computing three-dimensional structure-from-motion (Hoffman and Bennett, 1986; Marr, 1982; Ullman, 1979).

We next consider the characteristics of an "ideal" three-dimensional structure-from-motion neuron. At one extreme would be a cell that only responded if an object had a three-dimensional shape defined by motion. This cell would, in principle, not respond to the object's identity, its size, its color, or any other characteristic. One might even say it was a "grandmother" cell for a specific visual perceptual property (Barlow, 1972; Konorski, 1967). However, given the advances in our understanding of visual representations in cortex, it is highly unlikely that such a cell could exist, nor would it be possible to exhaustively test it (Van Essen, 1985). The more realistic point of view, both conceptually and pragmatically, is to expect that experiments to demonstrate that a putative three-dimensional structure-from-motion neuron has many of the appropriate characteristics, as demonstrated by observing the responses to a series of stimuli grounded in psychophysical experimentation. That is the approach taken here.

It is also realistic to expect that such a structure-from-motion neuron could be sensitive to other characteristics. One need only consider that individual MT/V5 neurons, perhaps the prototype of an exquisitely tuned neuron (i.e. to motion) also have selectivity to many other related visual perceptual attributes. They are selective to disparity, wavelength, and context (Albright and Stoner, 2002). Hence it was surprising to find an "ideal" neuron that only responded to the three-dimensional structure defined by motion independent of other visual (and perhaps non-visual) events. Other untested stimulus characteristics may modulate these cells. In summary, a putative structure-from-motion thus would be expected to have a response that distinguished between a two different levels of fraction-of-structure and have activity temporally correlated with the perceptual event of detecting changes in fraction-of-structure.

A substantial proportion of neurons were found in STPa that passed these stringent standards. These neurons are exquisitely sensitive to the speed gradients across the receptive field, which is precisely what defines three-dimensional structure-from-motion selectivity (Longuet-Higgins and Prazdny, 1980; Siegel and Andersen, 1988). Thus it is concluded that STPa neurons contain a representation of three-dimensions constructed from two-dimensional motion information. Whether or not these cells represent different shapes remains to be seen; testing with stimuli that have matched two-dimensional contours and different three-dimensional motion shapes could address this issue (Phinney and Siegel, 1999).

Possible sources of three-dimensional structure-from-motion selectivity in STPa

The analysis of the three-dimensional structure-from-motion may arise in STPa or may be carried from other cortical regions. Area MT has neurons that appear to differentiate the front and back of transparent objects as a function of the monkey's interpretation of the visual image (Bradley *et al.*, 1998). These cells could form a first step in the analysis of three-dimensional structure-from-motion, however given the small receptive field size of MT neurons; it is not clear how they could directly analyze the difference between planar and three-dimensional as seen in STPa neurons. The contextual interactions from beyond the classical receptive field (Albright and Stoner, 2002; Allman et al., 1985) may play a role in the initial processing of three-dimensional structure-from-motion. However to date MT neurons have not been demonstrated to distinguish structured and unstructured three-dimensional motion. MST contains groups of neurons sensitive to both optic flow (MSTd) and the movement of objects in depth (MSTl) (Desimone and Ungerleider, 1986; Duffy and Wurtz, 1991; Graziano et al., 1994; Newsome et al., 1988; Orban et al., 1992; Saito *et al.*, 1986; Tanaka and Saito, 1989; Tanaka et al., 1993) making it a putative source of three-dimensional structure-from-motion selectivity. Other regions that project to STP, such as the caudal intraparietal sulcus (Shikata *et al.*, 1996), or area 7a (Sakata *et al.*, 1994), could be part of the source if the structure-from-

motion selectivity. This possibility is difficult to assess since the three-dimensional structured and unstructured displays have not been tested in these areas with single unit recordings. The fMRI studies in monkey suggest a constellation of areas involved with the analysis of three-dimensional structure-from-motion, including STPa.

The alternative is that the selectivity to the fraction-of-structure of the three-dimensional sphere could arise by the convergence of signals from individual neurons sensitive to direction, transparency, and/or motion in depth, perhaps as found in MT or MST. These signals could then be combined locally to give rise to the representation of the three-dimensional structure of an object rotating in depth. A similar mechanism has been proposed for the selectivity of STPa cells to biological motion (Oram *et al.*, 1993; Wachsmuth *et al.*, 1994). In addition, there are static representations of form in nearby regions (TE in the lower bank of STS and IT) (Janssen et al., 2001; Kayaert *et al.*, 2003), which could contribute to the formation of the representation of structure-from-motion.

Invariance properties in STPa

Over half of the cells found to be selective for the three-dimensional motion displays also encoded the size and/or axis-of-rotation of the sphere displays, while the others were size and/or axis-of-rotation invariant. Size-invariance has been shown for shape-selective neurons in the inferotemporal (IT) cortex although changes in the size of the preferred shape can alter the absolute firing rate of IT neurons (Schwartz et al., 1983). The effects of size on the response of neurons in STPa may be similar, in that size affects the strength of the response, but not the overall selectivity of the neuron. In addition, the effects of the axis-of-rotation (or orientation) of the spheres on the firing rate of STPa neurons are similar to view-dependent IT neurons (Logothetis *et al.*, 1995). The finding that other STPa neurons respond equivalently regardless of the orientation of the sphere suggests that these latter responses are independent of the viewpoint of the observer relative to the object. Similar view-independent properties have been shown for face and object selective neurons in IT (Desimone et al., 1984; Perrett et al., 1985). STPa is connected with IT cortex via the fundus and dorsal bank of the STS (Cusick *et al.*, 1995; Jones and Powell, 1970; Morel and Bullier, 1990). Therefore signals from IT may contribute to the ability of STPa neurons to extract viewer-independent structural information from motion stimuli. Furthermore, STPa may be combining inputs from motion areas in STS with those from IT in order to encode the structure of moving objects. The finding of a more equal distribution of invariant and non-invariant responses to the size and orientation of a motion-defined object in STPa is consistent with its more numerous and direct connections with areas in the dorsal stream than with those in the ventral stream.

Utility of STPa neurons for behavior

Many cells in STPa responded to changes in the structure-from-motion stimuli by increasing or decreasing their firing when the stimuli became unstructured. STPa is interconnected with inferior parietal areas thought to be involved in the localization of stimuli for intended movements, including reaching, grasping, and object manipulation as well as sensorimotor transformations (Andersen et al., 1990; Baizer *et al.*, 1991; Goodale and Milner, 1992). The three-dimensional structure-from-motion selective neurons often also represent navigational optic flow indicating that STPa plays a role in encoding three-dimensional object movement in the environment relative to an observer (Zemel and Sejnowski, 1998). The confluence of these properties at this particular apex of the "what" and "where" pathways support the hypothesis that STPa plays a crucial function in the integration of spatial and form information and its transfer onto motor planning regions to guide or plan grasping and other reaching movements to moving objects in the environment.

References

Albright TD, Stoner GR. Contextual influences on visual processing. Annu Rev Neurosci 2002;25:339–379. [PubMed: 12052913]

Allman J, Miezen F, McGuinness E. Stimulus specific responses from beyond the classical receptive field: neurophysiological. AnnRevNeurosci 1985;8:407–430.

Andersen RA, Asanuma C, Essick GK, Siegel RM. Cortico-cortical connections of anatomically and physiologically defined subdivisions within inferior parietal lobule. JCompNeurol 1990;232:443–455.

Anderson KC, Siegel RM. Neuronal response to optic flow patterns in STP$_a$ in the behaving macaque. AbstrSocNeurosci 1995;21:664.

Anderson KC, Siegel RM. Neuronal response to 3D structure from motion (SFM) in the anterior superior temporal polysensory area (STPa) in a behaving monkey. InvestOphthalmolVisSci 1997;38:625.

Anderson KC, Siegel RM. Optic flow selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. JNeurosci 1999;19:2681–2692. [PubMed: 10087081]

Baizer JS, Ungerleider LG, Desimone R. Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J Neurosci 1991;11:168–190. [PubMed: 1702462]

Barlow HB. Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1972;1:371–394. [PubMed: 4377168]

Boussaoud D, Ungerleider LG, Desimone R. Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. JCompNeurol 1990;296:462–495.

Bradley DC, Chang GC, Andersen RA. Encoding of three-dimensional structure-from-motion by primate area MT neurons. Nature 1998;392:714–717. [PubMed: 9565031]

Britten KH, Shadlen MN, Newsome WT, Movshon JA. Responses of neurons in macaque MT to stochastic motion signals. Visual Neuroscience 1993;10:1157–1169. [PubMed: 8257671]

Bruce CJ, Desimone R, Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. JNeurophys 1981;46:369–384.

Cusick CG, Seltzer B, Cola M, Griggs E. Chemoarchitectonic and corticocortical terminations within the superior temporal sulcus of the Rhesus monkey: evidence for subdivisions of superior temporal polysensory cortex. JCompNeurol 1995;360:513–535.

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. JNeurosci 1984;4:2051–2062. [PubMed: 6470767]

Desimone R, Ungerleider LG. Multiple visual areas in the caudal superior temporal sulcus of the macaque. JCompNeurol 1986;248:164–189.

Duffy CJ, Wurtz RH. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. J Neurophysiol 1991;65:1345–1345.

Goodale MA, Milner AD. Seperate visual pathways for perception and action. TINS 1992;15:20–25. [PubMed: 1374953]

Graziano M, Andersen RA, Snowden RJ. Tuning of MST to spiral motions. JNeurosci 1994;14:56–67.

Gross CG, Rocha-Miranda CE, Bender DB. Visual properties of neurons in inferotemporal cortex of the macaque. JNeurophysiol 1972;35:96–111. [PubMed: 4621506]

Hoffman DD, Bennett BM. The computation of structure from fixed-axis motion: Rigid structures. BiolCybern 1986;54:71–83.

Janssen P, Vogels R, Liu Y, Orban GA. Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. J Neurosci 2001;21:9419–9429. [PubMed: 11717375]

Jones EG, Powell TPS. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 1970;93:793–829. [PubMed: 4992433]

Kayaert G, Biederman I, Vogels R. Shape tuning in macaque inferior temporal cortex. J Neurosci 2003;23:3016–3027. [PubMed: 12684489]

Konorski, J. Integrative activity of the brain; an interdisciplinary approach. Chicago: University of Chicago Press; 1967.

Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. Current Biology 1995;5:552–563. [PubMed: 7583105]

Longuet-Higgins HC, Prazdny K. The interpretation of a moving retinal image. ProcRoySocLondB 1980;208:385–397.

Marr, D. Vision. San Francisco: W.H. Freeman and Co; 1982.

Mishkin M, Ungerleider LG, Macko KA. Object vision and spatial vision: two cortical pathways. TINS 1983;6:414–417.

Morel A, Bullier J. Anatomical segregation of two cortical visual pathways in the macaque monkey. VisNeurosci 1990;4:555–578.

Morgan MJ, Ward R. Conditions for motion flow in dynamic visual noise. VisionRes 1980;20:431–435.

Newsome WT, Britten KH, Movshon JA. Comparison of neural and behavioral sensitivity to visual motion in alert monkeys. Soc Neurosci Abstr 1988;14

Oram MW, Perrett DI, Hietanen JK. Directional tuning of motion-sensitive cells in the anterior superior temporal polysensory area of the macaque. Experimental Brain Research 1993;97:274–294.

Orban GA, Lagae L, Verri A. First order analysis of optical flow in monkey brain. PNAS 1992;89:2595–2599. [PubMed: 1557363]

Orban GA, Sunaert S, Todd JT, Van Hecke P, Marchal G. Human cortical regions involved in extracting depth from motion. Neuron 1999;24:929–940. [PubMed: 10624956]

Perrett DI, Smith PA, Mistlin AJ, Chitty AJ, Head AS, Potter DD, Broennimann R, Milner AD, Jeeves MA. Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: a preliminary report. BehavBrainRes 1985;16:153–170.

Phinney RE, Siegel RM. Stored representations of three-dimensional objects in the absence of two-dimensional cues. Perception 1999;28:725–737. [PubMed: 10664767]

Phinney RE, Siegel RM. Speed selectivity for optic flow in area 7a of the behaving macaque. Cereb Cortex 2000;10:413–421. [PubMed: 10769251]

Ratzlaff EG, Siegel RM. A workstation interface for measuring spike intervals. JNeurosciMethods 1990;35:195–201.

Saito H, Yukie M, Tanaka K, Hikosaka K, Fukada Y, Iwai E. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. JNeurosci 1986;6:145–157. [PubMed: 3944616]

Sakata H, Shibutani H, Ito Y, Tsurugai K. Parietal cortical neurons responding to rotary movement of visual stimulus in space. ExpBrainRes 1986;61:658–663.

Sakata H, Shibutani H, Ito Y, Tsurugai K, Mine S, Kusunoki M. Functional properties of rotation-sensitive neurons in the posterior parietal association cortex of the monkey. ExpBrain Res 1994;101:183–202.

Schwartz EL, Desimone R, Albright TD, Gross CG. Shape recognition and inferior temporal neurons. ProcNatlAcadSciUSA 1983;80:5776–5778.

Seltzer B, Pandya DN. Further observations on parieto-temporal connections in the rhesus monkey. ExpBrainRes 1984;55:301–312.

Sereno ME, Trinath T, Augath M, Logothetis NK. Three-dimensional shape representation in monkey cortex. Neuron 2002;33:635–652. [PubMed: 11856536]

Shikata E, Tanaka Y, Nakamura H, Taira M, Sakata H. Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. Neuroreport 1996;7:2389–2394. [PubMed: 8951858]

Siegel RM, Andersen RA. Perception of three-dimensional structure from two-dimensional motion in monkey and man. Nature (Lond) 1988;3319:259–261. [PubMed: 3336437]

Siegel RM, Andersen RA. The perception of structure from motion in monkey and man. JCognitive Neurosci 1990;2:306–319.

Siegel RM, Read HL. Analysis of optic flow in the monkey parietal area 7a. Cerebral Cortex 1997;7:327–346. [PubMed: 9177764]

Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y, Iwai E. Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. JNeurosci 1986;6:134–144. [PubMed: 3944614]

Tanaka K, Saito H. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. JNeurophysiol 1989;62:626–641. [PubMed: 2769351]

Tanaka K, Sugita Y, Moriya M, Saito H. Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. JNeurophysiol 1993;69:128–142. [PubMed: 8433128]

Ullman, S. The interpretation of visual motion. Cambridge: MIT Press; 1979.

Vaina LM. Functional segregation of color and motion processing in the human visual cortex: clinical evidence. Cerebral Cortex 1994;5:555–572. [PubMed: 7833656]

Van Essen, DC. Functional organization of primate visual cortex. In: Peters, A.; Jones, EG., editors. Cerebral Cortex. NY: Plenum Publishing; 1985. p. 259-329.

Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, Todd JT, Orban GA. Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science 2002;298:413–415. [PubMed: 12376701]

von Helmholtz, LF. Helmholtz's treatise on physiological optics, translated from the third German edition. New York: Dover Publications; 1924.

Wachsmuth E, Oram MW, Perrett DI. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. Cerebral Cortex 1994;4:509–522. [PubMed: 7833652]

Wallach H, OConnell DN, Neisser U. The memory effect of visual perception of three-dimensional form. JExpPsychol 1953;45:360–368.

Zemel RS, Sejnowski TJ. A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. JNeurosci 1998;18:531–547. [PubMed: 9412529]
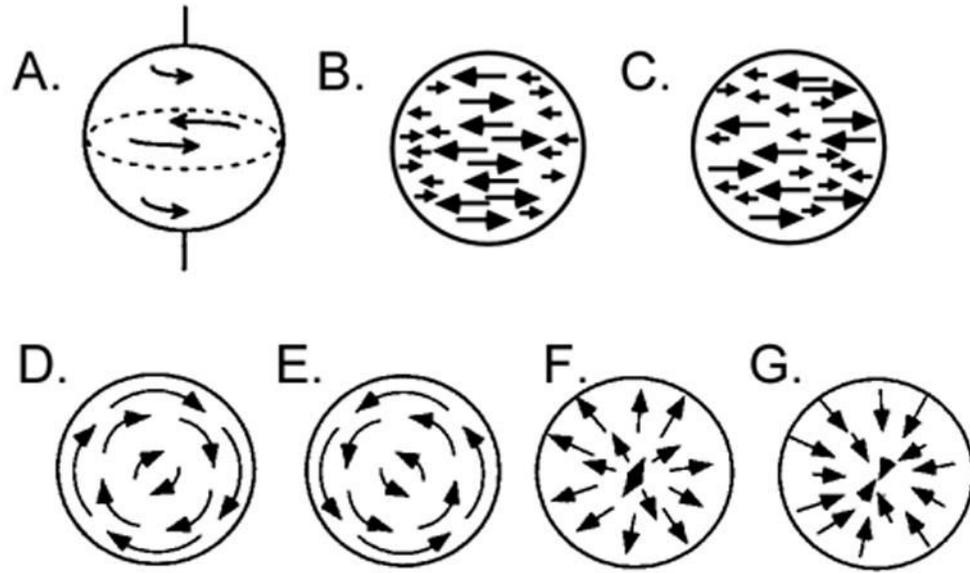
Figure 1.
Stimulus displays. A. Schematic of three-dimensional sphere with vertical axis of rotation. B. Schematic of speed gradient for structured sphere. C. Schematic of speed gradient for unstructured sphere.D–F. Four navigational optic flow displays. D. Clockwise rotation. E. Counterclockwise rotation. F. Radial expansion. G. Radial compression. All displays were matched for the number of points and point life.
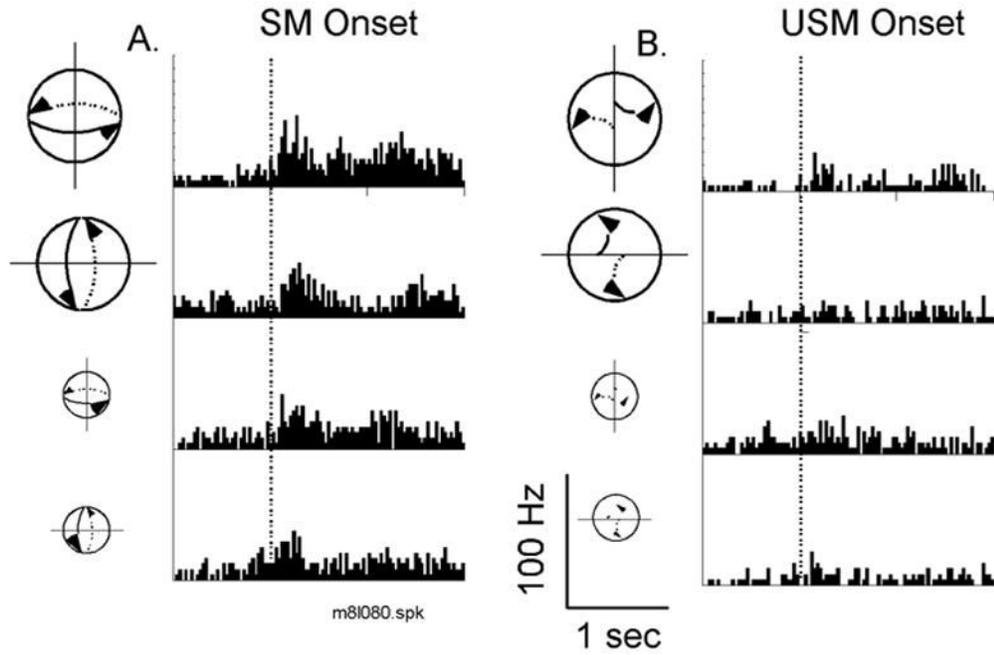
Figure 2.
Differential response of an STPa neuron to the fraction-of-structure (FOS) at stimulus onset (dotted line). A) Response of the cell to structured motion onset (SM Onset, FOS=1). Under these conditions, each stimulus is a field of evenly distributed dots with the speed at each location defined by its three-dimensional position. This cell responded strongly to the onset of the structured motion stimuli (P<0.05). Furthermore, the response significantly varied depending on the size and axis of rotation of the structured sphere displays. B) When the displays were unstructured and all the motion trajectories are randomly shuffled, the percept of a hollow sphere is lost. There was no significant response of this neuron to the onset of the unstructured displays. The distribution of speeds is the same as for the structured spheres. Eight repetitions of each condition were presented.
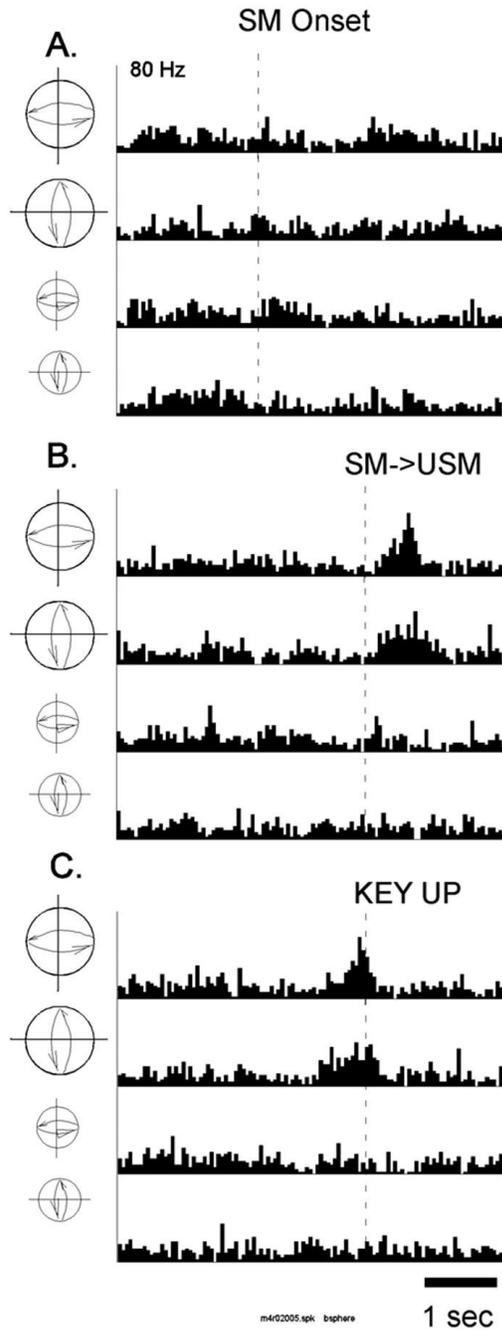
Figure 3.
Effect of decreasing the fraction of structure on the response of cells. A) This cell had no response at stimulus onset to the three-dimensional structured motion displays, regardless of the axis of rotation or the size of the display. B) When the data were synchronized to the time that the stimulus changed (dotted line), there was an increase in activity that corresponded to the loss in structure. Greater activity was shown in response to the larger spheres. C) In order to determine if the response followed the key release, the same data were synchronized to the key release for each trial. The change in activity preceded the key release (dotted line). Statistical analysis of these data only used trials for which there was behavioral control (i.e.,

the period following the change in the stimulus and preceding the key release on correct trials only). Eight repetitions of each stimulus condition were given.
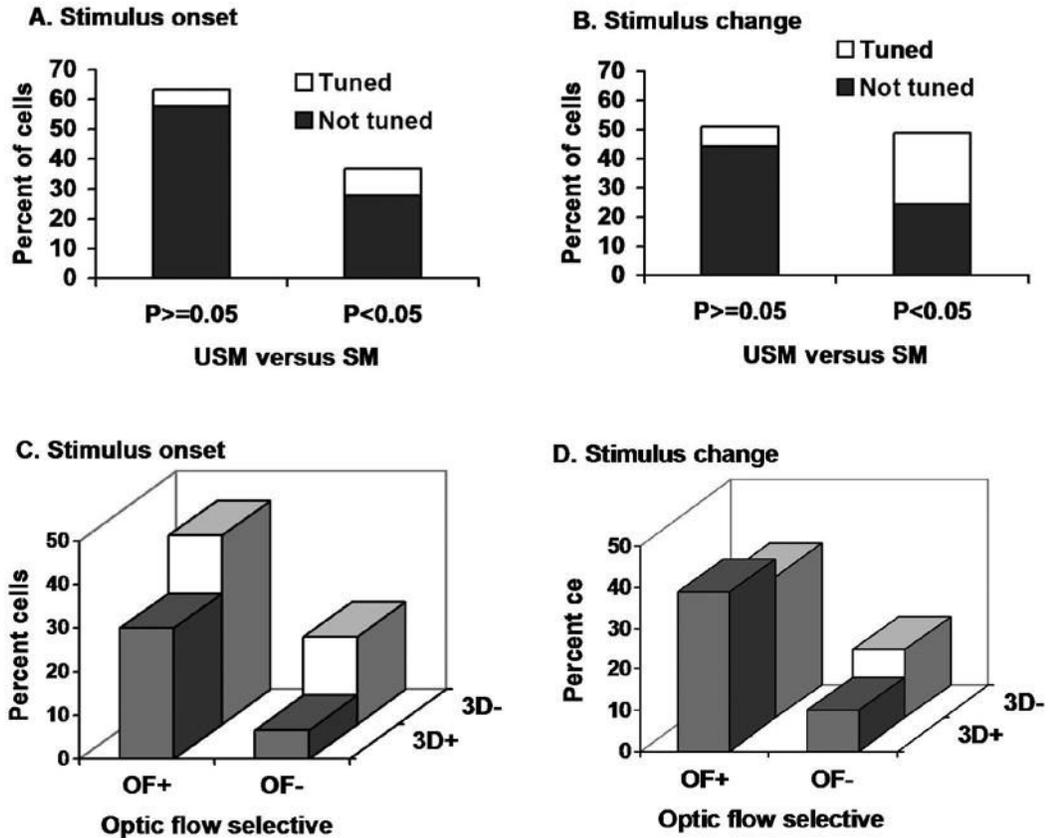
Figure 4.
Population distributions of STPa neurons' response to the fraction-of-structure for 90 cells tested with both the displays that began with structured motion (FOS=1) and those that began with unstructured motion (FOS=0). A) 37% of the cells tested were able to distinguish the structured from the unstructured motion at stimulus onset (P<0.05). One-quarter of these were able to provide significant information (P<0.05) about the characteristics of the stimuli (tuned). Of the cells that were not able to distinguish unstructured from structured motion (P>0.05), only 9 were able to indicate any information about the sphere characteristics (tuned). B) Responses at stimulus change. The proportion of cells tuned for the fraction of structure at change has increased relative to stimulus onset to almost 50%. The percentage of these cells providing information about the characteristics of the sphere increased as well. C) Distribution of the responses of the 90 cells tested with both the three-dimensional fraction-of-structure and the optic flow displays. Fraction-of-structure selective neurons often responded to two-dimensional optic flow. Neurons that responded only to three-dimensional stimuli were less common. D) Response for three-dimensional FOS at stimulus change as compared to the response to optic flow at stimulus change. OF+ means cells had a P<0.05 for optic flow stimuli, OF- means P>=0.05, 3D+ means P<0.05 for three-dimensional structure-from-motion, etc.
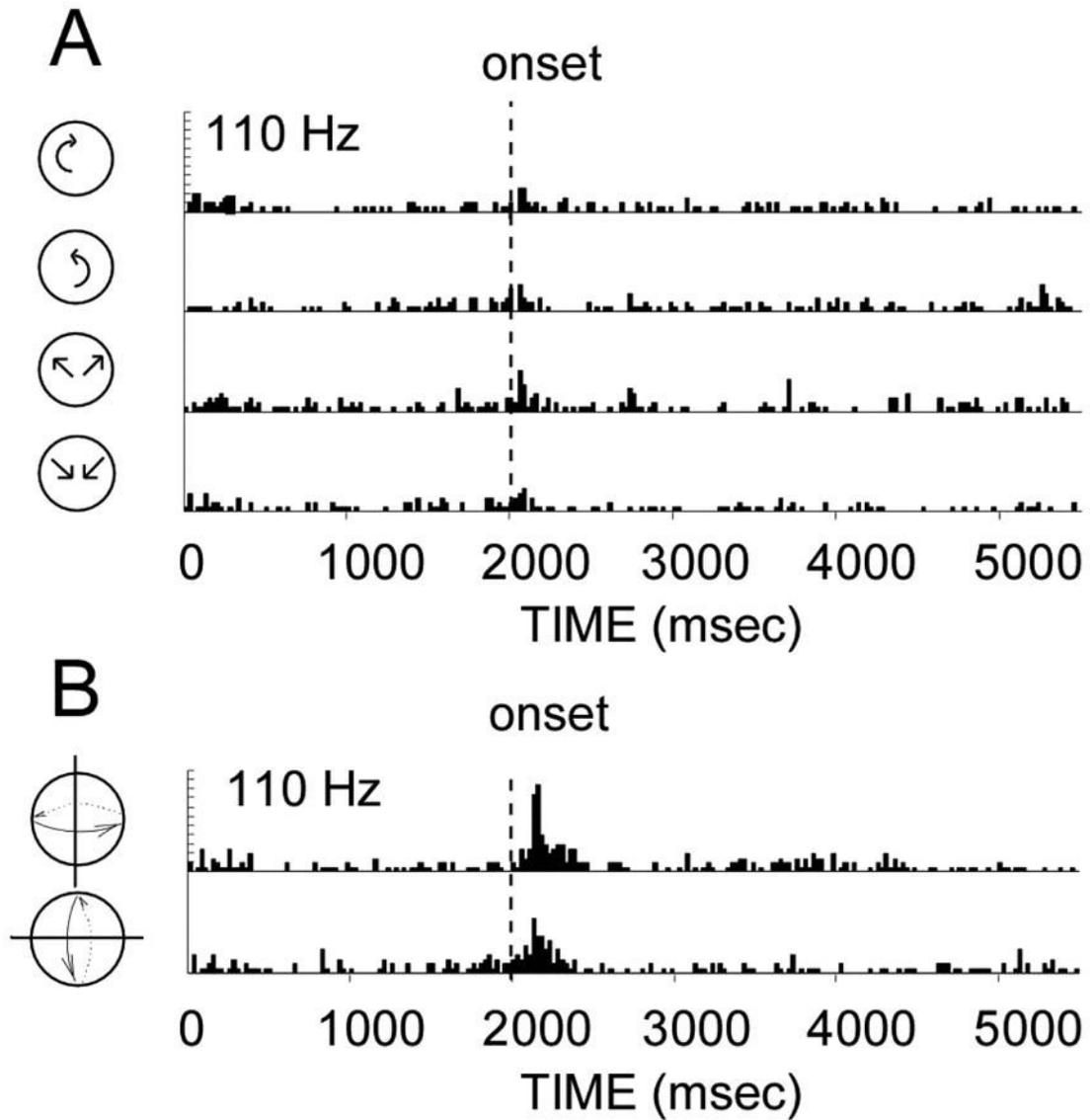
Figure 5.
Comparison of response to optic flow and three-dimensional structure-from-motion. (A) This cell had a strong response spheres for the two sphere displays. (B) The responses to the navigational flows were much weaker. The two sets of displays were matched in size, number of points, and distribution of speeds of the motion trajectories. Bin size = 25 msec; eight repeats for each histogram.