

University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

**CODING OF HEADS BY SINGLE NEURONS IN THE SUPERIOR
TEMPORAL SULCUS OF THE RHESUS MONKEY**

RHYS BEVAN

**DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF ST ANDREWS
FIFE, SCOTLAND, KY16 9JU**

1991



The Al₃Ge

Declaration for the degree of MSc

I, Rhys Bevan, hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Signed.

Date. 28/8/90..

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

CONTENTS

ABSTRACT	1
INTRODUCTION	3
AIMS	36
METHODS	38
RESULTS	59
DISCUSSION	75
ACKNOWLEDGEMENTS	89
REFERENCES	90
FIGURE LEGENDS	102
FIGURES	109
APPENDIX	i

ABSTRACT

This study investigated to what extent coding of objects (specifically heads) by neurons in rhesus monkey temporal cortex is viewer-centred or object-centred (view-independent). A second aim was to determine whether viewer-centred cells were preferentially tuned to four hypothetical "characteristic views" of the head.

Standard extracellular electrophysiological techniques were used to record from single neurons in the superior temporal sulcus (STS) of four rhesus monkeys (*macaca mulatta*).

Stimuli used were four or eight views of human and monkey heads and various objects as controls. Spontaneous activity was also measured.

For cells tested with eight views, linear regression was used to estimate the angle of view of the head to which the cell would respond optimally.

Eighty-eight cells were found to be viewer-centred, four were object-centred and three were classed as "object-centred/view-selective" (having both viewer-centred and object-centred properties).

Angles of optimal response of viewer-centred cells were distributed over a wide range and were not significantly clumped around the hypothetical characteristic views for the combined data. However, the optimal response angles of the cells from one

monkey did show a significant preference for these views and the combined data showed a trend in this direction which may be expected to reach significance with a larger sample. The data do however provide evidence against the hypothesis that only four views are coded.

The coding of many different head views is interpreted as having a role in the analysis of the direction of attention of other individuals.

Breadth of tuning varied considerably from very narrow to almost object-centred, with most cells giving half of their maximal response to angles of head rotation of forty-five - seventy degrees from the optimal angle.

INTRODUCTION AND REVIEW OF LITERATURE

Visual object recognition is a process of comparing sensory information with internal representations. These representations may be of specific objects, or of general categories of objects. Recognition may therefore involve either identifying an object as a member of a category or identifying an individual exemplar of a category. Thus according to circumstances we may recognise a particular object as, for example, "an animal", "a monkey", "a rhesus macaque" or "Becky".

For some types of objects which can be classified into a hierarchy of categories there is a "basic" level category, which is accessed before higher or lower levels. Which level corresponds to this basic level is learned and may depend on personal familiarity with the objects concerned, cultural factors, or expert knowledge (Rosch et al 1976). Thus the average person might recognize an object immediately as "a tree", whereas "a silver birch" would be the first words that came into the head of a botanist on seeing the same object.

Another complication is that the same category of object may be recognised visually in quite different ways under different circumstances. For example one might recognise both a detailed drawing in a textbook and a small black blob moving around a

lampshade in a certain way as exemplars of the "housefly" category.

This kind of learned flexibility is characteristic of humans. This can be contrasted with the performance of animals with somewhat simpler and more genetically pre-programmed visual recognition systems. A well known example is the "bug detecting" system of the frog. The retina and optic tectum of the frog contain cells which "recognise" a category of food items, responding selectively to small moving blobs (which in a frog's natural environment generally turn out to be flying insects) and activating appropriate feeding behaviour (Maturana et al 1960). But a frog will starve to death in front of a pile of equally nutritious dead insects, lacking the visual recognition apparatus to identify them as food. Since piles of dead insects are not frequently encountered in a frog's natural environment, there was never any selection pressure to develop a system to identify them. This example shows that it may often be useful to consider the role of evolutionary forces in determining what the visual system of a particular species may be capable of doing. This is the main reason for using non-human primates in research (including the experiments described below) on those aspects of visual physiology which are considered relevant to theoretical issues in human psychological research. A group of animals which are quite closely related

to us, which occupy an ecological niche presumed to be like that of our pre-human ancestors and whose brain anatomy and physiology are similar to ours are the best available models for the human system where invasive techniques are required.

The above examples show that visual recognition is not a simple, unitary process. However, if we confine our attention to the human or primate visual system, most cases of object recognition involve similar problems which must require similar solutions in terms of the type of stored representations used. The main constraint to the type of representation which must be involved in recognising objects is that it must be able to explain the phenomenon of object constancy, that is the ability to extract knowledge of the actual unchanging three dimensional structure of an object from a two dimensional retinal image which is liable to change radically in many different ways. Various "frames of reference" have been proposed to account for this ability (Feldman 1985, Marr 1977, Perrett et al 1987, 1989, 1990). It seems clear that the retinal image itself would not be viable as a frame of reference for actual recognition of an object, since there is an almost infinite number of possible retinal images which could represent the same object at different distances, in different orientations, under different lighting conditions and so on.

Possible frames of reference which could be used for

recognising objects include viewer-centred and object-centred systems.

A viewer-centred frame of reference depends on the position of the viewer relative to the object being recognized. A viewer-centred description of an object is specific to the particular viewpoint from which the object is seen. If the object is viewed from a different perspective, this representation no longer corresponds to the viewed stimulus in a meaningful way, and a completely separate representation is needed to enable recognition. Viewer-centred encoding would therefore seem to require that a very large number of different views of a particular object would have to be treated as separate objects. This seems inefficient in terms of storage space and because separate internal representations of each view would have to have their own separate associations to other memories, otherwise it would not be possible for an animal to behave appropriately and consistently when confronted with different views of, say, an item of food, a member of its own species or a predator. One particular form of this problem which has been pointed out (eg Marr 1982, Hasselmo et al 1989) is that if different views of an object were encoded separately, associations with memories formed through modalities other than vision would have to be made separately to each view of an object. Thus, for example, an association could be made between

one view of a novel food item and a reinforcing taste without any associations being made between this taste and other views. This would seem a serious problem for a biological recognition system, given the obvious importance of such associations between modalities to an animal which must learn quickly and efficiently about significant aspects of its environment to survive. (In some cases, however, there would clearly be a need to respond differently to different views of the stimulus. For example a lion facing forward and advancing would elicit a different response from one facing away and retreating. This example shows that viewer-centred encoding may well have important biological advantages if the problem of associations between views could be overcome).

Despite the apparent disadvantages, the direct approach of encoding a large number of separate views of an object has been used in neural network or mathematical models (Kohonen 1978, Hopfield 1982, Huberman and Hogg 1984, Willshaw et al 1969). These models make use of associative memory mechanisms which can store a large set of patterns, corresponding to various views of a particular object, which are associated in some way. An input pattern is compared with all these internal representations and the one most similar to the input is retrieved. These models all require a cumbersome set of associations between the various

patterns representing a particular object and so suffer from severe limitations in terms of storage space and processing time, especially if they were to be applied to natural objects which would require a very large number of separate views to be stored. To recognise reliably a natural object from any possible view using such a system would require a prohibitively large number of representations and associations. Such a direct system could only cope with views that were identical or at least very similar to one of the stored patterns. This contrasts with the actual performance of the human visual system which can often (although not always) identify objects from unfamiliar viewpoints. One way that this could be done is by the identification of certain regularities which belong to any view of an object.

Pitts and McCulloch (1947) proposed that there are indeed certain invariant properties common to all views of an object. A recognition system based on such properties would not depend at all on viewpoint, and would not therefore be viewer-centred. This is the idea behind the theory of high-order invariances proposed by Gibson (1950, 1979). Gibson suggested that invariant properties of an object could be reflected in higher order invariances in the so-called optical array. Objects could be recognised directly by identifying these invariances, which would be based on relatively low-

level visual information such as spatial and temporal gradients of texture density.

The invariant-properties approach has had some limited success in artificial vision systems designed to recognise restricted categories of simple objects, such as industrial parts (Bolles and Cain 1982). It has severe limitations when applied to more complex objects however, and can hardly explain the general recognition abilities of biological systems. It is difficult to imagine how such a system could distinguish between a dog and a cat for example. Ullman (1989, p 203) highlighted these limitations as follows: "The weakness of this approach is that in visual object recognition there is no particular reason to assume the existence of relatively simple properties (particularly global ones) that are preserved across the transformations that an object may undergo. It is not surprising, therefore, that despite considerable effort invariant properties of general applicability for visual object recognition proved difficult to find". The invariant properties approach does not therefore seem to provide a realistic alternative to a viewer-centred system in which a large number of separate views are encoded, with the inherent problems this would cause.

An alternative solution could be provided by an object-centred (view-independent) system of encoding. An object-centred description would be

based on certain characteristics of the object, regardless of the position of the observer relative to it. These characteristics would however, be specified in higher-order or more global terms than merely simple invariances visible in any view. Although the appearance of certain features of an object would change relative to the viewer when the angle of view changed, their positions relative to some significant point of reference on the object itself (usually the major axis) would remain constant, thus providing the information for it to be recognized whatever the viewpoint of the observer. With this system only one description of an object would have to be encoded for recognition to be possible from any view.

An object-centred system would not require multiple views to be encoded in memory. There is, however, empirical data which suggests that multiple views are encoded in the human and monkey recognition systems. Experiments conducted by Rock and DiVita (1987) showed that subjects performed poorly on a recognition task when there was a large spatial rotation separating a learned model from the viewed object. This seems to indicate viewer-centred encoding. There is also physiological evidence from the studies of Perrett et al (1985a, 1985b, 1989, 1990) for multiple views being encoded (see below).

Information about an object associated in memory to an object-centred representation could be

retrieved when any view was encountered. Although such a system would appear on a priori grounds at least more efficient than viewer-centred encoding, it presents formidable computational difficulties.

Marr & Nishihara (1978) and

Marr (1982) proposed that object recognition uses a single stored 3D object-centred representation which is compared with input constructed from a low-level viewpoint-dependent representation, known as the 2.5D sketch. Marr suggested that this sketch could be constructed by using retinal input alone without recourse to stored representations and would contain sufficient information (such as depth and orientation of local surface details) to allow computational matching to the correct 3D object-centred description of the appropriate object stored in the brain. The information in the 2.5D sketch is specified in terms of a retinal map. It could not itself serve as a high level viewer-centred stage from which a view of an object could be recognised without proceeding to the construction of the final object-centred description. It is important to note that actual recognition does not take place until the object-centred, view-independent representation has been accessed. The information specified in the 2.5d sketch, although containing depth cues (hence 2.5D rather than 2D) is dependent purely on retinal input. It does not have access to learned information about the 3D structure of objects and so

cannot provide knowledge of surfaces not visible in the retinal image. It is also too low-level to allow recognition by itself since it does not generalise sufficiently between different possible retinal images of a particular view of an object.

Marr and Nishihara adapted the proposal of Binford (1971) that an object could be represented in object-centred coordinates by a series of volumetric primitives known as generalised cones or cylinders. A single large-scale cone based upon the principal axis of the object would provide an object-centred frame of reference. Parts of the object, and parts of these parts could then be specified by smaller and yet smaller cones in a hierarchical way.

It is always the principle axis which defines the reference frame. Thus although the system of Marr and Nishihara is object-centred and view-independent and should therefore work equally well for nearly all conceivable views, a possible exception would be views in which the principle axis is difficult to see. Marr notes that if the principle axis was severely foreshortened relative to the viewer, the object would prove difficult or impossible to identify, since the pattern of the visual input would not correspond closely enough to the appropriate stored representation. This seems subjectively plausible in that, for example, long, thin objects seen end on are often difficult to identify. The main difficulty with interpreting any

perceptual difficulty with foreshortened views is that such views tend to have far fewer features visible than others, and so the effect can be equally well explained in terms of a model of recognition based on feature analysis as by the invisibility of the principle axis (Warrington and James 1986).

There is some evidence from neuropsychological studies that patients with certain forms of brain damage (particularly those with right posterior parietal lesions) who have difficulty identifying principle axes of objects perform poorly on recognition tasks (Warrington and Taylor 1973, Warrington and James 1986, Humphreys and Riddoch 1984). Humphreys and Riddoch (1984) described four patients with right hemisphere lesions who were very bad at recognising foreshortened views of photographs of objects, but scored at near-normal levels when the object photographs were rotated in their own plane through ninety degrees. The authors regard this as evidence for Marr's view that an object with a foreshortened main axis would be difficult to recognize. Warrington and James (1986), however, interpret these results as due to the lack of visibility of particular features in the unusual views with the foreshortened main axis, rather than to difficulty in retrieving the main axis itself. In their own clinical study, Warrington and James (1986) found that the degree of foreshortening of

the major axis of an object, measured by angle of rotation, did not have a systematic effect on recognition thresholds, and the authors interpreted their overall results as more consistent with a distinctive-features model of object recognition than with Marr's model. It is thus not possible to decisively either confirm or refute Marr's ideas about the role of the major axis in recognition by examining the clinical data.

Although the ideas of Marr & Nishihara have been very influential in directing work in the fields of artificial intelligence and cognitive psychology, there are many problems for their model which remain unsolved. For instance many objects have neither an obvious main axis or an obvious hierarchical structure, which are both requirements of this model of encoding. An example given by Marr & Nishihara themselves is a crumpled newspaper. Another example is a human head, in which the longest axis is not greatly different from the other axes. It is also extremely difficult to describe rules for how to compute a 2.5D sketch from the low-level information contained in the first stage of the Marr & Nishihara model - the "primal sketch", or for how to construct the final object-centred representation from the 2.5D sketch.

Lowe (1985, 1987) proposed an alternative object-centred recognition system which differs from Marr's model in not being dependent on the retrieval of a

central axis. Instead the relationships of parts are defined with reference to other parts. This system allows particular parts of an object to be more salient than others. Especially salient parts would be often used in recognition and so views in which they were occluded would be difficult to recognize.

Biederman (1987) has put forward a system of recognition by components (RBC). As with Lowe's scheme, some of these may be particularly salient. The components are described by a set of thirty-six volumetric primitives known as "geons". These geons are similar to the primitives used in several other models (Binford 1971, Marr and Nishihara 1978, Brooks 1981, Pentland 1986), consisting of simple three dimensional shapes such as boxes, cylinders and wedges. Biederman claims that by using this system most objects can be recognised by a combination of only one to three distinctive components. Complex objects are described by decomposing them into parts, each described by a particular geon, with a description of the spatial relations between them. Description of components is qualitative rather than quantitative. Perrett and Harries (1988) point out that this approach may be useful when "recognition" involves deciding whether an object belongs to one category or another, but seems inadequate in discriminating between individual exemplars of a particular category, since these kinds of intracategory discriminations would

often be mainly or purely quantitative in the case of homogenous categories. For example one might discriminate a dog from a cat or perhaps a greyhound from a poodle in a qualitative way, but recognising an individual greyhound would probably require a quantitative analysis.

Recognition of components in a qualitative way would show considerable tolerance to changes in viewing angle, but Biederman's system is to some extent viewer-centred in that there would be some viewpoints from which important components would not be visible at all.

Another recognition system based on decomposing an object into parts is that of Hoffman and Richards (1986) which is designed to recognise contours. This is complementary to Biederman's system, which assumes that an analysis of contours has already been done.

Viewer- and object-centred recognition systems each present their own problems: on the one hand the extravagant inefficiency of encoding a huge number of views and forming all the necessary associations, and on the other the enormous computational difficulties involved in a purely object-centred system such as that put forward by Marr. These problems have lead to the proposal of an alternative solution which represents a compromise. This is that recognition does depend on viewer-centred representations, but that instead of a very large

number of possible angles of view, only a relatively small number of "characteristic" or "prototypical" views of an object need to be represented in memory (Koenderinck and Van Doorn 1976, 1977, 1979; Callahan and Weiss 1985, Rieger 1986, Gray 1986, Perrett et al 1984, 1985a, 1987, 1989). All versions of this scheme recognise that the characteristic views involved must be high-level, generalizing over position, size, lighting, orientation and (obviously if the number is to be relatively few) to a certain extent over perspective. Which and how many views would be required would depend partly on the amount of visual information each view could provide and partly on the probability of seeing particular objects from particular views. For example you would not often need to recognise a telegraph pole from directly above. (If you did you could probably only do so by relating its position to other objects which were recognisable such as roads).

Koenderinck & van Doorn (1977) introduced the term "view potential" in this context to emphasise the fact that an object has only a limited number of qualitatively different views. Thus the qualitative character of many views of an object is often the same in that the same main features are present even if there are differences in the quantitative details of these features. It is only when the angle of view is changed sufficiently for significant features to appear or disappear or to seem radically rearranged

relative to each other that a qualitatively different view or "aspect" results. As an observer moves relative to an object, there are positions where a small change of view will produce only a quantitative change in the object's image (stable aspects) and others where a qualitative change is produced, and a previously hidden part becomes visible (unstable aspects). The "line singularities" of an object are its distinct stable aspects, which can be mathematically deduced (Koenderinck and Van Doorn 1976). Koenderick and van Doorn give a house as an example. Despite the infinitely large number of possible different views of a house in the horizontal plane they can all be assigned to one of only eight genuinely qualitatively different characteristic views: the front, two sides, back, and the four intermediate views. (This example may be slightly misleading, since both the house itself and certainly more complex objects would require many more views to provide a complete description using actual line singularities rather than just subjectively different views - see below).

If characteristic views could be used for recognition, a viewer-centred system of encoding would obviously be far more viable than if a huge number of views had to be encoded. Although the necessity for separate associations with other memories would still exist, this would be much less of a problem if only a comparatively small number of

associations needed to be formed. Moreover, a system of viewer-centred encoding based on only a small number of characteristic views with high-level properties could easily be used as an intermediate stage in building object-centred representations, as well as for direct recognition in its own right (Perrett et al 1985a, 1987, 1989). This would be much less of a problem than trying to construct an object-centred description from low-level input specified in terms of retinal coordinates, as in Marr's model.

Callahan and Weiss (1985) followed Koenderink and van Doorn in representing different aspects of an object by reference to an imaginary viewing sphere, with the object at the centre. This sphere can be split into areas over which a given aspect of the object remains visible. Lines separating these areas represent viewpoints where a qualitatively different aspect of the object emerges through a change in the type of singularity visible. Callahan and Weiss (1985) mapped out viewing spheres for a number of simple smooth objects, such as a sphere with a bump.

Using the aspects deduced from singularity theory to define characteristic views has one major problem. This is that complex objects would have a very large number of aspects. For example Koendrinck and van Doorn (1979) estimate that the sculpture "Astronomy" by Giovanni Bologna would have many thousands of aspects. It is possible, however, that

not all the different aspects need be equated with characteristic views. For instance a selection of scale could be used so that only those singularities resulting from substantial bumps on the surface of an object would be considered (Callahan and Weiss 1985).

As an alternative to the use of line singularities, several computational schemes have been developed which have defined characteristic views of an object in terms of elementary visual features present in particular views (Chakravarty and Freeman 1982, Thorpe and Shafer 1983, Gray 1986, Ikeuchi 1987). Gray (1986) used cluster analysis to define groups of adjacent views which share many features with each other, but few features with other clusters. The merit of any particular features in determining the characteristic views is assessed independently for each object, rather than seeking to define features or parameters of universal importance. Gray's system is therefore more empirical than those based on line singularities and has the advantage of producing a manageably small number of characteristic views.

Poggio and Edelman (1990) attempted to implement a viewer-centred recognition scheme in a computer simulation of a neural network which could learn to recognise three-dimensional objects. Wire-frame objects were used, in which all features are visible from any view. The authors claim that the scheme

could be generalised to opaque objects by "partitioning the viewpoint space for each object into a set of "aspects", corresponding to stable clusters of visible features" (cf Gray 1986 and singularity theory). Poggio and Edelman postulate that a "standard view" of any object can be mapped onto any possible view and that this standard view can be constructed by means of a multivariate function which can be derived from a relatively small set of prototypical views. Such a function would be object-specific, with different functions corresponding to different 3D objects. This function was calculated by Poggio and Edelman's network using a scheme known as Generalized Radial Basis Functions (GRBF).

Results showed that 80-100 random training views were necessary to achieve satisfactory performance in recognising novel views for the full viewing sphere. Limiting the attitude of the object to one octant of the viewing sphere produced a satisfactory recognition performance with 10-40 training views, with just two GRBF basis units required. Poggio and Edelman regard this as evidence that only a very small number of basis units would be needed for each aspect of an opaque object. It was also found that after training, the centres of the radial basis units corresponded to views that were different from any of the training views. This may indicate that

stored characteristic views of an object could be modified by learning.

Ullman and Basri (1989) proposed a mathematical model of object recognition in which a 3D object is represented by the linear combination of a number of 2D images. It is proposed that for a limited range of rotation (around 60 degrees) of the object only three 2D images would be required in the case of objects with sharp contours, and six for smooth objects. This linear combination approach requires that different 2D images can be aligned using points of reference which must be visible in all cases. For this reason the system does not work for views where these points are occluded. Ullman states that "To represent an object from all possible viewing directions (eg both "front" and "back"), a number of different models of this type will be required". This could perhaps be regarded as the equivalent of building one object-centred description from a relatively small number of high level viewer-centred descriptions (Perrett et al 1985a, 1987, 1989).

Perrett and Harries (1988) set out to empirically test whether characteristic views of objects were used in recognition by human subjects. A second aim of the experiments was to determine whether singularity theory could be used to explain which views subjects appeared to use. Subjects were required to learn the surface appearance of objects placed on a stand which was free to rotate and were

tested on a subsequent recognition task. Two separate experiments were done, each using one category of objects. These were tetrahedra, as examples of regular polygons, and potatoes, as examples of smooth but irregular objects. The recognition task was thus to discriminate between individual exemplars of a category rather than identifying the category. Subjects were allowed to rotate the turntable at will and the angle of the turntable was measured to determine whether any views were inspected preferentially.

During the learning phase, subjects concentrated the majority of their time on a small number of views (two to six, depending on the object). Different subjects tended to prefer the same views. These results held true for both tetrahedra and potatoes. This does seem to provide some evidence for the existence of preferred views. However, the views chosen were shown not to accord with either singularity theory or with Gray's (1986) computational model. In the case of potatoes the preferred views were found to correspond to the maximum and minimum stimulus width/area (ie the end view and side view). In the recognition phase of the experiment there was no evidence for preferred views, with no differences in recognition scores between different views of either class of object.

Harries et al (1990) did two experiments with similar aims to those of Perrett and Harries (1988),

but using clay models of human heads as stimuli. The first experiment used a similar paradigm to that of Perrett and Harries. Again results showed that when learning the appearance of the heads, subjects did not distribute their time evenly, but concentrated on two particular views: the face view and a view close to the profile (72 deg away from the full face). All subjects spent most time on the face view. Twelve out of fifteen subjects spent more time on the 72 deg view than the 90 deg (full profile) view, and this difference was statistically significant overall. There was also a slight tendency for the back of head view to be preferentially inspected compared with adjacent views. Each of the six individual heads used showed the same general pattern of inspection. The learning phase was followed by a recognition task in which subjects had to discriminate between familiar and novel heads in various views.

Results showed an effect of view on accuracy of recognition, with the back of head view eliciting significantly more errors than the face, profile and "half profile" (45 deg from face) views. There was no significant correlation between mean inspection time and mean accuracy scores for different views.

The second experiment used a different approach to deduce characteristic views of heads. This time inspection times for different views were measured while subjects attempted to recognise heads from

memory, rather than during learning. For the learning phase, subjects were shown a video film of three clay heads, rotating at a fixed velocity. They were thus evenly exposed to all views during encoding. This was followed by a recognition task in which subjects had to distinguish the three heads they had seen from three novel heads. The stimuli were placed on a turntable which could be controlled by the subjects. The objects were viewed on a monitor (via a video camera) to make the images equivalent to the learning phase. Results showed similar preferred views to experiment one, with a significant correlation between the distribution of inspection time in the two experiments. This indicates that subjects preferred the same views for recognition purposes as they preferred when attempting to encode objects into memory.

Harries et al note that the preferred views for heads cannot be explained in terms of maximum and minimum stimulus width/area, as was suggested for the preferred views of potatoes in the study of Perrett and Harries (1988), since this hypothesis would predict that the face and back of heads views (which have the same stimulus width/area) would be inspected equally. Harries et al propose that the presence of facial features in the likely explanation for preferred inspection of the face view. The results do seem to provide evidence for the encoding of prototypical or characteristic views

of heads and their subsequent use in recognition. However the fact that not very realistic clay model heads were used in the experiments means caution is required when attempting to generalise the findings to real heads as a category.

Heads are a useful category of objects for the physiological study of the mechanisms of object recognition. There are two main reasons for this. The first is that single unit electrophysiological recording techniques have shown that the temporal cortex of different species of macaque monkey contains neurons which are selectively responsive to faces and/or other views of the head (Bruce, Desimone and Gross 1981, Gross et al 1982, Perrett, Rolls and Caan 1982, Desimone et al 1984, Rolls 1984, Perrett et al 1985a, Bayliss, Rolls and Leonard 1985, Mikami and Nakamura 1988, Hasselmo et al 1989). Most of these cells have been found in the superior temporal sulcus (areas TPO, PGa and TEa of Seltzer and Pandya 1978, Pandya and Yeterian 1985). Secondly, case studies of brain-damaged patients suffering from a selective inability to recognise faces (prosopagnosia) provide reasonably convincing though not conclusive evidence that humans also possess an anatomically-localised neural mechanism specifically for recognising faces. (For a review of the relevant human case studies and their relation to monkey physiology, see Perrett et al 1988).

The main studies which have examined the effect of systematically varying the viewing angle on the responses of neurons selectively responsive to static heads are those of Perrett et al (1985a) and Hasselmo et al (1989). Studies by Perrett et al (1985b, 1989, 1990) on STS neurons selectively responsive to different body movements have also shown effects of head/body view and interactions between view and direction of movement.

Perrett et al (1985a) found evidence of both viewer-centred and object-centred encoding by cells in the STS, with the former predominating. Sixty-nine percent of all the cells recorded which were selectively responsive to at least one view of a head were defined as "view-sensitive", ie differentially responsive to the sight of the object from different vantage points. The majority of these cells responded to different exemplars of a particular head view despite changes in viewing distance and image size and despite isomorphic rotation. These two properties - selective response to different views but insensitivity to other large changes in the stimulus - are consistent with a system of high-level viewer-centred encoding. The authors point out that the fact that the responses of these cells were not significantly affected by large changes in the retinal image caused by changes in factors other than view (eg distance, orientation, colour, intensity) shows that they do

not respond in a way which corresponds to the 2.5D sketch proposed by Marr & Nishihara (1978), since this would need to provide detailed information about surface texture, distance from the viewer etc. Rather the responses of these cells must be seen as providing high level descriptions which could in themselves act as viewer-centred mechanisms for recognizing particular views of heads or could at least constitute an important part of such a mechanism. They could possibly also provide inputs for a still higher stage of object-centred descriptions which would enable recognition of the head from any view.

Perrett et al (1985a, p 314) report that one quarter of cells responding to the face were "relatively insensitive to viewpoint", responding to several views. These seem to provide evidence for object-centred encoding.

All the cells classified as view-sensitive (in the horizontal plane) could be assigned to one of four discrete populations, responding optimally to either the face, left or right profile or the back of the head. (This is actually a slight simplification in that most cells responsive to profiles did not discriminate between the left and right). Two further populations responded to the head tilted up and down. The important point is that no cells were found to respond more to intermediate views than to the face, profile and back of head views. These

results seem to provide strong evidence both for encoding of multiple views (and therefore a viewer-centred frame of reference) and the use of particular characteristic views in encoding heads.

Perrett et al suggest that any view of a head in the horizontal plane could be represented by four such populations of cells, each responding optimally to one particular characteristic view, but with sufficiently broad tuning so that an intermediate view would elicit significant responses. In fact an analogous form of encoding has already been shown to exist in the visual system. This is in the retina, where the whole visible spectrum is represented by the combined activity of three distinct populations of cone cells, each tuned to respond optimally to one of only three "characteristic" wavelengths, corresponding to the colours red, green and blue. A similar strategy for encoding heads could account for the more efficient processing of the "half-profile" relative to other views which has been shown in some psychological experiments (Woodhead and Baddeley 1979, Bruce and Valentine 1987, Krouse 1981, Logie, Baddeley and Woodhead 1987, Fagan 1979, Thomas et al in preparation). Given sufficiently broad tuning, the half-profile view would evoke significant responses from neurons tuned both to the face and the profile. Although these responses would be suboptimal, the combined activity could be more than that caused by only one population of cells

firing at their optimal rate. It should be noted, however, that the data on which views are easiest to recognise is inconsistent and the three-quarter view advantage was often weak and unreliable. View advantages often seem to have been dependent on the demands of a particular experimental task and tended to disappear when more natural tasks or stimuli were used (for review see Thomas et al, in preparation).

Hasselmo et al (1989) report results from two experiments which provide evidence for object-centred encoding by cells in the STS.

The first experiment used various head rotations performed by one of the experimenters as well as a variety of static and moving controls as stimuli. All of the ten neurons found which responded selectively to rotation responded to a particular movement independent of the orientation of the moving head in relation to the viewer, continuing to respond to this particular movement even when the head was inverted or shown from the back. This was taken by the authors as evidence that the movement was encoded in object-centred coordinates.

The second experiment used various static views of heads. Thirty-seven cells which had previously been classified as head-selective according to the criteria described by Rolls (1984) and Baylis, Rolls and Leonard (1985) were tested with several different views of the heads of different individuals. Eighteen of these cells showed

selectivity between different individual heads in all views tested. The authors interpret this result as evidence for object-centred encoding. (This does not seem conclusive evidence, however, since the possibility exists that differences in simple features visible in all views (such as the hair) could be responsible for the results). Fifteen of the cells which were sensitive to identity in all views also showed sensitivity to the viewing angle. There were therefore only three cells with responses unaffected by viewpoint.

Nineteen cells did not discriminate between different heads. Sixteen of these did show significant differences between responses to different views. These cells can presumably be regarded as providing viewer-centred encoding (provided that responses were higher than spontaneous activity).

Hasselmo et al estimated an "angle of orientation specificity" for each cell. The orientation specificity of each cell was represented as a vector in which the angle is the angle of view to which the cell would respond optimally and the length represents how much the response varied with the viewing angle (ie the tuning). Results showed that most cells were not sharply tuned to one view, but rather their responses generalised across views to a considerable extent.

When the angles of orientation specificity of all cells were plotted out, it was shown that although there were more neurons responding to the front half of the face than the back half, there was no tendency for cells to respond maximally to particular views - in fact angles of best response were distributed over a wide range of view. The authors concluded from this that there was no evidence that only a few characteristic views are encoded.

This conclusion is in marked contrast to that of Perrett et al (1985) who found that only certain views were encoded. It should be noted that the study of Hasselmo et al was more quantitative and thorough than that of Perrett et al and the results are probably more reliable. In particular Hasselmo et al subjected every cell recorded to a standard testing procedure, which included the presentation of a large number of views. Perrett et al presented fewer views and it is not clear that all cells were in fact tested equally with both those views hypothesised as being prototypical and the intermediate views, neither was there any attempt to calculate the precise angle of optimal response. The work of Perrett et al was in general more anecdotal and qualitative in character than that of Hasselmo et al.

The second major difference between the results and conclusions of Hasselmo et al (1989) and Perrett

et al (1985a) is that Hasselmo et al classified a much higher proportion of cells as object-centred. This may have been due at least in part to differences in the method of classification. Firstly Hasselmo et al classified cells as object-centred or not on the basis of an intra-category (identity) discrimination, while the study of Perrett et al was concerned with the recognition of a category. Secondly, Hasselmo et al classified cells as object-centred on the basis of an arbitrary statistical criterion, namely a significant effect of identity with two-way ANOVA on view against identity. This procedure would certainly not require that a cell discriminate between stimuli in all views to be classified as object-centred, and would also allow clearly view-selective cells to be classified as object-centred. Thus the criteria used by Hasselmo et al for object-centred encoding were much less rigorous than those of Marr and Nishihara (1978), which would require both a response to all "normal" views (with the principle axis visible) and absence of view sensitivity. Perrett et al (1985a), on the other hand, used a system of classification nearer to that of Marr and Nishihara, although they did not use objective quantitative criteria to classify object-centred cells.

Studies by Perrett et al (1985b, 1989, 1990) have found cells in the STS which are insensitive to static heads/bodies but which respond selectively to

these stimuli when moving, thus providing evidence for the convergence of static and dynamic information. Many of these cells showed viewer-centred properties, often being both view and direction selective, responding for instance to the left profile view of an experimenter walking to the monkey's left, but not to other views or directions. A much smaller number of cells showed object-centred properties, for instance responding selectively to an experimenter walking "forward" in any direction (following the nose). These object-centred cells demonstrate a form of encoding which accords with the system for the recognition of moving objects proposed by Marr and Vaina (1982).

In addition to the viewer-centred and object-centred coordinate systems, the studies of Perrett et al showed that the responses of a number of STS cells could best be explained by postulating a "goal-centred" frame of reference. For example one cell responded to the movements of an experimenter, but only when these movements led the experimenter to the door of the laboratory. Turning the monkey to face a different position in the laboratory made no difference to the direction selective response of this cell, when the movements of the experimenter were related to the door (Perrett et al 1990). Such a "goal" presumably has a learned significance for the monkey. Goal-centred representations may not be appropriate for actual recognition of objects but

are, as Perrett et al (1990, p 198) note, "...particularly useful for describing actions, since they make explicit the relationship between an agent performing an action and the goal of that act". A goal-centred frame of reference could also be used to describe the relationship between two static objects, for instance an experimenter facing toward the place where the monkey's food is kept. In this case the view would presumably serve to anticipate movement toward the goal.

AIMS OF THE PRESENT STUDY

The experiments reported below were designed to make use of electrophysiological methods to investigate in a quantitative and objective manner the way in which a specific category of objects, namely heads, are encoded in the brain of the rhesus monkey, with the hope that the results might make a contribution toward reaching a better understanding of the process of object recognition in general.

The study was limited to three principal aims. The first was to assess the extent to which the responses of head-sensitive cells indicated viewer-centred or object-centred coding. In the case of viewer-centred cells the second aim was to estimate the angle of view of the head which would elicit the optimal response of each individual cell. This was done in order to test the hypothesis that four hypothetical characteristic views: the face, left and right profiles and back of the head, are selectively or preferentially represented in the brain. These four views were chosen because a previous study (Perrett et al 1985a) had provided strong evidence that these particular views were coded preferentially, and indeed had found no evidence that any other views were represented at all.

The third aim of the study was to measure the breadth of tuning of the cells' responses to various

head views. This was important to determine the number of views that would theoretically need to be coded in order to allow recognition of an object from any viewpoint.

METHODS**Subjects**

One adult female and three juvenile male rhesus macaque monkeys (*macaca mulatta*), henceforth referred to as F, B, D and H. The female weighed approximately 4 kg and the males ranged from 5 - 10 kg. The four subjects were caged individually, but housed with other monkeys.

Recording area

The target area for recording was the anterior part of the upper and lower banks of the STS. This corresponds to parts of areas TPO, PGa and TEa of Pandya and Yeterian (1985) and Seltzer and Pandya (1978).

Construction of chronic recording implants

Two stainless steel rings (16mm ID, 19mm OD, 10mm deep), designed to fit a David Kopf hydraulic micro-drive, were placed onto a sheet of glass overlying a fixed piece of graph paper. In this way the rings could be fixed relative to one another in

appropriate positions with reference to two orthogonal axes which when positioned on the monkey's head would correspond to the saggital and interaural planes. For the final stereotaxic coordinates, the right ring (well) was centred 8mm and the left well 10mm anterior to the interaural plane; both wells were centred 12mm lateral to the saggital plane.

The edges of the rings were cemented to the glass plate with dental acrylate (Howmedica Int., London), care being taken that acrylic was not placed more than 2mm high on the walls of the rings, since this would interfere with the microdrive's subsequent positioning. Plastic tubes (7mm ID, 14mm OD, 61cm length) were placed in front of (approx. 38mm anterior) and behind (approx. 25mm posterior) the rings and cemented to them by a central line of acrylic which was placed around the centre of the tubes. (Horizontal bars could be inserted through these tubes and clamped to the sides of the primate chair for recording sessions, to ensure restraint of the head.)

When this construction was dry, it was floated off the glass with water and, in preparation for the following operation, placed in a stereotaxic holder (David Kopf Instruments, California) at predetermined antero-posterior coordinates. The laterality was adjusted with respect to the midline saggital sinus.

Surgical procedures

The monkey was first sedated with a weight-dependent dose of intramuscular ketamine (Ketelar, 100mg/kg, Parke Davis and Co., Gwent) and then anaesthetized with 1-2ml of a barbiturate anaesthetic (Sagatal, 60mg/ml, May and Baker Ltd, Dagenham), introduced through an intravenous cannula, until the cessation of the gabella reflex. 1.0 ml dupocilin, an antibiotic, was injected by the intramuscular route and 1.0 ml atropine (0.6mg/ml, Evans Medical Ltd, Liverpool) was given to reduce mucous secretion. The monkey's head was closely shaved and subsequently swabbed with alcohol and then iodine. A drop of paraffin oil was placed in each eye to prevent corneal drying or damage, the eyes taped shut and the monkey placed securely into a David Kopf stereotaxic instrument. Full sterile precautions were then employed throughout the surgical procedures. Respiratory rate (30-40/min) and anaesthetic level were maintained constantly throughout the operation with additional anaesthetic being administered when necessary through the intravenous cannula.

The initial incision was made longitudinally along the skull, i.e. in the antero-posterior direction. The skin was reflected and a suitable area of skull cleared of connective tissue to the size of the

implant. The implant, held at the preset coordinates in a stereotaxic arm, was then lowered to the skull and the size and position of the rings marked. The implant was raised and the circular areas of bone beneath the rings were removed by drilling around them with a dental drill. Following this suitable positions on the skull were marked around the periphery of the implant for small holes for the insertion of stainless steel screws and T pieces. With the screws and T pieces locked in position (forming a firm frame to which the implant could be attached) the implant was lowered to sit flush with the skull and cemented in the preset antero-posterior coordinates calculated from a macaque stereotaxic atlas (Snider and Lee, 1961) and histological brain sections of rhesus monkeys of equivalent weights. The implant was then built up around the wells (rings) and plastic tubes with dental acrylic. PEP, a topical antibiotic (2% powder, Intervet Laboratories Ltd, Cambridge) was applied to the exposed dura and plastic aerated caps fitted over the wells. The animal was subsequently removed from the apparatus and returned to its home cage where recovery occurred within a few hours.

At the beginning of each experiment both the plastic well caps were removed and the inside of the wells and the exposed dura swabbed with very dilute disinfectant (Savlon). 0.1-0.2ml lignocaine hydrochloride (Xylocaine, 40mg/ml, Astra

Pharmaceuticals Ltd, Kings Langley), a topical anaesthetic, was applied to the dura before inserting the guide tube 3-5mm through the dura. During periods when no recording took place the wells and well caps were regularly cleaned with dilute Savlon to prevent infection.

At the end of the experiments, the monkey was first sedated with ketamine, then killed with a lethal dose of barbiturate anaesthetic followed by transcardial perfusion (see below).

Electrode construction

Recording microelectrodes were constructed from 150mm straightened tungsten wire (125 microns OD) and fine capillary tubes of borosilicate glass (250 microns ID, 380 microns OD, 150mm length) using techniques adapted from Merrill and Ainsworth (1972). A piece of tungsten was electrolytically sharpened by repeatedly dipping one end of the wire into Levick's solution (after Merrill and Ainsworth, 1972) while a voltage (2-3V) between this tungsten wire and a neutral carbon graphite electrode was maintained using a Regulac variable transformer (Claude Lyons Ltd, Herts). When the tip was a suitable shape (discussed later), the wire was inserted (blunt end first) into a tube of insulating glass. The etched tip was drawn back 5.0cm inside

the tube, its position marked on the glass and a drop of rapid-drying cyano-acrylic superglue (Loctite UK, Herts.) placed on the exposed end of the wire at the glass-wire boundary.

The microelectrode was then firmly secured in the electrode-pulling apparatus with the tip a fixed distance above a nichrome heating coil. Glass was then collapsed onto the metal electrode by passing a current through the coil and following this, glass insulation was stripped from the very tip of the electrode. This was achieved (under microscopic control) by inserting the electrode tip into a bead of solder glass, heated on a platinum-irridium wire, allowing it to cool and then removing the electrode tip from the bead.

Using a small volume of Levick's solution held in a glass capillary tube, the tungsten metal could be electrolytically re-etched (as above) if necessary, to a final tip shape approximately 6-10 microns wide and 10-25 microns long. The tip shape determined the electrode's performance and a trade-off was often necessary between a long, fine point for good (large signal to noise ratio), stable, single cell recordings and a shorter, blunter tip for durability in traversing brain tissue.

Electrode placement

Microelectrodes were advanced into the brain through a trans-dural guide tube (0.5mm ID, 1.0mm OD, 5cm length) using a David Kopf hydraulic micro-drive and micropositioner adapted to allow electrode passage at angles of 0-15 degrees with respect to vertical. The height at which the guide tube touched the dura was used as a fixed zero reference for calculating the depth of travel of the electrode into the brain. The microelectrode was lowered by hand to a predetermined distance below the dura, judged from X-ray measurements to be approximately the depth of the temporal cortex. At this depth the electrode was attached to the microdrive using poster adhesive (Blu Tak, Bostic Ltd, Leicester) and advanced or retracted in 10 micron steps, using the hydraulic system, to investigate the temporal cortex.

Signal detection

Electrical activity from the microelectrode was preamplified by a field effect transistor (NeuroLog head stage NL 100) and further amplified by a second preamplifier (NeuroLog NL 104). The neural signal (1-5KHz) was filtered (NeuroLog NL 125) below 0.5KHz and above 15KHz and displayed at slow time base (5ms/cm) on the main oscilloscope (Telequipment

DM63). The activity of individual cells was selected and digitized with a voltage window (Digitimer D.130).

When two or more cells were encountered with similar action potential amplitudes, the microelectrode position was adjusted until it was possible to discriminate one spike as having a larger amplitude than the other. The voltage trigger level could then be altered to catch only one of these spikes and using an analogue delay of 1ms (NeuroLog NL 740) the entire wave form of the selected action potential could be monitored on a second oscilloscope at fast time base (0.5ms/cm). Selected signals exceeding the lower voltage window level were also amplified for audio output through a loudspeaker system. This audio output enabled value judgements on the firing rate of a cell to be made and provided a qualitative indication of neural responses to stimuli and an audible indication of electrode drift from a given cell.

Location of cells recorded

Frontal and lateral X-radiographs were taken of the position of the electrode at the end of each experiment. Reconstruction of the position of electrodes in the brain was achieved by reference to the positions of micro-lesions (10 microamp DC for

30 s) made at the end of some electrode tracks.

Microlesions were subsequently identified in histological sections. Additional markers used in calibration of electrode position were provided by micro-injection of anatomical tracers (horseradish peroxidase and the fluorescent dyes true blue and diamidino yellow) at the site of cell recording on some recording tracks. For these markers the X-ray position of injection could be compared to the anatomical location of injection revealed through normal or fluorescence microscopy.

At the end of experiments a sedating dose of ketamine was administered followed by a lethal dose of barbiturate anaesthetic. Phosphate buffered saline and 4% glutaraldehyde/paraformaldehyde fixative were then transcardially perfused. The brain was removed and sunk in successively higher concentrations (10, 20 and 30%) of sucrose solution or Dimethylsulphoxide (DMSO) and glycerol as cryoprotective agents. The brains were blocked at the occipital pole and sectioned whole in the coronal plane for routine histology. The brain was photographed during sectioning. Photographic slides and cresyl fast violet sections were used to construct and atlas of each brain, every 0.5 mm. The photographic slides provided an accurate record of the position of sulci and cortex without distortions introduced during histology mounting.

The antero-posterior within the STS cortex was specified relative to the middle of the lateral geniculate nucleus (LGN). This was defined as the brain section where (a) the LGN attained maximum size and/or (b) the midway between the most anterior and posterior sections with LGN cells evident.

Recording methods

Single unit activity was recorded from awake, behaving monkeys using physiological techniques that have been described previously (Perrett et al 1985).

Subjects were restrained in a primate chair for periods of two to five hours (average three hours) during which recording took place. An individual monkey was recorded up to a maximum of five times a week and never more than once a day.

Various types of visual stimuli were presented behind a large aperture shutter (6.5 cm). The shutter was opened for each trial to reveal the stimulus and remained open for a period of 1 s. Neuronal firing rates were measured in periods of 250 ms and 500 ms, in each case beginning 100 ms after the shutter opened. These data were analysed on line by a Hyundai microcomputer.

During some experiments eye movements were monitored using an infra-red corneal reflection system. This was done to determine whether any

results obtained could have been due to the monkeys selectively attending to some stimuli rather than others (eg face but not back of head etc). In the event it was shown when eye movements were monitored that results could not be explained by selective attention and that the monkeys did nearly always look in the direction of the stimuli. This was therefore assumed to be the case with all the experiments.

Before chronic implantation the subjects were trained to do a task designed to ensure their attention to visual stimuli. This task was to discriminate between red and green lights. A single LED light was used, which could be switched to either colour. This was situated level with the monkey's line of sight on a blank white wall 4 m in front of it.

The situation in which training took place was the same as for the actual experiments, except that before implantation there was no means of fixing the monkey's head rigidly in place.

The monkey was restrained in a primate chair. A lick tube was positioned in front of its mouth and the shutter fixed in place in front of its head. A tone sounded just before the shutter opened to signal to the monkey to attend. The LED came on simultaneously with the shutter opening and remained on for the 1 s period the shutter remained open. If the monkey licked the tube when the light was green,

it was automatically reinforced with a reward of dilute sweetened fruit juice. If it licked when the light was red it received aversive hypertonic saline solution. The monkey had to lick during 1 s during which the shutter was open to obtain liquid. The interval between trials was varied.

The subjects continued to do the task simultaneously with the experiments. Thus whenever the shutter was opened to reveal the experimental stimulus, the green or red light came on and fruit juice or saline were automatically delivered according to the monkey's response. The 2D images used as experimental stimuli were projected onto the wall on which the LED was located. The 3D stimuli were presented at varying distances in front of the wall and to either side of the LED. In this way the monkey's attention was directed toward the experimental stimuli.

Subjects were deprived of water for periods of up to twenty-four hours before each training and recording session to ensure sufficient motivation for the task. They were carefully monitored for signs of dehydration and were given free access to water at weekends and periods when no recording or training was done.

Stimuli used

The stimuli used for testing possible head-selective cells were various views of static heads and a large number of different control objects. Both real 3D heads (the experimenters) and 2D heads (video disk images and slides of the heads of the experimenters and various monkeys) were used. All of the human heads used were equally familiar to the subjects. Eight different views of most of these heads were available. These were the four hypothetical characteristic views, namely the face (0 deg), left profile (90 deg), back of head (180 deg) and right profile (270 deg); plus the four intermediate views: 45 deg, 135 deg, 225 deg and 315 deg.

A wide variety of real (3D) objects and 2D images of objects were used as controls, including images which could be generated during the experiment using a Fairlight Computer Video Instrument. Some controls were chosen specifically to control for the possibility that a cell which appeared to be responding to a complex stimulus could actually be responding to only a simple component. For example hair or a wig was used as a control to determine whether a cell which was responding selectively (compared with other controls) to all views of a head (and which therefore appeared to show object-centred encoding) was in fact only responding to hair, a simple feature visible in any view.

Procedure for stimulus presentation

Each cell recorded was first subjected to exploratory testing involving the presentation of a large variety of static and moving visual stimuli, including in nearly every case at least four views of a static head. Cells which showed an initial tendency to discriminate at least one view of a head from control objects were then tested with several trials of at least four views of a head and various controls. In most cases this was done by using one of two standardised computer-controlled protocols.

The first protocol consisted of five trials of each of the four characteristic views of the chosen stimulus plus five trials of various control objects and five trials where the shutter opened with no stimulus present as a measure of the cell's spontaneous (baseline activity). The second protocol was identical except that five trials each of all eight views (the characteristic views plus the four intermediate views) were used. The computer program randomised the order of presentation of all the trials in both protocols. In the more recent of the experiments the program controlled the presentation of video disk stimuli and synchronised this with the collection of data. This made it possible to dispense with the shutter. If it was decided to use

real 3D heads as stimuli however, the shutter was used. In this case the computer simply instructed the experimenter which view to present for a particular trial and the experimenter activated the trial with a control box when he/she was ready. In all cases when computer-controlled protocols were used the data was analysed immediately on line. This enabled a reliable assessment of the nature of the cell's response to be made in most cases while it was still being recorded. Some cells which showed significant results at this stage were retested with different individual heads. Some cells were tested with both 3D and 2D stimuli.

Cells responding selectively to static heads which also showed a tendency to respond to movement were tested for interactions between view and direction. Only data from responses to static stimuli are reported here however.

Many cells recorded were also tested with tactile and auditory stimuli, but the results are not reported here.

Data analysis

An ANOVA was done on the data from each cell tested with four or eight views. If the ANOVA was significant ($p < 0.05$), post hoc testing was done

using the protected least significant difference (PLSD) method (multiple t tests).

For only those cells tested with eight views a linear regression (see Appendix for details) was used to estimate the angle of view of the head which would evoke the optimal response. This was done by using the following equation:

$$R = \beta_1 + \beta_2 \cos \theta + \beta_3 \sin \theta + \beta_4 \cos 2\theta + \beta_5 \sin 2\theta$$

where R is the response and θ is the angle of view. This equation was chosen because it makes few assumptions about the nature of view tuning and allows for both unimodal and bimodal patterns of response. For a cell with a single preferred view the $\sin \theta$ and $\cos \theta$ terms specify the angle of optimal response and describe a monotonic decay of response with angular deviation from optimal view. The $\sin 2\theta$ and $\cos 2\theta$ terms allow the equation to have two peaks and determine their relative amplitude, separation and sharpness.

No attempt was made to estimate the angles of optimal response of those cells tested with the four characteristic views only. This was because an initial attempt to do so (using a simpler version of the above equation) revealed the possibility that the optimal angles calculated could have been biased toward these four views. Testing with the four

intermediate views controlled for this possible source of bias.

This method of regression was able to produce two angles of optimal response in the event of cells being found which were selectively responsive to two separate views. In the case of such bimodal cells a major mode angle (where the estimated maximum response occurred) and a minor mode angle (where a lower response peak occurred) were calculated.

The regression analysis was used to produce tuning curves for all the cells tested with eight views. The half width at half height of the curve was used as a standard measure of the breadth of tuning. This was defined as the number of degrees away from the estimated angle of optimal response at which the response was equal to the average of the estimated maximum and minimum responses. This meant that there were actually two angles, one each side of the optimal response angle. The mean of these two was taken as the average half width at half height angle of the cell.

The responsiveness of the cells was calculated from the regression estimates of their maximum (R_{max}) and minimum (R_{min}) responses using the equation

$$\text{Resp} = \frac{R_{Max} - R_{Min}}{R_{Max} + R_{Min}}$$

where R_{max} and R_{Min} were measured in spikes/sec.

Cell classification

All cells tested with four or eight views were classified according to the results of ANOVA and PLSD posthoc tests.

All cells which PLSD showed to discriminate at least one view of a head from both control objects and spontaneous activity (or no stimulus) were classified as **head-selective**. All head-selective cells were if possible divided into the following three categories:-

Cells which PLSD showed to discriminate between at least two different views of the same head and to discriminate at least one view but not all views from both controls and spontaneous activity were classified as **viewer-centred**.

Cells which PLSD showed to discriminate all views from both controls and spontaneous activity but did not discriminate between any of the views were classified as **object-centred**.

Cells which both discriminated all views from controls and spontaneous activity and discriminated between any two views were classified as **object-centred/view-selective**.

Any cell in the viewer-centred and object-centred/view-selective categories was defined as

bimodal if it discriminated two separate non-adjacent views from at least two intervening views (at least one on each side of the bimodal axis).

Head-selective cells which did not fit into the three categories defined above were classified as anomalies. Cells which showed any statistical effect of head view but which could not be classified as head-selective according to the above definition were also classed as anomalies.

Each cell was classified only once. If more than one test was done on a cell and the different tests gave different classifications, the cell was classified according to the most significant result, or if there was no difference in significance the result of the first test done was used for classification.

Analysis of optimal response angles of viewer-centred cells

All cells tested with eight views which had been classified as viewer-centred using ANOVA/PLSD and also gave significant ($p < 0.05$) regression results were divided into two categories: those with angles of optimal response within 22.5 degrees either side of one of the characteristic views and those with optimal angles somewhere outside this range.

A binomial test was done on these data to determine whether the overall distribution of optimal angles was significantly clumped around the four characteristic views. Separate binomial tests were also done on the individual data for monkeys B, D and H. (No cells from monkey F were tested with eight views).

Each cell was counted only once for the analysis of the distribution of optimal angles. In cases where more than one test was done on a cell, the angle produced by the most significant regression was used, or where there was no difference, the result of the first test done was used. In the case of bimodal cells the major mode angle was used.

Estimation of earliest response latencies

The earliest latencies of the onset of firing of cells in response to head stimuli were estimated by visually examining computer printouts which showed when individual spikes occurred with reference to when the shutter opened. The point where firing became quicker than spontaneous activity (ie where the spikes became clearly closer together than before) was taken as the onset of the response. The individual trial giving the earliest response to a head was used to determine the earliest latency of a cell, except that in the case of viewer-centred

cells, only trials where the optimal view was the stimulus were considered.

RESULTS

Cell classification

A total of one hundred and ten cells were head-selective or at least showed some statistical effect of head view. Fifteen of these were classed as anomalies. Eighty-eight cells were viewer-centred, four were object-centred and three were object-centred\view-selective.

Histology

At the time of writing histological results were available for monkeys B and D.

1916 cells were recorded in the STS from monkey D (of which 1397 were in the upper bank and 519 in the lower bank) and 1300 from monkey B (570 in the upper bank and 730 in the lower bank), making a total of 3216 recorded from the STS of both monkeys.

The majority (85%) of head-sensitive cells (ie all head-selective cells plus those anomalies which showed sensitivity to varying head view without discriminating any view from controls and/or spontaneous activity) were in the upper bank of the STS. The other head-sensitive cells were all in the

lower bank of the STS except for two cells from monkey D which were located in the tail of the caudate nucleus.

Fig 1 shows the location of all head-sensitive cells located in the top bank of the right hemisphere STS of one monkey (D). Figure 1 also shows that there appears to be some clumping of head-sensitive cells in two particular locations and that there is no tendency for different views to be represented in different locations.

Viewer-centred cells

Table 1 shows a summary of the eighty-eight cells classified as viewer-centred by ANOVA/PLSD tests. It can be seen from the fourth column that three cells in monkey D and one in monkey H which gave significant ANOVA/PLSD results were not significant with regression.

#	# VIEWS TESTED	# REGRESSION P < 0.05 #
#MONKEY#		# (CELLS TESTED WITH #
#	# FOUR# EIGHT	# EIGHT VIEWS ONLY) #
# D	# 16	# 34 # 31 #
# B	# 10	# 14 # 14 #
# H	# 3	# 6 # 5 #
# F	# 5	# 0 # - #
# ALL	# 34	# 54 # 50 #

Table 1 Viewer-centred cells. The fourth column shows how many of the cells tested with eight views listed in column three gave significant regression results.

The responses of the majority of the viewer-centred cells followed a unimodal pattern, with one view evoking the optimal response. Figure 2 shows a typical example.

Six viewer-centred cells were classified as bimodal. Figure 3 shows the responses of the clearest example of a bimodal cell, with the estimated response to the minor mode angle equal to ninety percent of that to the major mode angle.

In four of the bimodal cells the two views that evoked the highest responses were 180 degrees apart. In three cases these were the profile views. One cell met the statistical criteria for classification as bimodal for the face and back of head views

because of very low responses to the profiles, but in reality seemed unimodal, since the response to the back of head view was far lower than that to the face, and the back of head view actually inhibited the cell's responses relative to the no stimulus condition. Two cells were bimodal for views only 90 degrees apart: the two "half-profile" views (45 and 315 deg). The response patterns of the last three cells mentioned were unusual and caused problems for the regression. (See next section and Figures 4 and 5).

Although only six cells could be classified as bimodal using the stringent criteria defined above, a number of other cells showed to varying degrees a non-significant trend toward bimodality. In particular with some cells selectively responsive to one profile there was a tendency for the responses to the other profile to be slightly higher than the responses to the face and back of head views. Overall there seemed to be a continuum from many purely unimodal cells at one end of the scale to a very few genuinely bimodal cells in which there was no real difference between responses to the two optimal views at the other.

The responsiveness of viewer-centred cells tested with eight views formed a continuous range from 0.25 to 1.0 in cases where the estimated minimum response was 0 spikes/sec. The mean was 0.70 ± 0.03 SEM. An

ANOVA on the separate data showed no significant differences in responsiveness between monkeys.

Limitations of the regression method used

In most cases the estimates of optimal response angles, maximum and minimum responses and breadth of tuning produced by significant regressions seemed to accurately reflect the measured responses to different views and were in line with ANOVA/PLSD results. However, for a small number of cells some of the regression estimates were clearly wrong. In most cases this only applied to the estimates of the breadth of tuning and seemed to be the result of very sharp tuning, where a small change in view caused a large change in response.

To determine the extent of this problem a chi-squared goodness of fit test was used to check that the curve produced by each significant regression was an accurate reflection of the observed data. In the event of a bad fit this test showed at what data point(s) it occurred. (This could be seen easily anyway from plot outs of the regression curve and data points produced by computer). Ten significant regression results were shown by chi-squared to be bad fits.

In five of these cases the bad fit occurred at data points far away from the (major mode) optimal

angle. There was a tendency for the regression to produce a slightly bimodal curve with a minor mode about 180 degrees away from the major mode optimal angle even when the observed data showed that the cell was purely unimodal. In fact the majority of regressions produced minor modes, although the observed responses showed that most of these cells were actually unimodal. In a few cases this tendency was enough to cause a bad fit with chi-squared. This was not a major problem since the site of the bad fit was too far away to affect either the optimal response angle calculated or the estimate of the breadth of tuning. In these cases the chi-square result was disregarded and the regression results included in the overall analyses.

A more serious problem was bad fit caused by narrow tuning close to the view that evoked the highest response. This could affect the estimate of the breadth of tuning and possibly in some cases that of the angle of optimal response. In calculating the breadth of tuning, the regression could not give a half width at half height angle for any cell of less than forty-nine degrees, even for three cells for which the real figure must have been much less. Figure 4 shows the responses of one of the most sharply tuned cells recorded (see next section). The regression estimate of the half width at half height of this cell (50 degrees) is obviously far too high. Note however that the

estimate of the optimal response angle on the other hand appears to be a very accurate reflection of the data. Figure 5 shows one of three examples shown by chi-square to be bad fits where the estimated angle of optimal response appeared to be wrong. This cell was shown by ANOVA/PLSD results to be bimodal, responding selectively to the two "half-profile" (45 and 315 deg) views, but the regression estimated the optimal response angles as being close to the profile views. This result is difficult to explain. Although this cell was very sharply tuned between the two half-profile views and the face view, to which there was virtually no response, this does not seem to account for the optimal angles calculated. One other cell showing the same pattern of response (bimodal for the 45 and 315 deg views) also gave incorrect optimal angles. This is because the equation used cannot deal with bimodal cells where the two optimal views are as close together as ninety degrees.

Tuning to different views

Figure 6 shows the range of half width at half height values for all cells tested with eight views giving significant regression results (including viewer-centred, object-centred/view-selective and anomalies). This shows that the majority of cells

had half width at half heights of between forty-five and ninety degrees. Regression estimates of half width at half height values did, however, form a continuous range from 49 deg to 121 deg. The mean was 69.0 deg +/- 2.8 SEM. (This mean value must have been biased slightly upwards by the overestimate of the breadth of tuning of a small number of cells. The real half width at half height of three cells was clearly well below forty-five degrees). Breadth of tuning seemed to form a continuum, from very sharp tuning where there was a change from the maximum response to next to no response in forty-five degrees or even less, to tuning showing a pattern which was almost that of an object-centred cell, with little difference between responses to different views and only one or two views failing to evoke higher responses than controls or the no stimulus condition. An ANOVA on the separate data showed no significant difference in breadth of tuning between monkeys. Another ANOVA showed no significant difference in breadth of tuning between cells tuned to different views. This result could also have been biased by the overestimation of a small number of values.

Very sharp tuning was always associated with the face view, either with the face as the optimal stimulus or with a sharp fall in response to the face compared with adjacent views. Three cells had half width at half values well below forty-five

degrees. Two of these responded optimally to the face view, with a sharp fall in response to the 45 deg and 315 deg views (one is shown in Figure 4), while the other was bimodal for the 45 and 315 deg views, with no response to the face (Figure 5). Three cells were significantly inhibited relative to the no stimulus condition by non-preferred views (Figure 4 is the best example). In contrast the response pattern of a very broadly-tuned cell is shown in Figure 7. For this cell all views except the right profile (270 deg) and the adjacent 315 deg view evoke higher responses than controls and no stimulus and there are no significant differences between responses to the face, the left profile and the back of the head.

Although there were a few instances of inhibition to non-preferred views in the case of sharply tuned cells, there were no examples of cells actually coding the "optimal" view by inhibition. In other words there were no cells where one or more minimal views gave significantly lower responses than no stimulus, while other views were no different to it. It was always the case that when one or more views significantly inhibited a cell, others excited it. Optimal view appeared to be always coded by excitation.

**Optimal response angles in relation to
characteristic views**

The optimal response angles of three of the fifty viewer-centred cells tested with eight views giving significant results were found to be incorrect (see above). Table 2 shows the relation of the optimal response angles of the other forty-seven cells to the four hypothetical characteristic views. The probability values given by binomial tests on the numbers of cells which responded optimally to angles within 22.5 degrees of one of the characteristic views against those that responded optimally to viewing angles outside this range are also shown.

# MONKEY # CHRC VIEW # OTHER VIEW #BINOMIAL P#						
# D # 20 # 8 # 0.018 #						
# B # 5 # 9 # 0.21 #						
# H # 3 # 2 # 0.50 #						
# ALL # 28 # 19 # 0.12 #						

Table 2 Relation of optimal response angles of viewer-centred cells to the four hypothetical characteristic views. "Chrc view" = optimal response angle of cell within 22.5 degrees of one of the characteristic views. "Other view" = optimal angle outside this range.

It can be seen from Table 2 that for the combined data there was no significant clumping of optimal response angles around the characteristic views, although there was a non-significant trend in that direction. For monkey D only, optimal response angles were significantly clumped around the characteristic views. For monkey B there was a slight non-significant trend in the opposite direction.

Figure 8 shows the distribution of the optimal response angles of all cells tested with eight views and giving significant regression results (viewer-centred, object-centred/view-selective and anomalies) from monkeys B, D and H.

Object-centred cells

Four cells were classified as object-centred, three in monkey D and one in monkey B. Two were tested with four views and two with eight. The two cells tested with eight views both gave non-significant regression results. Figure 9 shows the responses of an object-centred cell.

Object-centred/view-selective cells

There were three cells in this category, all in monkey D. Two were tested with eight views and one with four. One of the cells tested with eight views gave a significant regression, with an optimal response angle within 22.5 degrees of a characteristic view. The other regression was non-significant. Figure 10 shows the responses of one of these cells.

Anomalies

The fifteen cells classed as anomalies could be divided into three distinct types.

Ten cells were view-selective (and therefore in some way "head-sensitive") but not head-selective. ANOVA/PLSD showed that these cells discriminated between at least two views in the same way as viewer-centred cells, but unlike viewer-centred cells they did not discriminate any view from both controls and no stimulus.

Two cells were head selective but could not be placed into one of the three categories of head-selective cells defined above. These cells discriminated at least one but not all views from both controls and no stimulus, but did not discriminate between any of the views. One of these two cells nearly met the criteria for inclusion in

the object-centred category, showing no differences between responses to different views and with the responses to the least effective view significantly higher than those to no stimulus and almost higher than those to controls (PLSD: $p = 0.059$).

The third group of anomalies comprised three cells which showed an effect of head view by giving significant regression results but did not discriminate between any of the views tested according to PLSD results. These cells did not meet the criteria to be classified as viewer-centred but did seem to be view-sensitive in some way.

Discrimination between different individual heads

Fifteen of the cells which were classified as viewer-centred, object-centred or object-centred/view-selective had been tested with an equivalent test using more than one individual head. (An equivalent test means that nothing was varied apart from identity. This does not include, for example, instances where one test was on a live head and the other on a 2D head). Five of these fifteen cells gave significantly (PLSD) different responses to two or more individuals in at least one view.

It is not, however, possible to claim that all these five cells were definitely identity sensitive.

This is because when two or more individual heads were used, they were nearly always tested separately one after the other. This obviously meant that there could have been a general reduction or change in the cell's response between tests, so any apparent identity difference may not have been real. The more effective method of randomly interspersing the trials of two heads in one test was not used because the large number of trials involved (100) brought the very real risk of the cell being lost before the test was complete. Given that testing for identity sensitivity was not a major aim of this study, it was not considered worthwhile to risk losing valuable data in this way.

For two viewer-centred cells identity sensitivity consisted of significant differences between the responses to the optimal view only of two different heads. The optimal views were the same for both individuals. An example is shown in Figure 11. This cell showed the same pattern of view-selectivity for two experimenters, ST and RB, but in the optimal view responses to ST were significantly higher than those to RB. A two-way ANOVA on view against identity gave significant effects for both view and identity, but no significant interaction. For this cell further testing was done using interspersed trials of only the optimal view of the two experimenters, which confirmed significant identity discrimination.

Another form of identity discrimination was shown by two cells which responded to one view of one head but not at all to any view of a second head. The responses of one of these cells are shown in Figure 12. Responses to one experimenter (RB) showed the typical pattern of a viewer-centred cell, with just one of the four views tested discriminated from controls. But the cell failed to distinguish any view of the head of a second experimenter (MH) from control objects. A more dramatic example of a cell which responds to one individual but not another is shown in Figure 13. This was one of the four cells classed as object-centred. Responses to four views of JH followed a classic object-centred pattern: all were much greater than those to controls and no stimulus and there were no significant differences between them. However, the cell showed no sign at all of "recognising" another individual (DP), with no view evoking responses which were even slightly higher than controls and no stimulus.

Reliability of estimated optimal response angles

Seven viewer-centred cells tested with eight views were tested with two individual heads and gave significant regression results for both tests. This includes two cells where one test was on a live (3D) head and the other on a 2D head and five cells where

the two tests were of the same type. The optimal response angles produced by the two tests were compared to determine to what extent the results generalized between heads. Differences in optimal angle between the two tests for these seven cells ranged from 0 deg to 16 deg, with a mean difference of $6.9 +/ - 2.1$ deg. The differences in optimal angle between tests were no different for the two cells that were tested with one 2D and one 3D head from where the same type of test was used. These data indicate a high level of reliability for the method used.

Earliest response latencies

Some latency data were lost because of computer malfunction. Data are available for sixty-seven viewer-centred cells, two object-centred cells and three object-centred/view-selective cells.

The mean earliest latency for viewer-centred cells was $139.0 +/ - 5.6$ ms, for object-centred cells $135 +/ - 35.0$ ms and for object-centred/view-selective cells $116.7 +/ - 8.8$ ms. T tests showed no significant differences between any of these means.

DISCUSSION

Histology

All but two of the head-sensitive cells from monkeys D and B were located in the STS. This allows direct comparison of these results with the results of earlier studies of STS cells (eg Hasselmo et al 1989, Perrett et al 1985a).

Figure 1 shows evidence that head-sensitive cells in monkey D were organized into patches, with one large and one smaller clump separated by 3 mm. (The figure also shows that electrode tracks were relatively even throughout the area and that these patches were not artefacts of testing bias). There was, however, no evidence of more complex organization than this general clumping. There was, for example, no apparent tendency for different views to be represented at different locations. Instead a full range of views was found to be analysed within one patch. It is of course possible that there is a more microscopic level of anatomical organization where cells coding different views are spatially separated within patches of head-sensitive cells, but testing this hypothesis would require

methods of much greater resolution than those employed here.

Predominance of viewer-centred encoding

The results show the importance of viewer-centred encoding in the STS, with the vast majority of cells recorded being strongly view-selective. This is in line with the results of the study of Perrett et al (1985a) using similar static stimuli and other results with moving head/body stimuli (Perrett et al 1985b, 1989, 1990).

The rarity of cells showing fully object-centred properties is striking and contrasts with the results reported by Hasselmo et al (1989). However, this could well be due to the different methods used to classify cells. As mentioned above, the method of classification adopted by Hasselmo et al did not require a cell to be either insensitive to view or for every view tested to elicit higher responses than other stimuli. The method of classification of object-centred cells used here accords with the model of object-centred encoding proposed by Marr and Nishihara (1978), while Hasselmo's definition does not. The vast majority of cells classified by Hasselmo et al (1989) as object-centred were clearly view-selective using the authors' own analysis (and so failed to meet Marr's criterion of view-

insensitivity for object-centred encoding). The most that could be argued is that these cells showed both object-centred and viewer-centred properties. It also seems likely that many of these cells would not have been classified as object-centred at all if the method of classification used in this study were applied to the data.

It could of course be argued that the methods used here were too stringent and that some cells which showed a very large degree of response generalisation across view could more realistically be regarded as object-centred than viewer-centred. Even taking this argument into account, however (and it would only affect the classification of a small number of very broadly tuned cells) viewer-centred encoding was clearly dominant among the sample of STS cells recorded in this study.

It is, of course, possible that the sample of cells recorded was not representative of the STS as a whole. This seems unlikely in view of the fact that previous studies of STS neurons have produced similar results (Perrett et al 1985a, 1985b, 1987, 1989, 1990). These studies combined should have examined a fair sample of STS cells. It is also a possibility that object-centred encoding is more common in other brain areas, although there is as yet no evidence for this.

A noticeable feature of the sample of viewer-centred cells recorded was the very small number of

genuinely bimodal cells. This contrasts with a previous study in which the majority of cells selective for profile views responded to both profiles rather than one (Perrett et al 1985a).

The response properties generally shown by the cells recorded were similar to those shown by cells recorded by Perrett et al (1985a). These included generalization of response to quite different (eg both human and monkey) heads in the same view, to heads at different distances, and to both 2D and 3D heads. Perrett et al note that these properties do not conform with what would be expected from a very limited low-level form of viewer-centred representation such as Marr's 2.5D sketch. The degree of generalisation for a certain class of objects which is shown across many kinds of large change in the retinal image shows that viewer-centred cells in the STS represent a high-level stage in the recognition process.

Advantages of a viewer-centred system

There are various theoretical advantages in a system of recognition using high-level viewer-centred encoding. For example it would be useful for an animal in some circumstances to recognise the view of an object at the same time as recognising its category. For instance in avoiding predators -

probably the first thing you would need to know about a predator as soon as you recognised it for what it was would be whether or not it was facing toward you. There are also many forms of interaction between individual animals of social species such as primates where quick recognition of view is important. Situations such as this suggest a role for viewer-centred cells in the coding of the direction of another animal's attention. If this were the case, then it seems likely that direction of eye gaze would also play a role in the responses of these cells. That direction of gaze can indeed sometimes play such a role in activating or inhibiting STS cells, sometimes interacting with head view, sometimes independent of it, has been shown by previous work (Perrett et al 1985a).

Since direction of gaze would seem to be a more reliable indication than head view of the direction of another animal's attention it would be a worthwhile project to investigate whether some viewer-centred cells could code for direction of attention by using either head view or gaze direction, but with gaze direction overriding head view if the two were not consistent. There is evidence to support this idea (Perrett et al 1985a). For example a viewer-centred cell coding for direction of attention responded to the face view both with the eyes looking directly forward and with the eyes completely covered (responding to some

global aspect of head view) but the response was abolished with the eyes visible and averted 45 degrees to one side, (which would provide evidence that attention was not directed toward the observer). A similar cell might theoretically also respond to the eyes looking directly at the subject (showing direction of attention toward the observer), but with the head at 45 degrees.

Object-centred encoding

To emphasise viewer-centred encoding and to regard it as a high-level form of recognition in its own right is not to say that a further stage of object-centred processing does not then take place. Indeed, there seem good reasons to believe that it does, both for the logical a priori reasons suggested above and because a small number of cells with such object-centred properties were indeed found in this study.

It would seem reasonable to postulate that the properties of these object-centred cells derive from the combination of inputs from several viewer-centred cells. The fact that there were far fewer object-centred cells than viewer-centred ones in the sample recorded in this study is consistent with this. This is, however, only a conjecture and the limited data available from cell response latencies does not support it.

If object-centred cells do only fire when they have received inputs from viewer-centred cells, then obviously object-centred cells will on average take longer to begin to respond than viewer-centred cells. There was, however, no significant difference between the response latencies of viewer-centred and object-centred cells in this study. In fact if the three cells showing both object- and viewer-centred properties are regarded for this purpose as object-centred, there is a non-significant trend for object-centred cells to respond faster than viewer-centred ones.

These data present a problem for any interpretation of the results in terms of discrete viewer- and object-centred stages of processing. It should be noted, however, that the sample of object-centred cells was small and the comparison of the response latencies of the two categories cannot be said to have provided conclusive results. If we assume only one synapse is necessary to form an object-centred cell from several viewer-centred inputs, the difference in latencies would obviously be small and a large sample of cells in both categories would probably be required to show a statistically significant difference between average latencies.

Breadth of tuning

Most viewer-centred cells tolerated a considerable change in view, as the mean half width at half height value (69 deg) showed. This meant that only a few views would have to be encoded to cover the whole range, although there was no evidence that this was the case (see below).

The most interesting fact to emerge about tuning was its range, with the half width at half height varying from well below thirty degrees, as in the example shown in Fig 4 to over one hundred and twenty degrees. It seems a general principle of sensory coding that breadth of tuning of different neurons to particular classes of stimuli forms a continuum from those which are very sharply tuned and specific to a particular stimulus to those which give very general responses, and encoding of head view is clearly no exception to this principle.

Although most cells recorded were successfully classified according to a fairly rigid system, it seems clear that in reality many cells are not definitely one thing or the other. This was shown firstly by the existence of the object-centred/view-selective category of cells. These are truly object-centred, in that actual recognition of the head clearly does not depend at all on viewpoint and yet are definitely not "view-insensitive" - a term often used almost as a synonym for object-centred. These

cells may represent a half way stage between the two forms of encoding (cf Hasselmo et al 1989). Perrett et al (1989) compare the finding of these cells with an analogous discovery of neurons showing the property of end stopping which led to the modification of the original simple hierarchical model of simple, complex and hypercomplex cells in the visual cortex proposed by Hubel and Wiesel (1968). The existence of these cells shows that when dealing with biological systems concepts such as viewer-centred and object-centred are difficult to apply as neat and precisely definable categories like those of computational and mathematical models (eg Marr 1982).

Even within the viewer-centred category itself there were problems of classification. It is difficult to believe there is not a qualitative difference between a dramatically viewer-centred cell such as that in Figure 4 and the broad-tuned cell shown in Figure 7. The latter cannot be classed as object-centred according to the method used here, and certainly not according to the scheme of Marr (1982), which required an object-centred description to be the same for all vantage points of the viewer. For some views this cell was clearly view-selective, and there was a range of view it didn't "recognise" at all, so it should be classed as viewer-centred. And yet the responses of this cell were the same to the face as to the back of the head. How could a

system of encoding based purely on viewpoint account for this degree of response generalization? One possibility is that this cell is combining inputs from more sharply-tuned viewer-centred cells, and could (like the cells in the object-centred\view-selective category) be regarded as representing some kind of half-way stage between purely viewer-centred and object-centred stages of processing. Perhaps the only way to describe such a response pattern is as object-centred, but only for a particular range of views, but this makes little sense in terms of a classification system based on mathematical models such as Marr's. Perhaps such cells can be regarded as 'failed' object-centred cells which have not combined enough viewer-centred inputs to cover the whole range of view. Such a cell could combine its output with that of another (possibly narrow-tuned) viewer-centred cell to create a truly object-centred cell. As noted above, however, the response latency data provides no support for the idea that cells do combine in this way.

Taken as a whole the evidence from examining breadth of tuning shows that heads may be encoded in a way that varies continuously from being extremely specific for one view to purely object-centred, with no rigid borders between different categories of cell, and that some cells in the middle of the continuum show a mixture of both types of encoding.

Are only characteristic views encoded?

The distribution of the angles of optimal response of viewer-centred cells provides strong evidence against the hypothesis that only the four characteristic views hypothesised above are encoded in the brain. In complete contrast to the findings of Perrett et al (1985a), who found no cells responsive to views other than these views, in this study many cells were found which responded optimally to angles closer to the four intermediate views than to the four so-called characteristic views. This supports the findings of Hasselmo et al (1989).

There was a significant clumping of optimal angles around the characteristic views in one monkey only, but not for the combined data.

However, the fact that the tuning of monkey D's cells was significantly clumped around the characteristic views, and that the trend for the combined data was in this direction indicated that there may be something special about these views, even if all cells are not tuned to them. More data could be expected to show significant clumping.

There is also evidence of individual differences between monkeys (B and D) as to whether clumping around certain views occurs or not.

Even if the trend toward clumping around the characteristic views did reach significance for the combined data, this would not support the claim that only the four hypothesised characteristic views are encoded. This claim would require that no cells were tuned to respond maximally to intermediate views, just as no cone cells in the retina are tuned to intermediate wavelengths. Significant clumping may, however, be able to account for some of the psychological data mentioned above, if we assume some kind of additive process. But as long as some cells are tuned to respond optimally outside their range, the hypothesis that recognition depends on the four hypothetical characteristic views is untenable.

Of course the fact that cells are not tuned to only the four views in question does not rule out the possibility that there are in fact characteristic views, but there are more than four of them and/or they are different to the ones suggested. The data reported here provide no clues as to what such views could be however. Fig 8 shows that distribution of angles of optimal response follows a very wide range.

Identity sensitivity

Examples of both purely viewer-centred and purely object-centred cells were found which gave significantly different responses to different individuals. These results are not as easy to interpret as those relating to responses to heads in general as a category. It is often tempting to read more into the results than is supported by the evidence. For none of these cells is there any evidence that they are actually recognizing a particular individual. An example is the cell shown in Fig 12 , which responds to one view of RB but none of MH. This result tells us nothing except that one view of RB can be distinguished from control objects but that no view of MH can and that in this view RB can be discriminated from MH. We have no evidence that this cell is involved in recognizing RB as a familiar individual. To show that would have needed a large number of different heads. In fact all the cells which showed identity effects were tested with a very small number of heads, usually only two, so all we can ever say about the results is that they show identity-sensitivity, never that they are selective for a given individual.

It is in fact dubious that responses specific for one individual could in principle be shown by a single cell. The old idea of grandmother cell encoding, defended most vigourously in recent years

by Barlow (1972, 1985) states that one cell could code the identity of a specific individual. Nobody seriously suggests that there is literally only one cell in the whole brain coding each individual. Rather there is presumed to be an enormous amount of redundancy but only one cell would actually be necessary. Konorski (1967) supported this idea, and suggested that the number of cells coding an individual would depend on factors such as level of familiarity and importance.

In contrast to the grandmother cell model, which has declined in influence, models of individual recognition based on distributed processing have been claimed to provide a more credible explanation of how individual cells selective for heads as a category may combine in a network which could recognize an individual. For example Rolls (1987) has put forward a model of a recognition system using ensemble coding which he claims can explain a pattern of single cell responses to different faces which would be problematical for a grandmother cell model.

ACKNOWLEDGEMENTS

This research was supported by grants from the SERC Image Interpretation Initiative (GR/E 43881), the Japanese New Energy and Industrial Technology Development Organization (NEDO) and the ESRC (XC15250005). Personal finance was provided by an SERC earmarked award linked to the Image Interpretation Initiative. Dr P Jupp of the Department of Mathematics at St Andrews University devised the method of linear regression used.

Thanks also to the following:

Dave Perrett, my supervisor for some useful, if over-aggressive, comments on previous drafts of this work.

Phil Winn, my second supervisor, for intervening in a moment of severe crisis, for much valuable and fair-minded advice and for persuading me to complete this thesis.

Mark Harries, for a vast amount of friendly, patient and invaluable help, advice and encouragement and for generally making life tolerable and completion of this thesis possible.

Sharon Thomas, for support and encouragement, for providing much-needed humour and for sharing the many deep humiliations of being a research student.

Phil Benson, Mike Oram and Jari Hietanen for help with many aspects of the work.

REFERENCES

Barlow H B (1972). Single units and sensation: A neuron doctrine for perceptual psychology.
Perception 1, 371-394.

Barlow H B (1985). The twelfth Bartlett Memorial Lecture: The role of single neurons in the psychology of perception. *Q J Psychol* 37A, 121-145.

Baylis G C, Rolls E T, Leonard C M (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res* 342: 91-102.

Biederman I (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review* 94, 115-147.

Binford T O (1971). Visual perception by computer. *Presented to the IEEE Conference on Systems and Control, Miami.*

Bolles R C and Cain R A (1982). Recognizing and locating partially visible objects: The local-feature-focus method. *Internat J Robotics Res*, 1, 57-82.

Brooks R (1981). Symbolic reasoning among 3-dimensional models and 2-dimensional images. *Artificial Intelligence* 17, 385-349.

Bruce C J, Desimone R and Gross C G (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46, 369-384.

Bruce V and Valentine T (1987). The basis of the three-quarter view advantage in face recognition. *Applied Cognitive Psychology* 1, 109-120.

Callahan J and Weiss R (1985). A model for describing surface shape. In *Proceedings of IEEE Conference of Computer Vision and Pattern Recognition*. New York: IEEE Computer Society.

Chakravarty I and Freeman H (1982). The use of characteristic views as a basis for recognition of three dimensional objects. *SPIE vol 336: Robot Vision* 37-45.

Desimone R, Albright T D, Gross C G, and Bruce C (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 8, 2051-2062.

Fagan J F (1979). The origins of facial pattern recognition. In M H Bornstein and W Kessen (eds) *Psychological Development from Infancy: Image to Intention*. Hillsdale N J Erlbaum (83-113).

Feldman J A (1985). Four frames suffice: A provisional model of vision and space. *Behav Brain Sci* 8, 265-289.

Gibson J J (1950). *The perception of the visual world*. Houghton Mifflin, Boston.

Gibson J J (1979). *The ecological approach to visual perception*. Houghton Mifflin, Boston.

Gray M (1986). Recognition Planning from solid models. *Proceedings of the Alvey Computer Vision and Image Interpretation Meeting, Bristol, September 1986*, pp 41-43. Sheffield: Sheffield University Press.

Gross C G, Bruce, C J, Desimone R, Fleming J and Gatass R (1982). Cortical visual areas of the temporal lobe. In *Cortical sensory organization* (ed C N Woolsey), vol 2 (*Multiple visual areas*), 187-216. Clifton New Jersey: Humana Press.

- Harries M H, Perrett D I and Lavender A (1990). Visual inspection during encoding and recognition of 3D heads. *Perception* in press.
- Hasselmo M E, Rolls E T, Baylis G C, and Nalwa V (1989). Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp Brain Res* 75, 417-429.
- Hoffman D and Richards W (1986). Parts of recognition. In A P Pentland (Ed) *From pixels to predicates*. Ablex, Norwood, NJ.
- Hopfield J J (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc Nat Academy of Sciences, USA*, 79 2554-2558.
- Hubel D H and Wiesel T N (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, Lond* 195, 215-243.
- Huberman B A and Hogg T (1984). Adaptation and self-repair in parallel computing structures. *Physics Review Letters* 52, 1048-1051.
- Humphreys G W and Riddoch M J (1984). Routes to object constancy: implications from neurological

impairments of object constancy. *Quart J Exp Psychol* 36A, 385-415.

Ikeukchi K (1987). Precompiling a geometrical model into an interpretation tree for object recognition in a bin-picking task. *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, 321-339.

Jolicoeur P (1985). The time to name disoriented natural objects. *Memory and Cognition* 13, 289-303.

Koenderinck J J and van Doorn A J (1976). The singularities of the visual mapping. *Biological Cybernetics* 24, 51-59.

Koenderinck J J and van Doorn A J (1977). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In *Kybernetik 1977* (eds G Hauske and E Butenandt), 224-47. Oldenburg.

Koenderinck J J and van Doorn A J (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics* 32, 211-216.

Kohonen T (1978). *Associative Memory: A system theoretical approach*. Springer Verlag, Berlin.

Konorski J (1967). *Integrative activity of the brain*. Chicago, Illinois: University of Chicago Press.

Krouse F (1981). Effects of pose, pose change and delay on face recognition performance. *Journal of Applied Psychology* 66, 651-654.

Logie R H, Baddeley A D and Woodhead M M (1987). Face recognition; pose and ecological validity. *Applied Cognitive Psychology* 1, 53-69.

Lowe D (1985). *Perceptual Organization and Visual Recognition*. Boston, Kluwer Academic Publishers.

Lowe D (1987). Three dimensional object recognition from single two-dimensional images. *Artificial Intelligence* 31, 355-395.

Marr D (1977). Analysis of occluding contour. *Proc R Soc London B*, 197, 441-475.

Marr D (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W H Freeman.

Marr D and Nishihara H K (1978). Representation of the spatial organization of three dimensional shapes. *Proc R Soc B* 200, 269-294.

Marr D and Vaina L (1982). Representation and recognition of the movements of shapes. *Proc R Soc (London)*, B, 214, 501-524.

Maturana H R, Letvin J Y, McCulloch W S and Pitts W H (1960). Anatomy and physiology of vision in the frog (*Rana pipiens*). *J Gen Physiol* 43, 129-175.

Merrill, E.G., and Ainsworth, A. (1972) Glass coated platinum-plated tungsten microelectrodes. *Med. biol. Eng.* 10:662-672.

Mikami A and Nakamura K (1988). Selective neuronal activities in the superior temporal sulcus of the macaque monkey. *Soc Neurosci Abstr* 8, 4.

Palmer S, Rosch E and Chase P (1981). Canonical perspective and the perception of objects. In J Long and A Baddeley (Eds) *Attention and Performance*. Hillsdale NJ, Erlbaum.

Pandya D and Yeterian E H (1985). Architecture and connections of cortical association areas. In *Cerebral Cortex*, vol 4 (eds Peters and Jones), 3-64. New York: Plenum Press.

Pentland A (1986). Perceptual organisation and representation of natural form. In M A Fisculer, O

Firschein (Eds) *Readings in Computer Vision - Issues, Problems, Principles and Paradigms*. Los Altos, CA: Morgan Kaufman.

Perrett D I and Harries M H (1988). Characteristic views and the visual inspection of simple smooth and faceted objects: 'tetrahedra and potatoes'.

Perception 17, 703-720.

Perrett D I, Harries M H, Benson P J, Chitty A J and Mistlin A J (1990). Retrieval of structure from rigid and biological motion: an analysis of the visual responses of neurones in the macaque temporal cortex. In A Blake and T Troscianko (Eds) *AI and the Eye*. John Wiley and Sons.

Perrett D I, Harries M H, Bevan R, Thomas S, Benson P J, Mistlin A J, Chitty A J, Hietanen J K, and Ortega J E (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J Exp Biol* 146, 87-114.

Perrett D I, Harries M H, Mistlin A J and Chitty A J (1989). Three stages in the classification of body movements by visual neurons. In H Barlow, C Blakemore and M Weston-Smith (Eds) *Images and Understanding*. Cambridge University Press, Cambridge.

- Perrett D I, Mistlin A J and Chitty A J (1987). Visual cells responsive to faces. *Trends Neurosci* 10, 358-364.
- Perrett D I, Mistlin A J, Chitty A J, Harries M H, Newcombe F and De Haan E (1988). In C Kennard and F Clifford Rose (Eds) *Physiological Aspects of Clinical Neuro-Ophthalmology*. Chapman and Hall, London.
- Perrett D I and Rolls E T (1983). Neural mechanisms underlying the visual analysis of faces. In *Advances in vertebrate neuroethology* (ed J P Ewert, R R Caprinica and D J Ingle), 543-566. New York: Plenum Press.
- Perrett D I, Rolls E T and Caan W (1982). 'Visual neurones responsive to faces in the monkey temporal cortex'. *Exp Brain Res* 47, 329-342.
- Perrett D I, Smith P A J, Mistlin A J, Chitty A J, Head A S, Potter D D, Broennimann R, Milner A D and Jeeves M A (1985b). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey. A preliminary report. *Behav Brain Res* 16, 153-170.
- Perrett D I, Smith P A J, Potter D D, Mistlin A J, Head A S, Milner A D, Jeeves M A (1985a). Visual

cells in the temporal cortex sensitive to face view and gaze direction. *Proc R Soc B* 223, 293-317.

Pitts W and McCulloch W S (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics* 9, 127-147.

Rieger J H (1986). Families of views of piecewise smooth objects. *Proceedings of the Alvey Computer Vision and Image Interpretation Meeting, Bristol, 1986*. Sheffield: Sheffield University Press.

Rock I and DiVita J (1987). A case of viewer-centred object perception. *Cognitive Psychology* 19, 280-293.

Rolls E T (1987). Information representation, processing, and storage in the brain: analysis at the single neuron level. In J P Changeux and M Konishi (eds). *The Neural and Molecular Base of Learning*, 503-540.

Rolls ET (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum Neurobiol* 3: 209-222.

Rosch E, Mervis C B, Gray W D, Johnson D M, and Boyes-Graem P (1976). Basic objects in natural categories. *Cog Psychol* 8, 382-439.

Seibert M and Waxman A M (1990). Learned aspect graph representations from view sequences. In D S Touretzky (Ed) *Advances in Neural Network Information Processing Systems, vol 2*. San Mateo, CA: Morgan Kaufman (in press).

Seltzer B and Pandya D N (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149, 1-24.

Snider R S and Lee, J C (1961). *A Stereotaxic atlas of the monkey brain (Macaca mulatta)*. Chicago University Press, Chicago.

Thorpe C and Shafer S (1983). Correspondence in line drawings of multiple views of objects. In *Proceedings of the 8th International Conference on Artificial Intelligence, vol 2*, 959-965.

Ullman S (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition* 32, 193-254.

Ullman S and Basri (1989). Recognition by linear combinations of models. *AI Memo No 1152.*

Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Warrington E K and James M (1986). Visual object recognition in patients with right-hemisphere lesions: axes or features? *Perception* 15, 366-366.

Warrington E K and Taylor A M (1973). The contribution of the right parietal lobe to object recognition. *Cortex* 9, 152-164.

Willshaw D J, Bruneman O P and Longuet-Higgins H C (1969). Non-holographic associative memory. *Nature* 222, 960-962.

Woodhead M M and Baddeley A D (1979). The choice of poses for faces photographed for subsequent recognition. Report prepared for the APRC 79/17.

FIGURE LEGENDS**Figure 1**

Serial sections (every 1.0 mm) of the upper bank of the STS from 3 mm posterior to 9 mm anterior to the mid geniculate level from the right hemisphere of one monkey (D). Left column indicates the position of all recording tracks. Columns 2-5 indicate the position of cells responsive to different views. The preferred angle of view for each cell (to the nearest 23 degrees) is indicated by the direction of the 8 arrow types. Column 2: face and back of heads views (up and down arrows); column 4: left and right profile views; columns 3 and 5: intermediate views. This figure includes all cells in any category which showed significant view selectivity (whether with regression or ANOVA) tested with eight or four views.

Figure 2

The responses of a typical viewer-centred cell to eight views of a head. The y axis of the graph measures response in spikes/sec. The short horizontal lines represent mean responses to the different views +/- 1 SEM. The curve is the

regression second order cardioid best fit. The horizontal dashed lines show the level of response to the control and no stimulus or spontaneous activity (SA) conditions. The angles refer to the eight head views tested. For this cell, responses to the back of head (180 deg) and 135 deg view were higher than those to all other conditions (PLSD $p < 0.05$). ANOVA: $F = 20.3$, $DF = 9, 40$, $p < 0.001$.

Figure 3

The responses of a bimodal viewer-centred cell. Responses to the two profile views were higher than to all other conditions except the 225 deg view (PLSD $p < 0.05$). ANOVA: $F = 4.8$, $DF = 9, 40$, $p < 0.001$.

Figure 4

The responses of a viewer-centred cell with very narrow tuning around the optimal view. The face view evoked far higher responses than any other condition (PLSD $p < 0.001$). All other views and controls evoked lower responses than no stimulus (PLSD $p < 0.05$). The cell was bimodal, with higher responses to both the face and the back of head view than to both profiles (PLSD $p < 0.05$). Thus despite the fact

that the back of head view actually inhibited the cell's spontaneous activity, inhibition to the profiles was significantly greater. Note that the regression curve cannot fit the narrow tuning of this cell, although the optimal response angle (359.7 deg) is clearly appropriate. ANOVA: $F = 53.5$, $DF = 9, 27$, $p < 0.0001$.

Figure 5

The responses of a viewer-centred cell showing a bimodal response to the 45 and 315 deg "half-profile" views but also discriminating between them. The 315 deg view evoked higher responses than all other conditions, including 45 deg (PLSD $p < 0.05$), which in turn evoked higher responses than all other conditions except 270 (PLSD $p < 0.05$). Tuning was very narrow, with 45 deg and 315 deg the optimal response views, but virtually no response to the intervening 0 deg (face) view. The face view did not appear to actually inhibit the cell however, since there was no spontaneous activity: all trials of controls and no stimulus failed to elicit any response above 0 spikes/sec. Note that the regression curve for this cell is an extremely bad fit: the data points for the 45, 315 and 0 views are all well outside it. ANOVA: $P = 14.5$, $DF = 9, 38$, $p < 0.001$.

Figure 6

Breadth of Tuning - range of half width at half height values produced by regression for all cells tested with eight views and giving significant regression results.

Figure 7

The responses of a broadly tuned viewer-centred cell. There were no significant differences between responses to the face, 45 deg, left profile, 135 deg and back of head views. All views evoked higher responses than controls and no stimulus except for the right profile and 315 views (PLSD $p < 0.05$). The regression curve is a good fit. ANOVA: $F = 9.9$, $DF = 4, 35$, $p < 0.001$).

Figure 8

The distribution of estimated angles of optimal response of all cells tested with eight views and giving significant regression results (including viewer-centred, object-centred/view-selective and anomalies). Each line represents the major mode

optimal response angle of one cell. No cell is counted more than once.

Figure 9

The responses of an object-centred cell. There are no difference between responses to any views and all views elicit higher responses than both controls and no stimulus (PLSD $p < 0.001$). ANOVA: $F = 14.6$, $DF = 9,40$, $p < 0.001$.

Figure 10

The responses of an object-centred\view-selective cell. Responses to all views were higher than to controls and no stimulus ($p < 0.05$). The face view evoked higher responses than the 315, 135, 270 and 90 deg views (PLSD $p < 0.05$). ANOVA: $F = 8.8$, $DF = 9,40$, $p < 0.001$.

Figure 11

The responses of a viewer-centred cell sensitive to identity. For the heads of both ST and RB the back of head view (180 deg) evoked higher responses than the face view (PLSD $p < 0.001$). Responses to the

back of ST's head were higher than those to the same view of RB ($p < 0.05$). ANOVA: $F = 14.2$, $DF = 9,55$, $p < 0.001$. A two-way ANOVA of view against identity showed a significant effect of view ($F = 15.9$, $DF = 3,43$, $p < 0.01$) and a significant effect of identity ($F = 6.6$, $DF = 1,43$, $p < 0.05$). There was no significant interaction ($p = 1.0$).

Figure 12

The responses of an identity-selective cell which was viewer-centred for one head, but not head-selective at all for another. The 90 deg (left profile) of RB evoked higher responses than the other views, all views of MH and controls (PLSD $p < 0.05$). There were no other differences between any conditions, with no view of MH evoking higher responses than controls. ANOVA: $F = 3.2$, $DF = 8,34$, $p < 0.01$. A two-way ANOVA showed a significant effect of view ($F = 5.0$, $DF = 3,32$, $p < 0.01$) and a significant interaction effect ($F = 3.0$, $DF = 1,32$, $p < 0.05$). The identity effect, however, did not reach significance.

Fig 13

The responses of an identity-sensitive object-centred cell. Responses to four views of JH were not different from each other, but all were significantly higher than those to controls and all views of another experimenter, DP (PLSD $p < 0.001$ each comparison). No view of DP elicited responses greater than controls or spontaneous activity. 1-way ANOVA: $F = 17.4$, $DF = 9, 57$, $p < 0.001$. A 2-way ANOVA showed a significant effect of identity ($F = 112.7$, $df = 1, 47$, $p < 0.001$) but no significant effect of view or interaction effect.

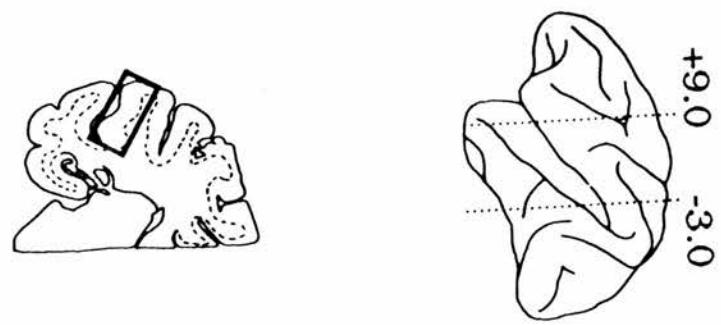
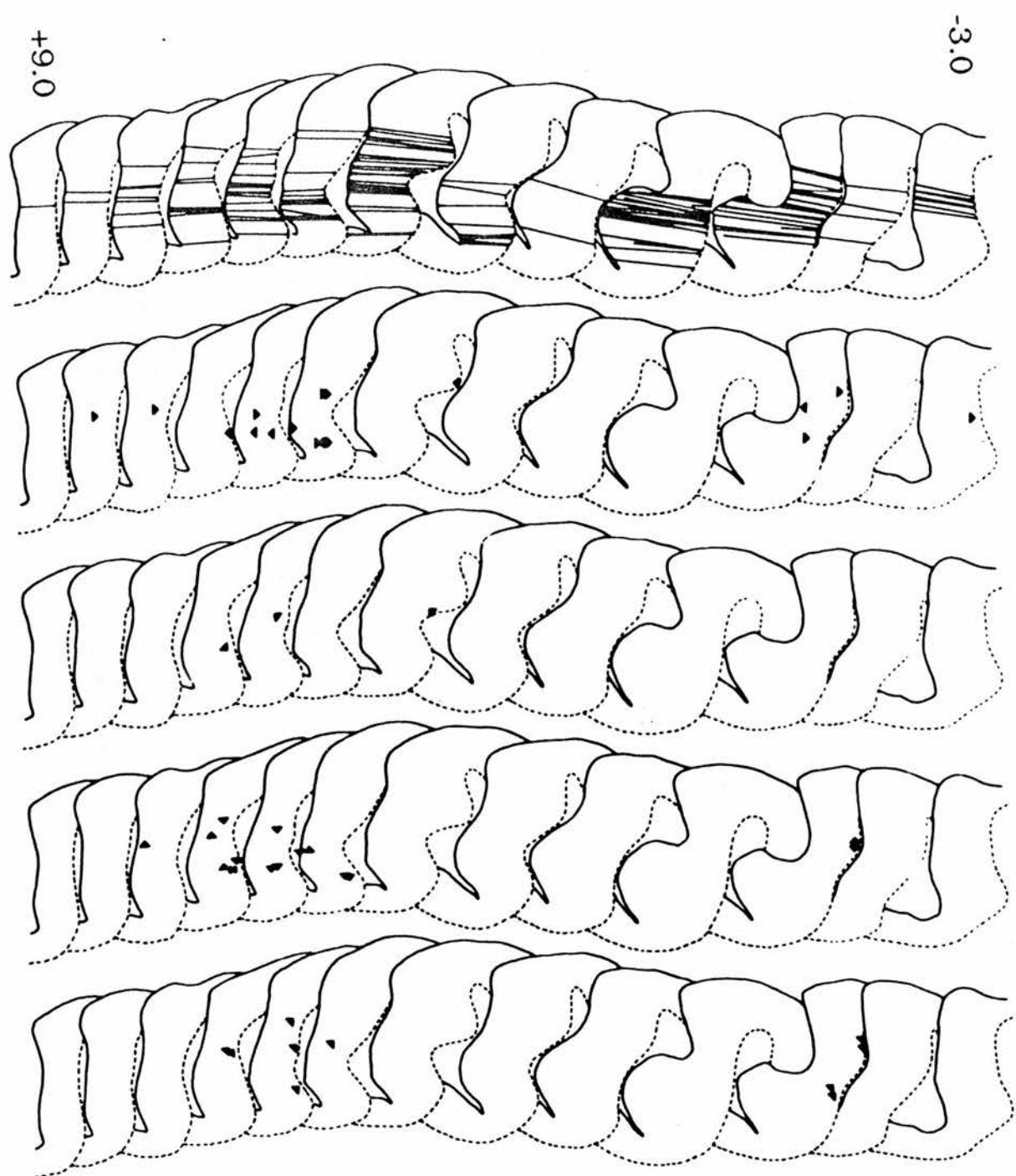


FIG 1

FIG 2

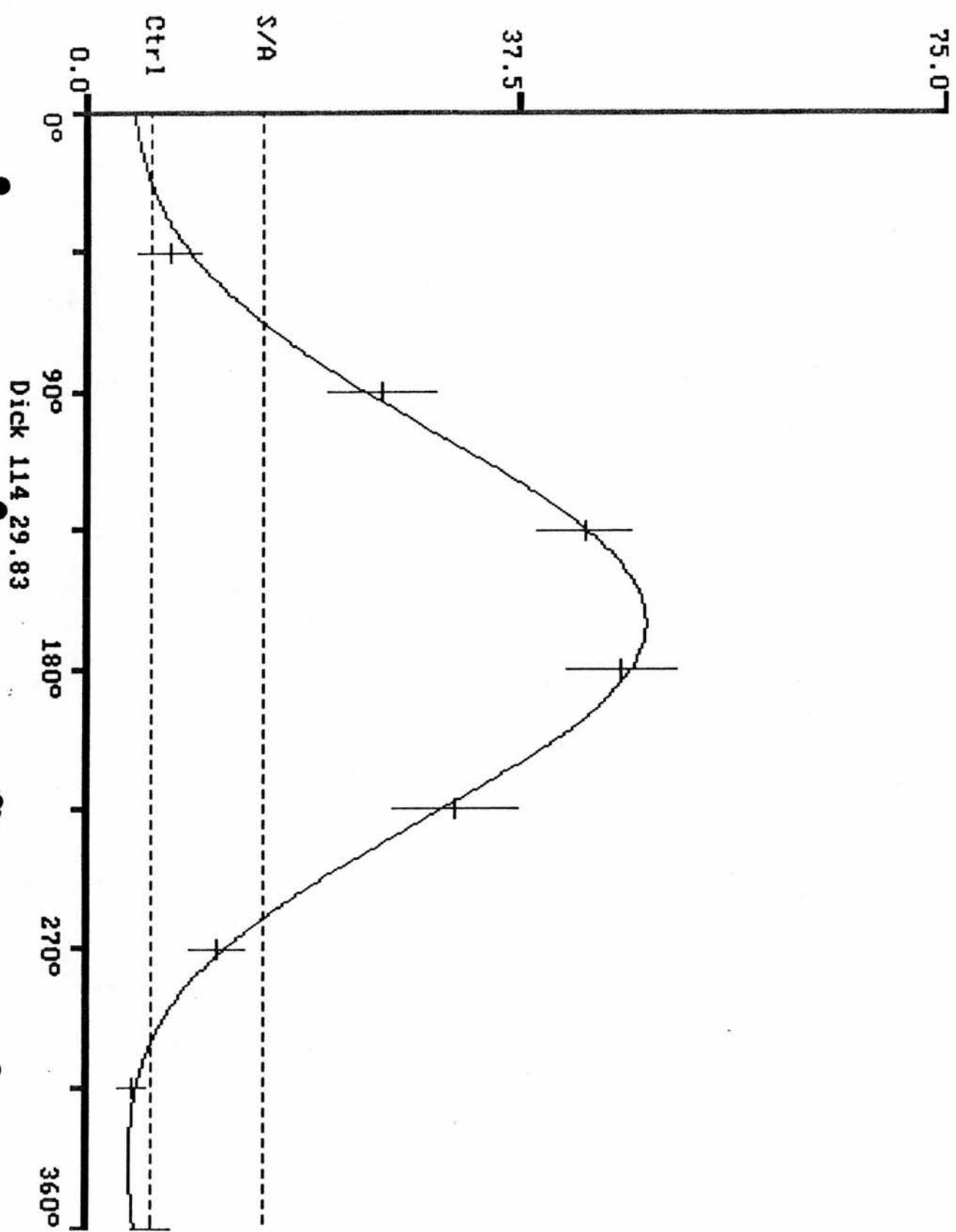


FIG 3

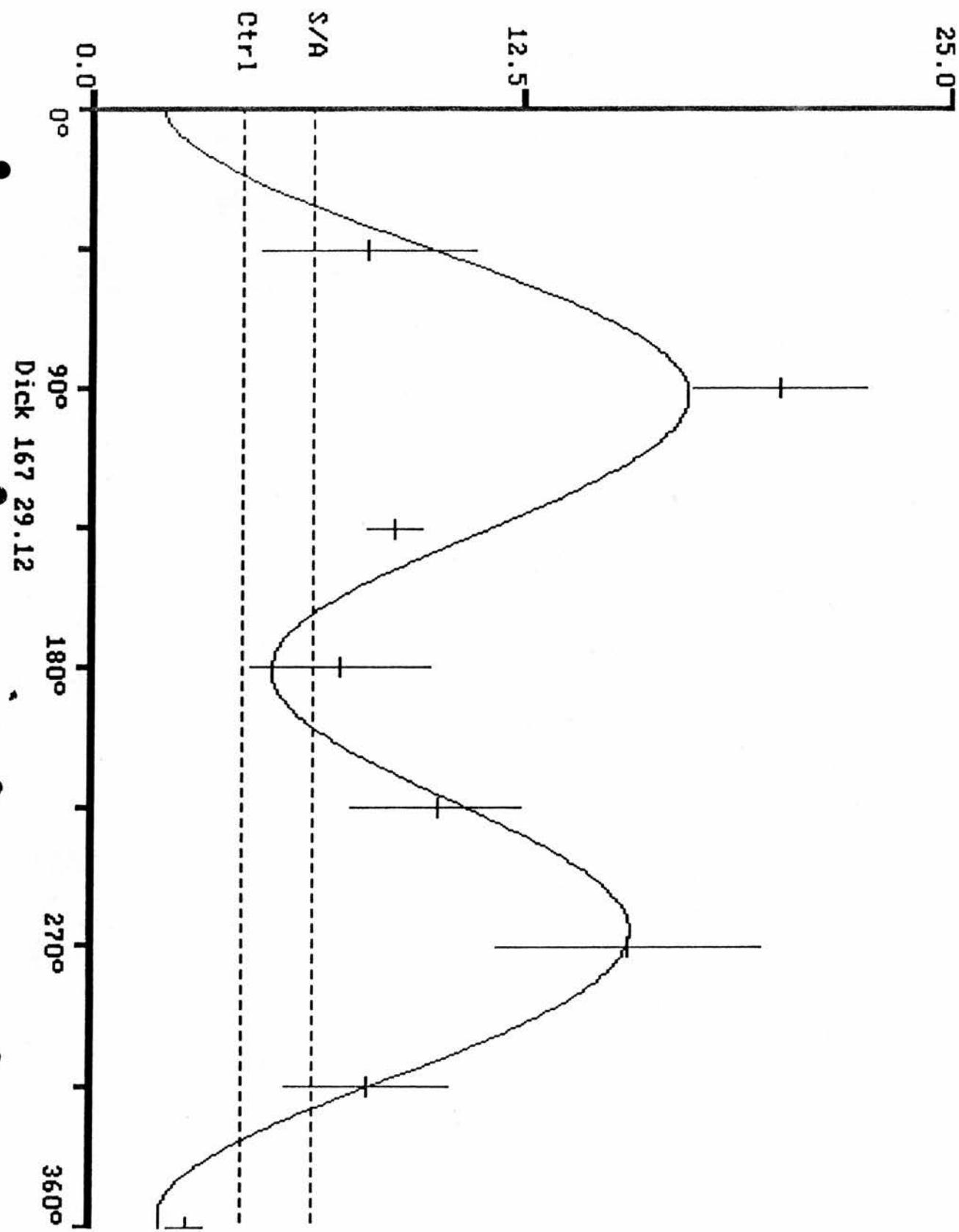


FIG 4

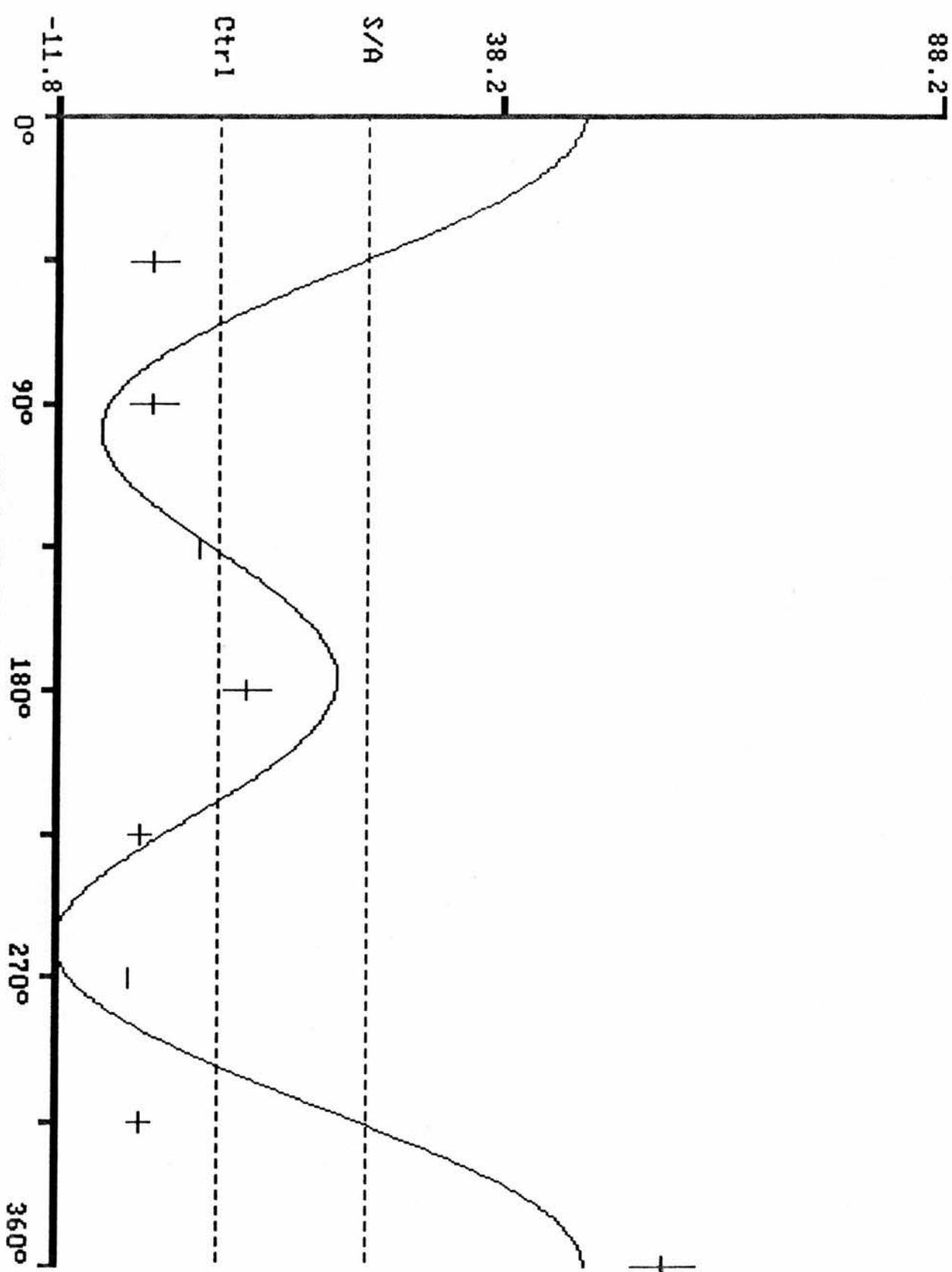


FIG 5

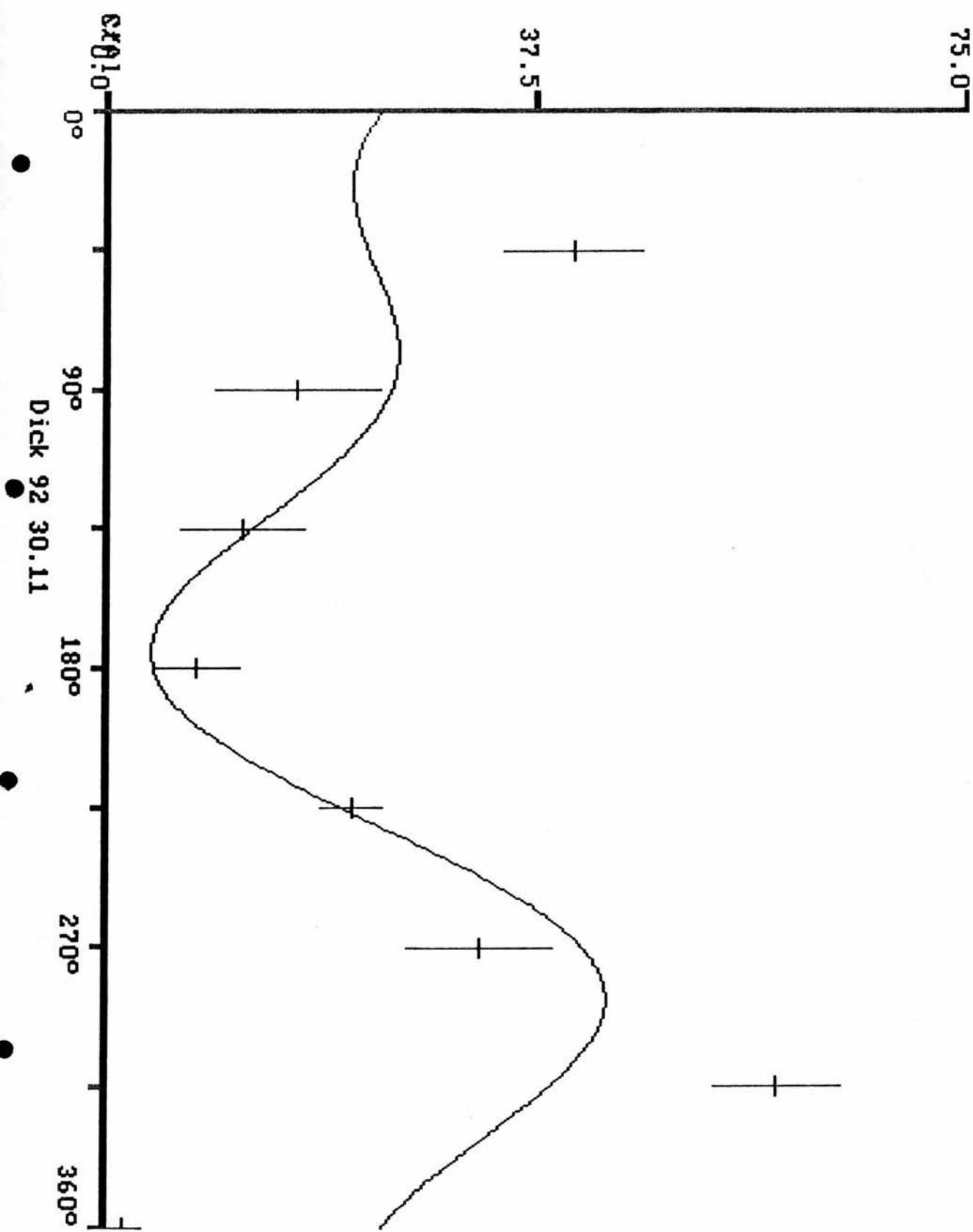


FIG 6

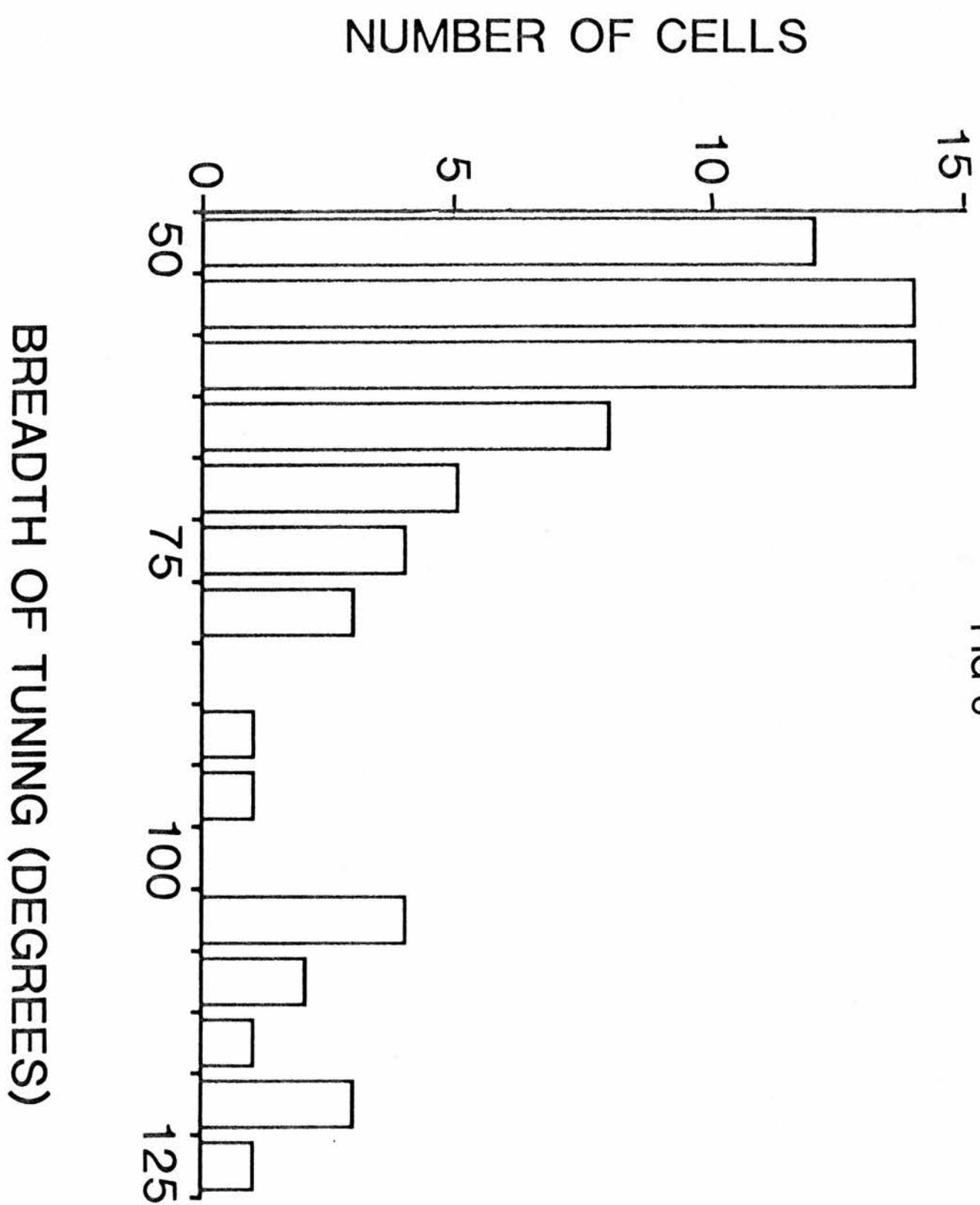
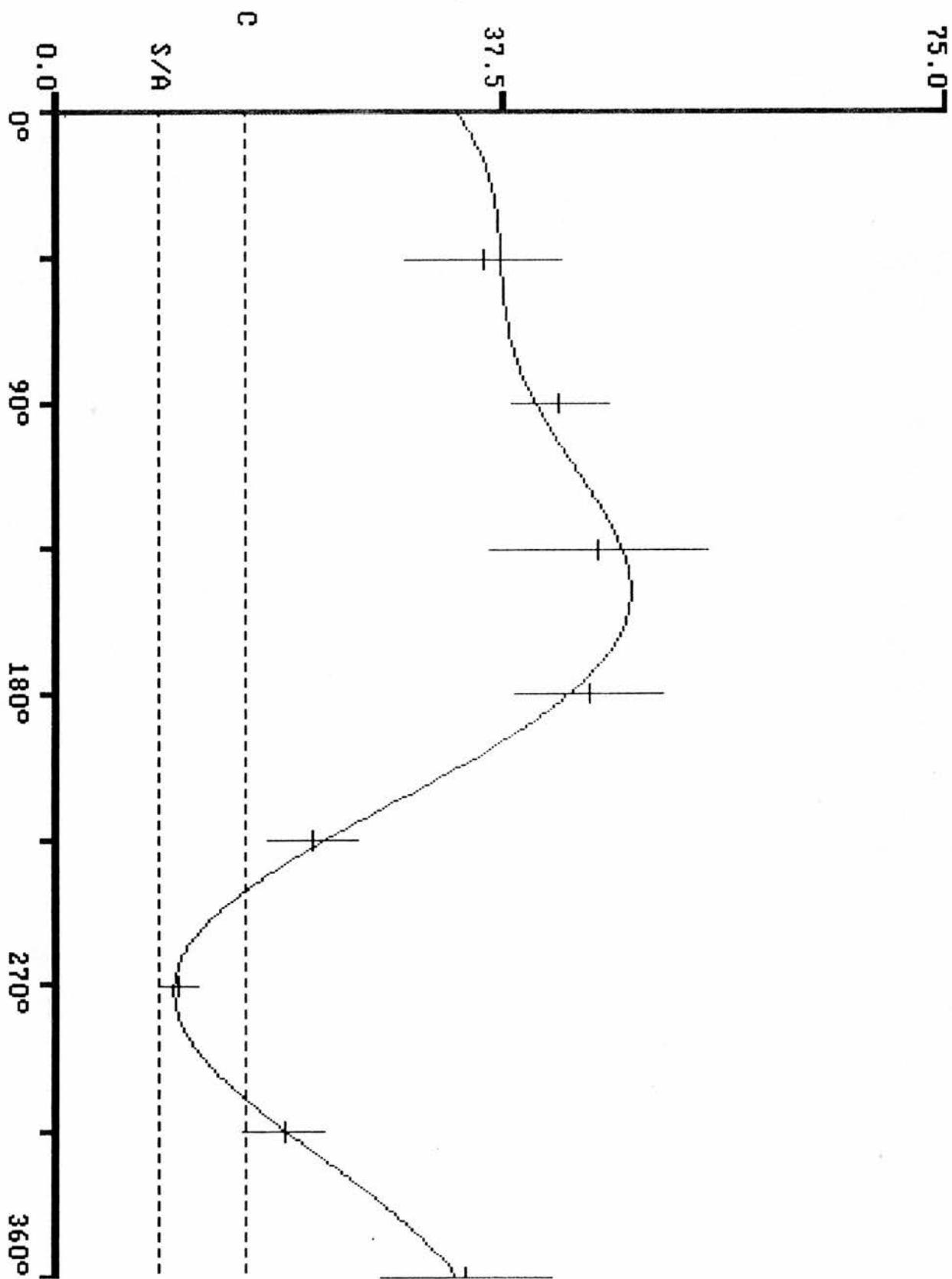


FIG 7



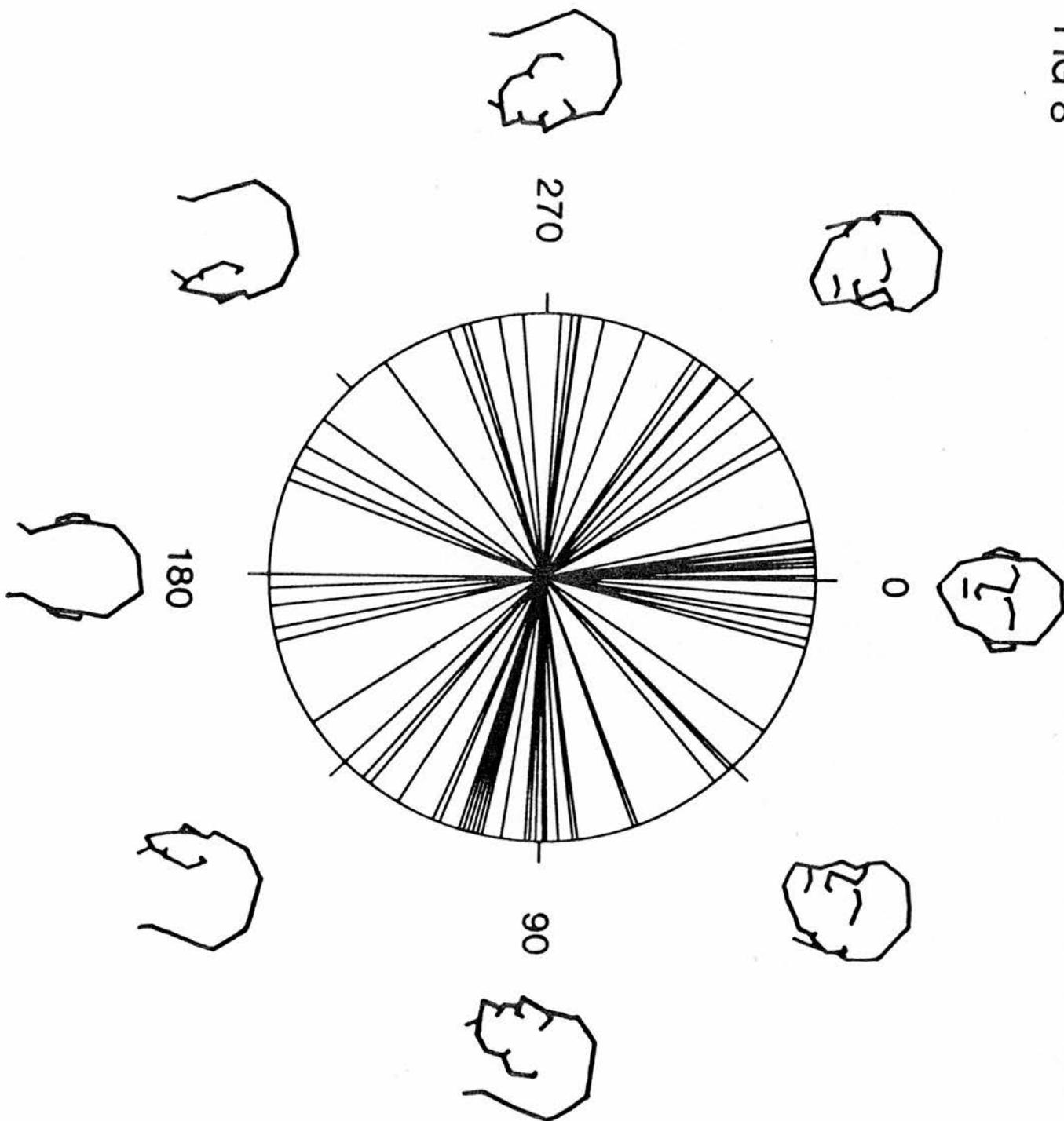


FIG 9

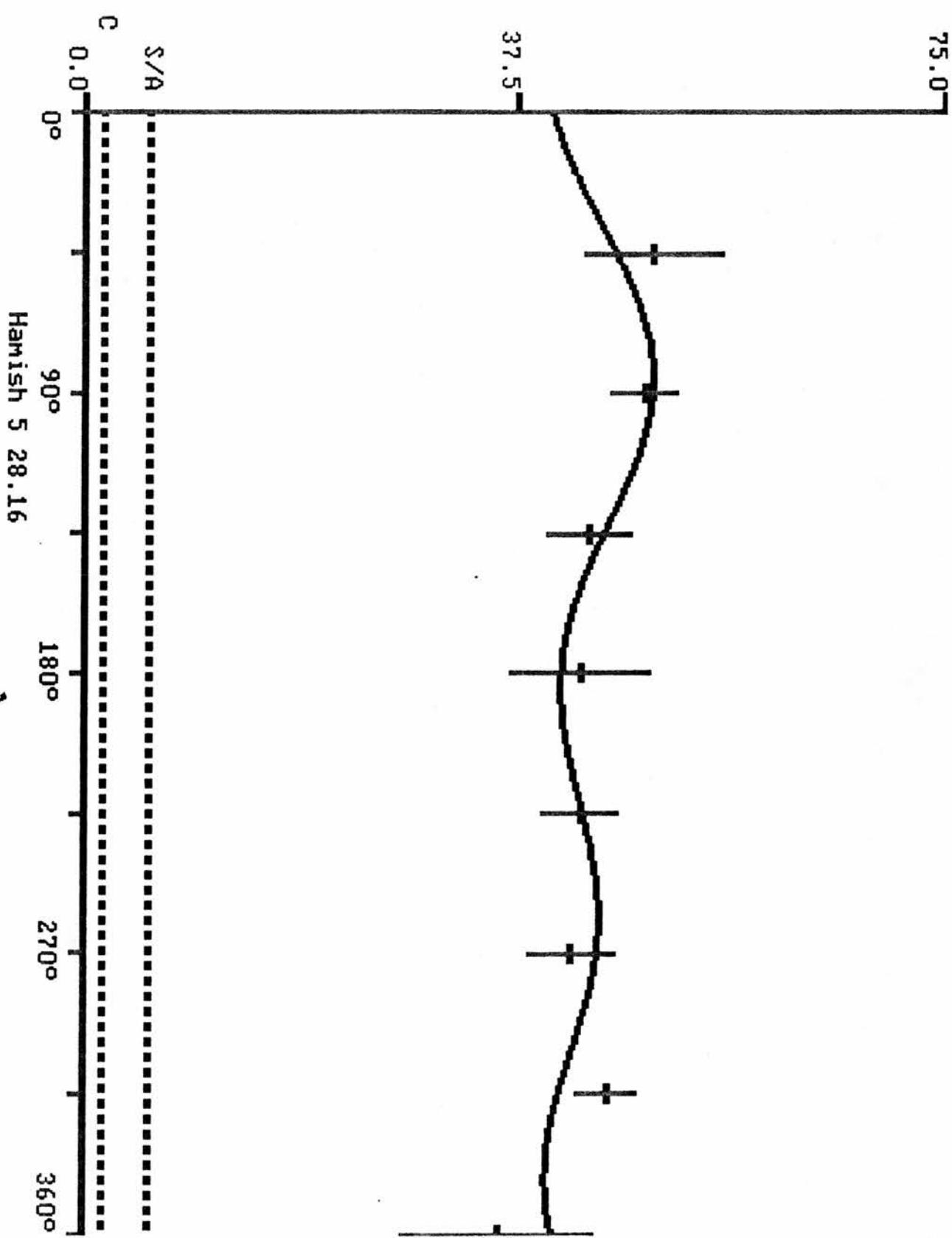


FIG 10

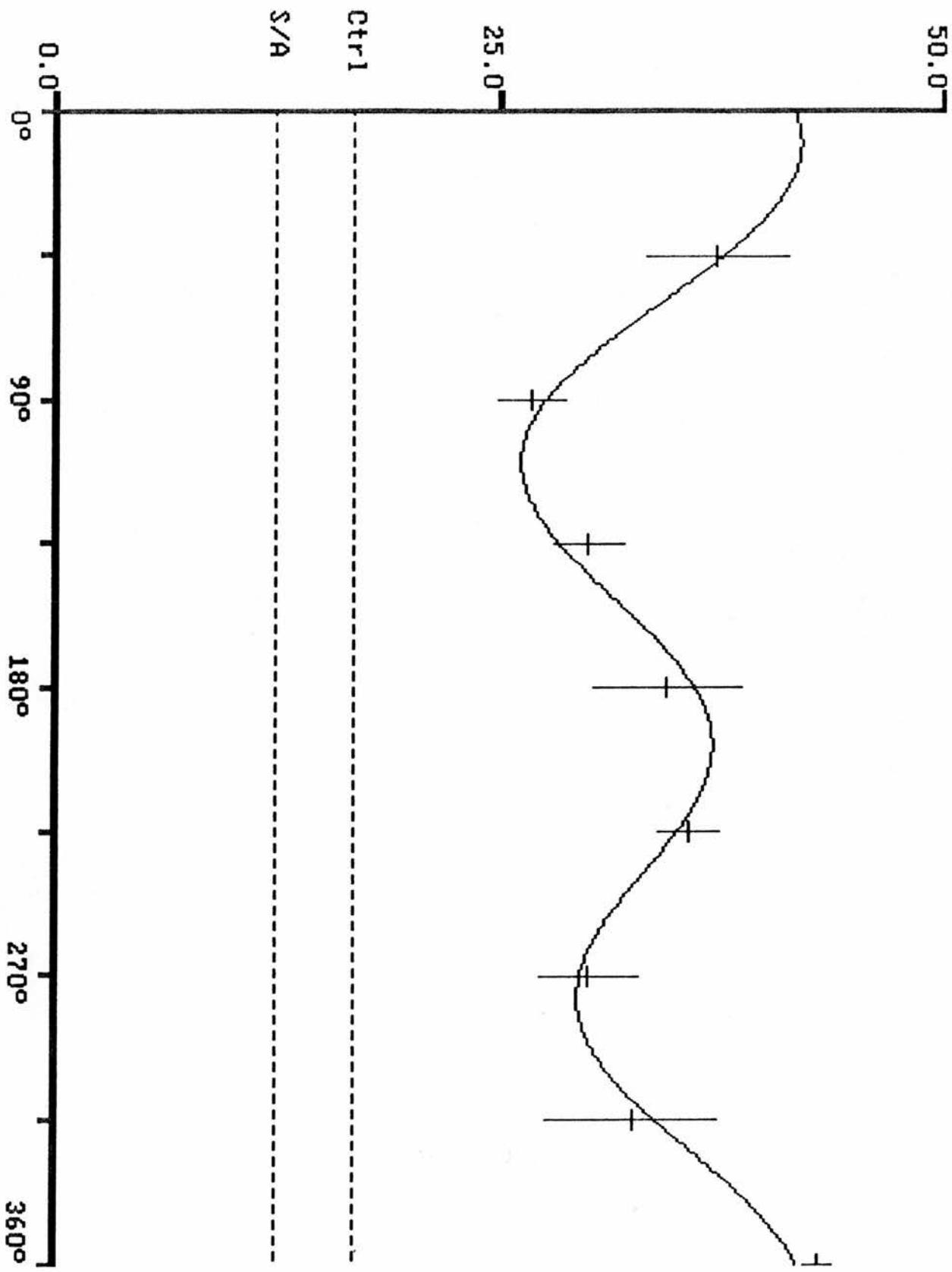
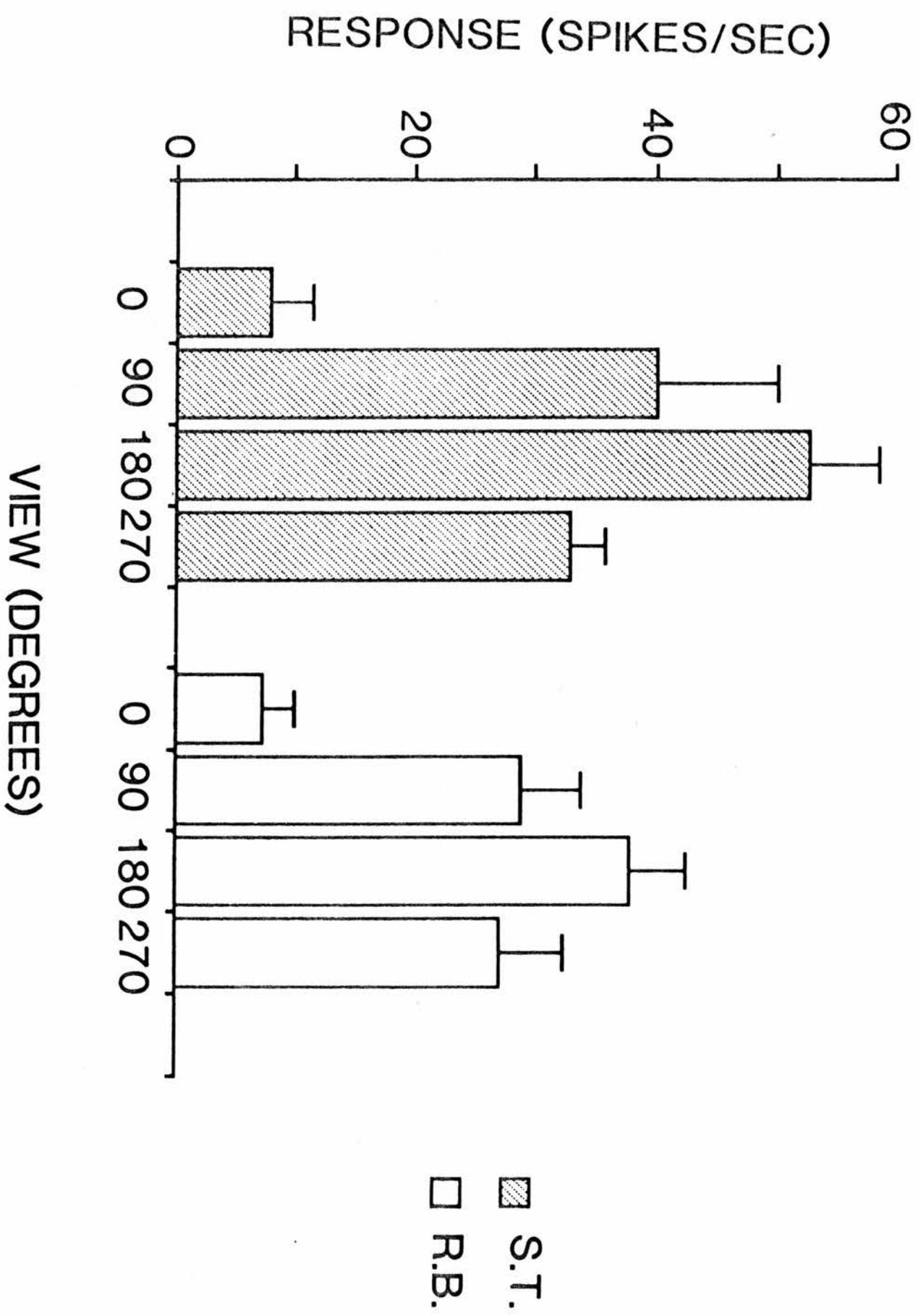


FIG 11



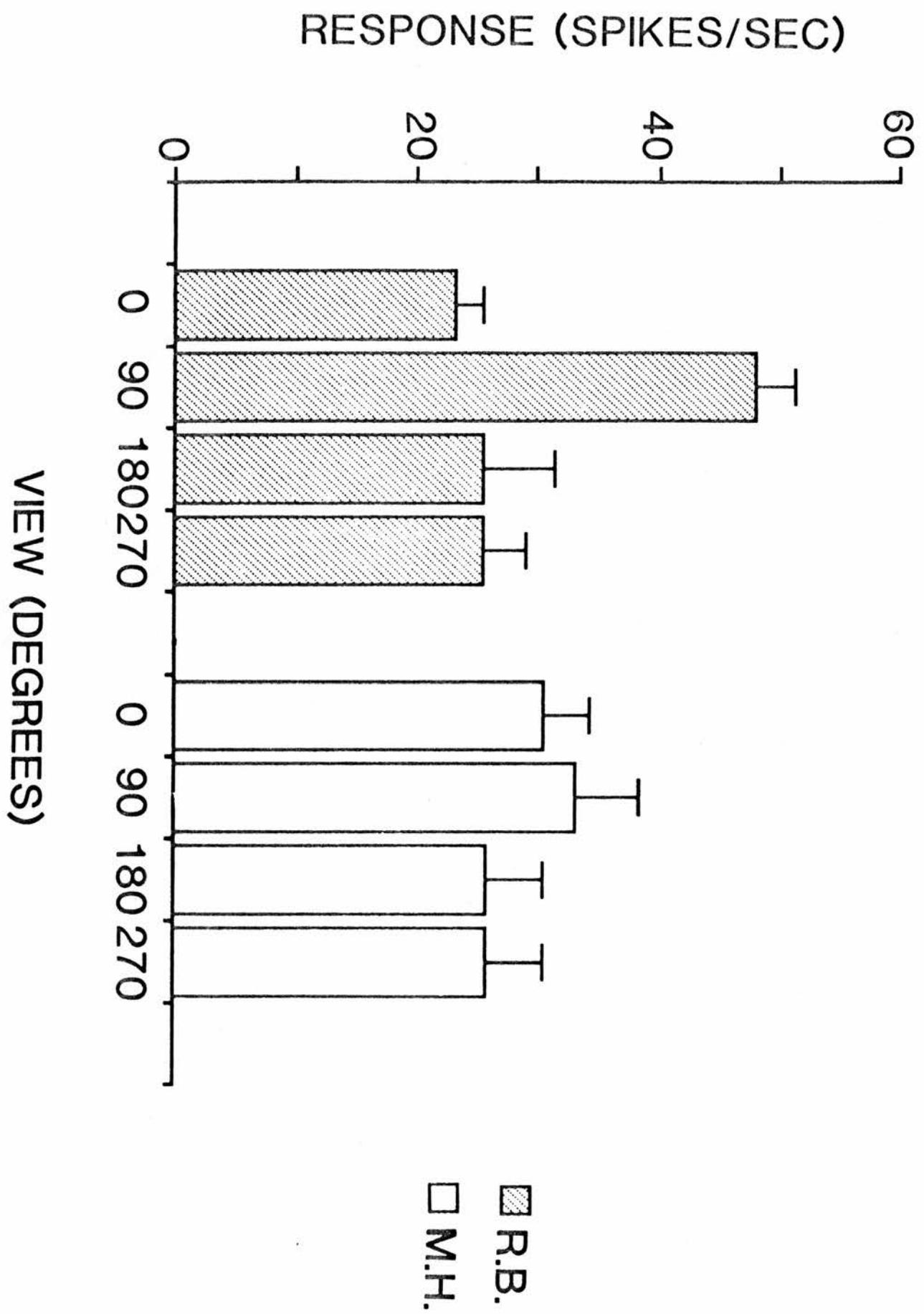
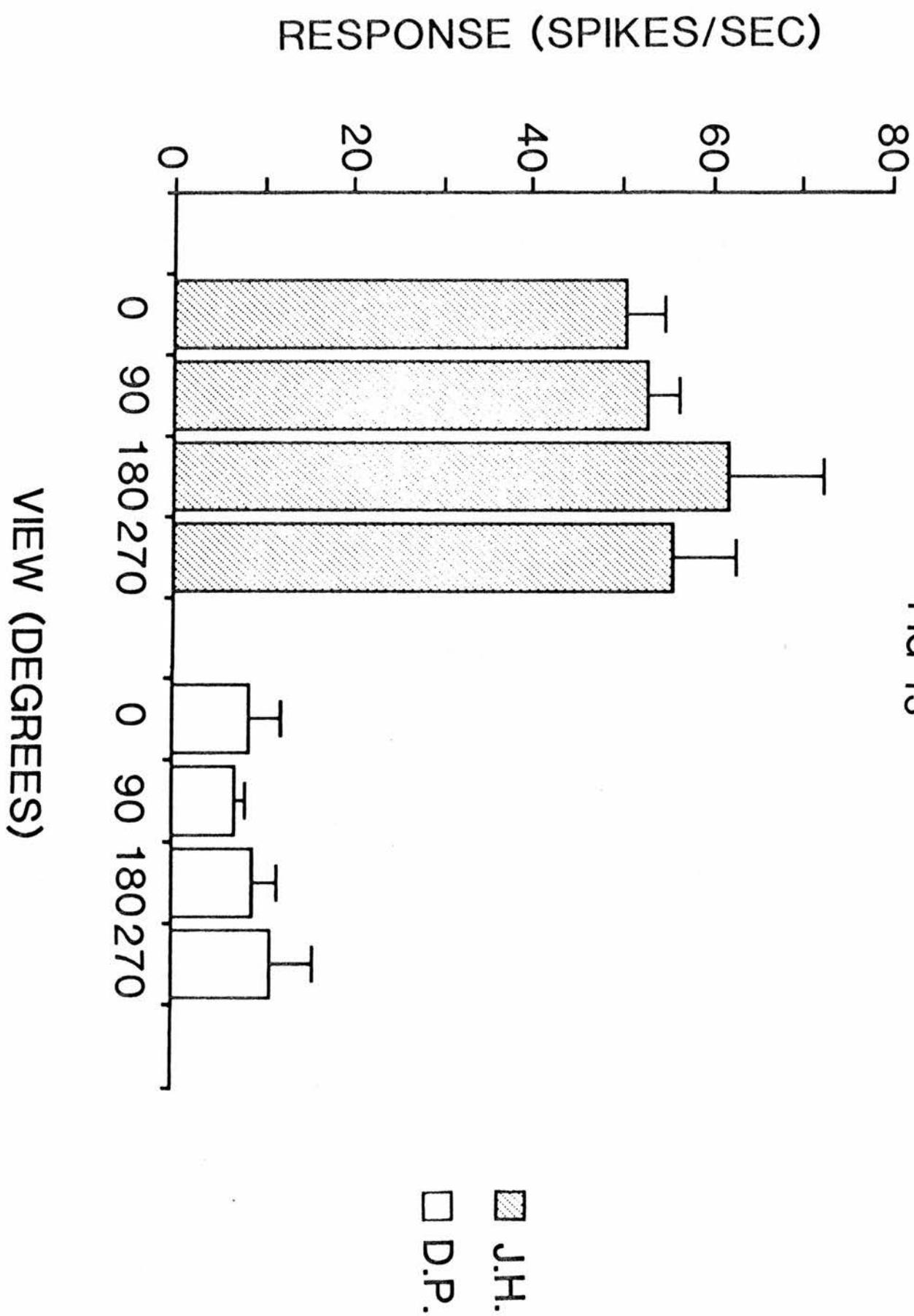


FIG 13



APPENDIX

METHOD OF REGRESSION USED AND R^2 VALUES OBTAINED

Method of regression

Observations consisted of neuronal responses R_1, \dots, R_n at angles of view $\theta_1, \dots, \theta_n$. In order to fit a theoretical tuning curve to these data, the technique of multiple (least squares, linear) regression was used. It was assumed that there is an equation of particular form (known apart from the value of certain constants) which gives the average value of R resulting from observations on any fixed value of θ . The equation used was

$$R = \beta_1 + \beta_2 \cos \theta + \beta_3 \sin \theta + \beta_4 \cos 2\theta + \beta_5 \sin 2\theta$$

where R is the neuronal response, θ is the angle of head view and β_1, \dots, β_5 are constants.

The values of θ are regarded as fixed by the experimenter (and so known exactly) while the values of the responses (R) are obviously variable. The values of the constants β_1, \dots, β_5 are unknown but there are various ways in which they might be estimated. In this study the method of least squares was used. This is used widely and has considerable theoretical justification (eg Freund 1973, Draper and Smith 1981). In this method β_1, \dots, β_5 are

calculated so as to minimise the squared sum of errors.

The above equation was chosen to provide a simple family of periodic (360 deg equivalent to 0 deg) functions which includes both unimodal and bimodal shapes. The condition of periodicity is the reason for the use of the periodic functions $\cos \theta$ and $\sin \theta$. (A simpler form of this equation which can be used with unimodal shapes only is given in Batschelet 1981, chapter 8.) The equation was designed to make the fewest possible assumptions about the nature of view tuning, allowing a tuning curve to have either one or two modes which need not be 180 deg apart. The curve also does not have to be symmetrical.

R² values of significant regressions

R² values of all significant ($p < 0.05$) regressions performed on the experimental data ranged from 0.19 to 0.90. The mean value was $0.43 +/ - 0.02$ SEM. The median value was 0.42.

Residual plots were not done. However, chi-squared tests were used to test goodness of fit (as described in the Results section above). In most cases where r^2 was low (although $p < 0.05$), chi-squared tests showed a bad fit. In these cases the optimal angle calculated was rejected if necessary

(see Results section for the criteria used in deciding on whether to accept the optimal angle in such cases).

REFERENCES

Freund J E (1973). *Modern Elementary Statistics* (4th ed). Prentice-Hall.

Draper N R and Smith H (1981). *Applied Regression Analysis* (2nd ed). Wiley.

Batschelet E (1981). *Circular Statistics in Biology*. Academic Press.