Chapter 5

Visual Representations in the Human MTL

5.1 Introduction

The work presented in the last chapter afforded us a look at how attention modulates brain activity. Although it was a big step forward in directly observing cortical activity, fMRI has its limitations in terms of the temporal and spatial resolution it offers. The fMRI response occurs over several seconds in comparison to the millisecond timescale of the electrical activity of individual neurons, and thus it is always unclear how the dynamics of the BOLD signal relate to the underlying neuronal mechanisms (Logothetis, Pauls et al., 2001; Logothetis & Wandell, 2004). In particular, given this long delay, one cannot assume that the signal being measured is due to a strictly local response to the stimulus. Instead, it is likely that several feedback signals from other processing centers in the brain participate in complex ways to constitute the BOLD signal. In terms of spatial resolution, the smallest voxels we could record from in our fMRI study (3.1 mm^3) would have included approximately 3 x 10^5 neurons[1] and thus the measured signal would reflect the averaged activity over a large neuronal population. While this average signal allows us to sample large parts of cortex, and could potentially represent general properties of a particular brain area, in order to understand how attention affects neuronal responses, it is imperative to study the electrical activity of individual neurons.

[1] 1 mm^3 of cortex is estimated to contain approximately 10^5 neurons.

As we have mentioned several times already, most of our understanding of attention has come from electrophysiological studies in awake behaving monkeys as they performed various tasks. By inserting electrodes into the brains of these animals, investigators could measure the underlying neural signal and examine how manipulating the focus of attention changed the response of the neurons. However, the effects of attention on individual neurons in the human brain have not been examined this far. Since the monkey brain is anatomically very similar to the human one, it is generally assumed that findings on monkey and human brains should be comparable. This assumption nevertheless does require one to extrapolate from one system to another, and consequently it is desirable whenever possible to record from the human brain and verify that the assumption is a reasonable one. In addition, certain experiments can be performed with humans that could never be feasible with monkeys (i.e., in relation to conscious perception, memories, etc.), and thus it is important to be able to conduct recordings in the human brain as well.

To complete this dissertation on attention, I will therefore describe work done on human subjects who performed an attentional paradigm while we recorded the activity of single neurons in their brains. However, since an understanding of the effects of attention on visual responses first requires an understanding of the visual responses themselves, in this chapter I will describe experiments that investigated the nature of the visual representations in the neurons we recorded from. The next chapter will describe the effects of attention on these visual responses. Parts of the recordings and analyses discussed in this chapter were performed by Rodrigo Quian Quiroga and are to appear in (Quian Quiroga, Reddy et al., 2005).

5.1.1 Subject population

Our subjects were epileptic patients at the University of California, Los Angeles (UCLA), who suffered from pharmacologically intractable epilepsy. In about 25% of the epileptic population, frequent seizure activity cannot be controlled by existing drugs (Engel, 1996; Ojemann, 1997), and thus the diseased portion of the brain where the focus of the seizure originates is often resected during surgery in a number of these patients. To locate the seizure focus, several non-invasive techniques such as structural MRIs, positron emission topography, and scalp EEG and MEG recordings are first employed. However, in some proportion of patients, these external measures do not yield sufficiently conclusive evidence about the location of a single epileptogenic focus, and patients must therefore be chronically implanted with up to 12 depth electrodes to locate the seizure focus (Fried, MacDonald, & Wilson, 1997; Fried, Wilson et al., 1999). Following electrode implantation, patients are monitored in the neurosurgical ward while their seizure activity is recorded via contacts on the electrodes. Patients typically spend a week to 10 days in the hospital—the length of time depends on clinical constraints, such as whether a sufficient number of seizures have been recorded for the doctors to acquire enough data to identify the seizure focus. During this period, the patients are willing to participate in experiments.

The surgeries at UCLA were performed by neurosurgeon Itzhak Fried. Dr. Fried was also involved at each stage in the planning of these experiments and in their actual implementation. Through the central lumen of each implanted (macro) electrode (1.25 mm in diameter), nine microwires (40 μm in diameter) were inserted from eight of which electrical activity of individual neurons could be recorded (Figure 5.1). The ninth microwire had lower impedance and served as a reference. The macro-electrodes contained 6–7 platinum contacts, which were used to acquire EEG data continuously during the entire length of time the patient spent on the ward. This EEG data was used

by doctors to examine seizure activity. All studies conformed to the guidelines of the Medical Institutional Review Board at UCLA.

5.1.2 Electrode locations

The placement of the electrodes was based solely on clinical criteria and was determined by the data obtained from the EEG scalp recordings and the other localization tests performed. The electrode locations were verified by MRI or by computer tomography co-registered to preoperative MRIs. It should be noted that the MRI images are obtained at 1.5 T, which provides sufficient resolution to identify the tip of the macro-electrodes but not the trace of the microwires. Thus the locations we report correspond to the macro-electrodes. Generally, the microwires extended 3–4 mm beyond the tip of the macro-electrode. In most epileptic patients, the medial temporal lobe (MTL) is the primary location in which seizures originate, and consequently most of our electrodes were placed in MTL structures such as the hippocampus, entorhinal cortex, parahippocampal gyrus, and the amygdala. Occasionally, electrodes were also implanted in the orbito-frontal cortex, the supplementary motor area, and the anterior cingulate cortex. See Table 5.1 for location of electrodes.

5.1.3 Anatomy of the MTL

In the monkey brain, the MTL system (also called the limbic system) marks the convergence of inputs from purely visual areas (Felleman & Van Essen, 1991) to the polymodal centers in the brain that receive somatosensory, olfactory, and auditory information as well (Figure 5.2) (Suzuki & Eichenbaum, 2000). In all primates, the MTL

consists of several distinct anatomical components that include the hippocampal formation (the dentate gyrus, areas CA3, CA1, and the subiculum); the entorhinal, perirhinal, and parahippocampal cortices; and the amygdala (Van Hoesen, Pandya, & Butters, 1972; Van Hoesen & Pandya, 1975a; Van Hoesen, Pandya, & Butters, 1975; Van Hoesen & Pandya, 1975b; Insausti, Amaral, & Cowan, 1987; Squire & Zola-Morgan, 1991; Suzuki & Amaral, 1994; Lavenex & Amaral, 2000). In the hierarchical organization of these components, the perirhinal and parahippocampal cortices are the structures that first receive input into the MTL. This input consists of information from both higher order areas such as the prefrontal cortex and also from the unimodal sensory areas. In the monkey brain, the perirhinal cortex receives particularly strong projections from visual area TE reflecting the role of this region in visual processing (Saleem & Tanaka, 1996; Suzuki, 1996b; Murray & Richmond, 2001), while the parahippocampal cortex receives prominent projections from dorsal areas and is more involved in spatial processing (Suzuki & Amaral, 1994; Malkova & Mishkin, 2003). The multimodal input from the perirhinal and parahippocampal cortices then feeds into the entorhinal cortex (EC) (Brown & Aggleton, 2001), which also receives some visual input from TE (Suzuki, 1996b). Extensive feedback connections exist between the entorhinal cortex and TE and the perirhinal and parahippocampal cortices.

The EC is the major source of cortical input to the hippocampal structures. Weaker projections from the perirhinal and parahippocampal cortices to the subiculum and CA1 layer of the hippocampus also exist in monkeys, but visual areas such as TE do not project directly to the hippocampus (Suzuki & Amaral, 1990). Instead, the EC appears to gate the flow of information in and out of the hippocampus.

The amygdala receives cortical input through the perirhinal, parahippocampal, and entorhinal cortices (Stefanacci, Suzuki, & Amaral, 1996; Suzuki, 1996a) although it also has other connections with different cortical and sub-cortical structures (LeDoux,

2000). For instance, direct connections to and from TE have been identified (Cheng, Saleem, & Tanaka, 1997).

Given the different amounts and sources of cortical input in each of these areas, it is likely that the different structures process information in distinct ways (e.g. (Baxter & Murray, 2001)). Since we have not recorded from sufficient numbers of neurons in each of these areas, we cannot make qualitative or quantitative distinctions in the nature of the visual responses we observe. However, it is important to keep this fact in mind in the later sections of this chapter where we present the visual properties of our MTL neurons.

5.1.4 Memory and MTL

Ever since the dramatic case of amnesia observed in patient H.M. after bilateral damage to the medial temporal lobe (Scoville & Milner, 1957; Milner, 1972), MTL structures have been linked to memory formation and retrieval. In particular it is now believed that these structures are important for the function of declarative memories (the memory for facts and events as opposed to non-declarative memory of skills and abilities) (Eichenbaum, 1992; Squire, 1992; Squire & Zola, 1996; Eichenbaum, 2000). Evidence for this view comes from patients and monkeys with lesions in MTL structures who reveal severe declarative memory impairments (e.g. a loss of autobiographical memories in humans) (Zola-Morgan, Squire, & Amaral, 1986; Zolamorgan, Squire, & Ramus, 1994; Stefanacci, Buffalo, Schmolck, & Squire, 2000).

Not all aspects of declarative memory function are damaged, however, as a result of these lesions. Immediate memories for items just brought into consciousness are largely intact, even in patients with extensive damage in the MTL (Drachman & Arbit, 1966). Animal studies have also shown that rats with hippocampal lesions can perform normally on tasks that require memory formation during short delays (~4s). However,

when the delays become longer (1 or 2 minutes), performance is impaired (Clark, West, Zola, & Squire, 2001). Thus, it appears that MTL structures are involved in the formation of memories over longer intervals of time. Memory of remote events and facts are also usually spared following lesions, although memories acquired just prior to the lesion are generally lost (Squire, Stark, & Clark, 2004). The observation that remote memories are not lost following lesions suggests that the MTL is not the permanent storage site of memory but that long-term memories must be stored elsewhere. Higher visual areas such as TE have been implicated in such visual memory storage, and it is supposed that the medial temporal lobe works together with area TE to establish these visual memories (Mishkin, 1982; Miyashita, 1993; Higuchi & Miyashita, 1996).

Given the multimodal input into the MTL (as discussed in the last section), the hippocampus may also be involved in tasks that depend on pooling information from various sources. Thus the formation of associative memories (i.e. pairing information about different items, such as connecting a face with a name) has been linked to the hippocampus (Stark, Bayley, & Squire, 2002; Stark & Squire, 2003; Wirth, Yanike et al., 2003). Consistent with this role of the MTL in establishing associative memories are findings that report that MTL structures participate in the recall of visually associated information and that lesions to these structures abolish the ability of TE neurons to represent associations between stimuli (Sakai & Miyashita, 1991; Higuchi & Miyashita, 1996).

5.1.5 Visual responses in the MTL

In accordance with the fact that the MTL receives high-level visual input, stimulus specific responses to complex stimuli have been observed in these structures. In monkeys, about 66% of cells in the perirhinal cortex respond selectively to specific visual

stimuli while in the entorhinal cortex 12% of cells are stimulus selective (Miller, Li, & Desimone, 1993; Suzuki, Miller, & Desimone, 1997). (This difference in the proportion of selective responses could be related to the fact that the perirhinal cortex receives stronger direct visual input from area TE compared to the entorhinal cortex, as discussed in Section 5.1.3). fMRI studies in the human parahippocampal cortex have revealed that this region responds specifically to visual images of buildings, scenes, and spatial layouts (Aguirre, Detre, Alsop, & Desposito, 1996; Epstein & Kanwisher, 1998; Epstein, Graham et al., 2003).

Relevant to our purposes, previous single unit recordings in humans have also reported visual responses to stimuli such as faces and objects in the MTL, and the nature of these responses tends to be fairly complex (Fried, MacDonald et al., 1997; Kreiman, Koch et al., 2000a). For instance, in the study by Fried et al., some neurons in the hippocampus and entorhinal cortex responded differentially to faces depending on the conjunction of facial expression and gender of the face. In other words, these neurons signaled an association between these different attributes. In another instance of complex representations, Kreiman et al. (2000) reported that visually responsive neurons in the MTL could base their responses on broad category level descriptions of stimuli. Thus, they observed an individual neuron in the entorhinal cortex that responded selectively to all images of animals that were presented, but not to any of the seven other stimulus categories used (i.e. emotional faces, objects, cars, spatial layouts, patterns, famous faces, and drawings of famous faces). Another neuron in the anterior hippocampus responded strongly to all drawings of famous faces, and to a lesser extent to pictures of famous faces, but not to the other categories of stimuli. Overall, 12% of the recorded neurons exhibited similar category selective behavior. Distinguishing images based on their category is clearly a high-level semantic distinction, and these

results suggest that rather than base their responses on low-level similarities between different images, MTL cells can encode abstract links between visual stimuli.

In another remarkable study, Kreiman and colleagues (Kreiman, Koch, & Fried, 2000b) also reported that MTL neurons were not only responsive to visual stimulation, but could also signal imagery or recall of the stimuli they represented. Thus when subjects were prompted to imagine a particular stimulus they had previously viewed, the corresponding neuron would show enhanced activity, even though the relevant image was not physically presented. Importantly, in 88% of the neurons that fired selectively during both imagery and recall, these two selectivities were comparable. These results provide further evidence for the observation that MTL neurons encode high-level, conceptual information since their responses were modulated even in the absence of visual input.

In light of these high-level visual responses and the putative role of the MTL in maintaining associations between different stimuli, it is interesting to ponder how abstract representations between various items are organized in the MTL structures. Do individual neurons carry explicit information about several aspects of a particular concept so that they could respond across different expressions of this concept? In other words, are responses invariant to different stimulus manipulations as long as the represented concept is left unmodified? Or is such information present only over distributed networks of neurons? For instance, does the neuron in my brain that codes for my dog do so for several different views of my dog? Or is it tuned for just one particular image? What about the dog's name? Can this semantic information also be represented in the same neuron? Or will all the different information be spread out over a larger neuronal population? This is the question that we aim to address in this chapter. As is described below, by presenting our patients with several different pictures of individuals,

landmarks, and objects, we tried to determine whether MTL neurons encoded information in an invariant manner.

5.2 Methods

5.2.1 Patients and recordings

The data reported here was obtained in 8 patients (8 right-handed, 3 male, 17 to 47 years old) with pharmacologically intractable epilepsy. As described in Section 5.1.1 and 5.1.2, these patients were implanted with chronic depth electrodes and were monitored in the hospital ward for 7–10 days, during which period this study was performed. The electrodes were located in the hippocampus, amygdala, entorhinal cortex, and parahippocampal gyrus (see Table 5.1 for summary of electrode locations).

Our data acquisition system was obtained from Neuralynx, Tucson, Arizona, and allowed us to record data from 64 microwires (in 8 macro-electrodes as discussed earlier) at a time. In 6 of the 8 patients, all 64 microwires were used. The remaining 2 patients were implanted with only 7 macro-electrodes for clinical reasons. The data from the active channels was amplified, usually with an amplification factor of 10,000, although it did change occasionally depending on the quality of the signal. The sampling frequency was always 28 kHz. The data was band-pass filtered by hardware between 1 and 9000 Hz. This continuous broadband signal was then digitally acquired and stored (using the Neuralynx Cheetah software).

5.2.2 Spike sorting

The full details on the spike-sorting procedure are described elsewhere (Quiroga, Nadasdy, & Ben-Shaul, 2004). A briefer description is presented here. Two steps were involved in the processing of the broadband data from the Neuralynx system before we obtained the action potentials. Firstly, from the continuous signal that contained both spikes and background electrical activity and noise, all the spikes that occurred had to be detected. Spike detection was achieved by band-pass filtering the signal between 300 Hz and 1000 Hz and marking all crossings of an amplitude threshold. The threshold was proportional to the median of the signal divided by a constant value. All events in the band-pass filtered signal that crossed this threshold were labeled as candidate spikes, and the times at which the maximum of these events occurred were recorded.

In the second step, the data had to be clustered to separate out spikes that corresponded to different neurons (recorded from the same microwire) or to artifacts. For the purposes of clustering, a different band-pass filter, between 300 Hz and 3000 Hz was necessary. The primary reason for changing this filter setting was to be able to get better resolution for observing the spike shapes. In other words, while the 300–1000Hz filter allowed us to reject all higher frequency artifacts, it also smoothed out the data excessively, which prevented effective clustering (since clustering is based on differences in spike shapes). Thus to get better resolution, the original signal was re-filtered, and datapoints around the spike timestamps (determined in the previous step) were stored for further analysis. For each spike, 64 datapoints (encompassing about 2.3 ms of data; 20 points before and 44 after the maximum value of the spike) were saved. All spikes were aligned to each other such that their maximums occurred at data point 20. Spike maxima were determined from cubic spline interpolated waveforms 256 samples in length. Once the spike shapes were obtained, clustering based on the wavelet transform and super-paramagnetic clustering was performed as described in (Quiroga, Nadasdy et al., 2004). Briefly, for each spike, 64 wavelet coefficients were

obtained, and the 10 coefficients that best accounted for the different spike shapes were fed into the clustering program. The clustering algorithm was based on super-paramagnetic clustering, a stochastic clustering procedure, which groups the data into clusters as a function of a parameter called the temperature. The temperature determines the probability of different spikes belonging to a particular cluster. At low temperatures, the probability of spikes being clustered together into one cluster is high whereas at very high temperatures, the clusters can "fragment" and several clusters with a small number of spikes each can form. At some optimal temperature between these two extremes, spikes corresponding to different neurons will form distinct clusters.

As shown in Figure 5.3, the temperature was a parameter that could be changed by the experimenter if the automatic clustering results were not satisfactory. About 20% of the time, we modified the output of the automatic (i.e. unsupervised) clustering program by i) combining clusters from two different temperatures by decreasing this parameter; or ii) splitting existing clusters into two or more smaller clusters by increasing the temperature; and/or iii) assigning membership of unclustered spikes to nearby clusters via a template matching strategy based on the Euclidean distance between the center of each cluster and the unclustered spike.

Once this procedure was complete, we classified each cluster into single or multi-units. Multi-unit clusters were those that reflected the contribution of two or more neurons to the cluster and whose spikes could not be further differentiated due to low signal to noise ratios. This classification into single and multi-units was done visually based on the following criteria: i) the spike shape and its variance; ii) the ratio between the peak of the spike and the noise level; iii) the presence of a refractory period for single units such that less than 1% of the spikes occurred within 3 ms of each other. Additionally, for both single- and multi-units, the ISI distributions were examined for peaks at 16.6 ms that correspond to 60Hz noise.

The results of clustering data from one microwire are shown in Figure 5.3. The blue cluster corresponds to a multi-unit, and the other 3 clusters correspond to single units. Over all 8 patients, a total of 998 units were recorded, of which 346 were classified as single units and 652 as multi-units. On average 47.5 units (16.5 single units, 31 multi-units) were obtained in a single recording session.

5.2.3 Stimulus presentation and behavioral task

Subjects lay in bed while stimuli were presented to them on a laptop Macintosh OS9 G3 powerbook. Each stimulus was presented at the center of the screen for 1 s. The size of the stimuli was about 1.5 degrees of visual angle. Each image was presented 6 times in a pseudo-random order, and the subject had to press either one of two keys (‘Y’ or ‘N’) to report whether the presented image was of a human face or not. This simple task ensured that subjects paid attention to the images.

The pictures were obtained from the web and consisted of images of animals, famous and non-famous places and buildings, and faces of famous and non-famous people. These categories of stimuli were chosen based on previous reports of visual responses in the MTL (Kreiman, Koch et al., 2000b, 2000a; Kreiman, 2002) and also the observations made by other researchers about the nature of visual responses in high level areas. In particular, since the hippocampus and parahippocampal areas have been implicated in navigation and the processing of spatial layouts and buildings, we included pictures of this type (O'Keefe & Dostrovsky, 1971; O'Keefe & Conway, 1978; Muller, Kubie, & Ranck, 1987; Wilson & McNaughton, 1993; Aguirre, Detre et al., 1996; Epstein & Kanwisher, 1998; Epstein, Harris, Stanley, & Kanwisher, 1999; Burgess, Maguire, & O'Keefe, 2002; Epstein, Graham et al., 2003). Similarly, since several studies in monkeys have reported that cells in V4 and IT respond to highly familiar objects such as

paperclips and other stimuli that the animals had been exposed to over a long period of time (Miyashita & Chang, 1988; Logothetis, Pauls, & Poggio, 1995; Kobatake, Wang, & Tanaka, 1998; Yang & Maunsell, 2004), we included pictures of current celebrities who patients (especially around Hollywood!) would have been familiar with. Based on the reports by Perrett, Rolls, and colleagues that cells in monkey IT and amygdala respond to different faces, we included images of non-famous faces (Rolls, 1984; Leonard, Rolls, Wilson, & Baylis, 1985); (Perrett, Rolls et al., 1982; Perrett, Smith et al., 1984, 1985; Perrett, 1987; Perrett, Hietanen, Oram, & Benson, 1992). Frequently, in discussions with the patients, we were able to find out what their interests were, or what their favorite movies were, etc., and to include relevant pictures based on these conversations. Figure 5.4 represents the overlap in images presented during the various screening sessions for all eight patients.

5.2.4 Screening/testing experiment format

A large proportion of our success in this project was due to a new experimental format that we implemented. In this format, we had two separate sessions, which I will refer to as the screening and testing sessions. In the screening session, we presented subjects with an average of 88.6 (range: 70-110) images from the categories described above. The data obtained from this session was rapidly analyzed, and the stimuli that elicited a response from any of our cells were selected. It should be noted that for this analysis, we did not cluster the data into separate units. Instead, we simply detected the spikes and looked at the multi-unit data obtained from each microwire. Typically, this analysis took about 3–5 hours. Once the selective stimuli for the cells were determined (typically only 3.1% (range: 0.9%–18.0%) of the pictures in a session elicited a response), we prepared the testing session. In this session, we presented between 3

and 8 different views of the stimuli that were selective from the screening session. All the images that elicited a visual response in the screening sessions were included in the testing sessions. Thus the screening session served as a screen for later sessions since, based on their results, we knew the selective stimuli and the responsive cells every testing day. In the testing session, in addition to presenting multiple views of selective stimuli, we also presented a random number of stimuli that did not elicit selective responses from any of the cells. The total number of stimuli presented was determined by the time available with the patient for a single recording session (~30 minutes). We had a total of 21 testing sessions.

5.2.5 Determining visual responsiveness

All trials of the experiment were aligned to stimulus onset (0 ms). The median number of spikes occurring across trials between [300 1000] ms following stimulus onset was determined for each stimulus. Baseline activity was computed as the average number of spikes over all stimuli that occurred in the [-1000 -300] ms interval before stimulus onset. A unit was considered responsive if the activity to at least 1 picture fulfilled two criteria: i) the median number of spikes during stimulus presentation was larger than the average number plus 5 standard deviations of spikes that occurred during the baseline interval; ii) the median number of spikes was ≥ 2. The measure of activity was computed using the median rather than the mean for this experiment to diminish the contribution of outliers.

5.2.6 ROC analysis

The degree of invariance observed for neurons that had selective responses to a given stimulus was analyzed by means of an ROC analysis (Green & Swets, 1966). The idea was to determine how selectively a neuron would respond to *different* pictures of a particular individual (for instance) but not to any other pictures that were presented. This was accomplished by measuring the number of responses to this individual (hits) and the number of responses to other stimuli (false alarms) using different thresholds (T) for the prediction. A response was said to occur when the neuron's activity crossed the threshold (here the activity was defined as the median number of spikes that occurred over all trials on which a particular picture was presented). The hits were defined as the number of responses to an individual divided by the total number of pictures of this individual. The false alarms were defined as the number of responses to other pictures divided by their total number. By setting different values of the threshold T for counting a response, the number of hits and false alarms could be determined. Thus, at very high thresholds, activity to the individual and to the other pictures would be less than the threshold, and therefore neither hits nor false alarms would occur (lower left hand corner in the ROC plots shown in Figures 5.5–5.9). As the threshold would be decreased, the number of hits and false alarms would change. For a cell that responded exclusively to pictures of a particular individual, hit rates of 1 and 0 false alarms would be observed as the threshold was lowered. This would correspond to steep increases in the ROC curve. On the other hand, for units that responded to a random selection of pictures, similar numbers of hits and false alarms would be observed that would correspond to an ROC curve close to the diagonal. The degree to which a unit could be considered invariant was defined on the basis of the area under the ROC curve. For highly invariant cells, the area under the ROC curve would be close to 1, whereas cells that responded randomly would have areas of about 0.5.

For each cell, we compared the ROC curve obtained as described above with 99 surrogate ROC curves. The surrogate curves were each obtained by randomly choosing a set of *n* different pictures and comparing the hit rates obtained for this set of pictures with the false alarms obtained for the remaining pictures. The value for *n* was identical to the number of pictures of the individual for which invariance was tested for this cell. If a cell was selective for only pictures of a particular individual, then the area of the surrogate curves would be much less than 1. A unit was considered to be invariant to an individual or object if its area was larger than the area of all 99 surrogate curves with $p < .01$.

The ROC analysis could also have been performed by considering responses on a single trial (rather than taking the median across trials). In this case, hits would correspond to the proportion of trials (on which the individual was presented) on which a response occurred. The false alarms would correspond to responses on trials on which other pictures were presented. Similar results were obtained using both ROC measures[2].

A one-way analysis of variance (ANOVA) also yielded similar results. In particular, we tested whether the distribution of median firing rates showed a dependence on the factor identity (i.e. the individual, landmark, or object shown). The different views of each individual or object were the repeated measures[3]. The ANOVA analysis, however, does not explicitly show how invariant the responses were across different views of the individual or object. On the other hand, the ROC analysis explicitly tests the presence of an invariant as well as sparse representation.

The ROC and ANOVA analyses were performed by Rodrigo Quian Quiroga.

---

[2] As is discussed later (Section 5.3), 52 of 137 responsive units showed invariance as defined by means of the ROC analysis computed over the median across all trials. On a single trial level, 56 units were found to be invariant.

[3] According to the ANOVA analysis, 50 units had a significant effect for factor identity with $p<.01$.

5.3 Results

The response of a single unit in the left posterior hippocampus to 30 out of the 87 images presented is shown in Figure 5.5. The cell did not show a statistically significant response to the other pictures. The unit fired selectively to very different images of the actress Jennifer Aniston but not to other famous or non-famous faces, landmarks, objects, or animals. It is interesting to note that this cell did not respond to pictures where Jennifer Aniston appeared with the actor Brad Pitt. The response to Jennifer Aniston was characterized by a sharp burst of 5 to 10 spikes between 300 and 600ms post-stimulus. During the pre-stimulus interval, the cell was almost silent (average of 0.02 spikes in the baseline interval). Figure 5.5b shows the results of the ROC analysis obtained for this cell. The red line corresponds to testing invariance for all seven pictures of Aniston, while the grey lines are the ROC values for the surrogate curves (see Section 5.2.6). As expected from the data shown in Figure 5.5a, the surrogate ROC curves are close to the diagonal while the curve testing for invariance to Aniston has an area of 1.0. Thus this unit fired selectively only to pictures of Aniston and not to a random assortment of stimuli.

Figure 5.6 shows the activity of another selective unit in the right anterior hippocampus. The preferred stimuli of this cell were different images of the actress Halle Berry. There are several interesting points to make about this cell. Firstly, this cell not only responded to various pictures of Halle Berry but also to a caricature of her (although not to other caricatures). Given the large differences between the pictures and caricature, this result is striking. Secondly, the cell responded to images of Halle Berry in character as “Catwoman,” her role in a recent film (but not to other pictures of Catwoman). Since in the images of Catwoman, the actor is masked and unrecognizable

in her costume, it is likely that this response arises as a result of the patient's knowledge of the movie. Finally, the cell also responded to the letter string "Halle Berry"! These results clearly suggest that this invariant pattern of responses is not due to low-level similarities between the images. Rather they represent high-level, semantic knowledge about the relationship between the images. The ROC curves for this cell are shown in Figure 5.6b. The area under the curve was 0.99.

The responses of a cell to famous landmarks are shown in Figures 5.7 and 5.8. The unit in Figure 5.7 was located in the left anterior hippocampus and responds to pictures of the Sydney Opera House and the Bahai Temple in New Delhi. Interestingly the patient identified the images of the Bahai Temple to us as the Sydney Opera House. This unit responded to the letter string "Sydney Opera" as well although not to other words (such as "Eiffel Tower" or "Bahai"). The unit in Figure 5.8 responded to several pictures of the Tower of Pisa.

Interestingly, in contrast to the cell shown in Figure 5.5, which only responded to Jennifer Aniston alone, we observed another cell in the same experimental session that only responded to images of her with Brad Pitt. This cell, shown in Figure 5.9, was located in the right posterior hippocampus (the other side of the brain compared to the Jennifer Aniston cell). The ROC area for this cell was 1.0.

Over all 8 patients, we recorded from a total of 998 units. Of these, 137 neurons (67 single and 70 multi-units) were visually responsive to one or more images according to the criteria defined in Section 5.2.5. The proportion of visually responsive cells we observed (13.7%) is comparable to previous reports (14 %, (Kreiman, 2002)). Of the 137 visually responsive cells, 52 (37.9%, 31 single units and 21 multi-units) showed invariance to a particular individual (38 units), landmark (6 units) animal (6 units) or object (2units). Eight of these invariant units were invariant to two different concepts

(e.g. two different people, or a person and an object) in that they responded to all the presented views of these two concepts.

In all 52 cases, the area under the ROC curves was significantly higher than that of the 99 surrogate curves as described above ($p < .01$). The median value for the area under the ROC curves over all cells was 0.94, and these values ranged from 0.76 to 1.0. The distribution of the areas under the ROC curves is shown in Figure 5.10.

The locations of the invariant units as a proportion of the number of cells in each area are shown in Figure 5.11. Relative to the number of cells we recorded from in each area, 45% of cells in the hippocampus, 54% of cells in the parahippocampal gyrus, 25% of amygdala cells, and 22% of entorhinal cortex cells showed invariant responses. As we noted earlier, given the anatomical differences between these areas and observations from monkey electrophysiology about the roles of these areas, it is likely that significant differences exist in the nature of their visual responses. However, we do not have sufficient data to make conclusive claims of this nature.

One of the most striking examples of invariance that we observed was the cell that responded to visual pictures as well as the letter string with the name of the individual. In 18 of the 21 testing sessions, we presented patients with letter strings in addition to the different images. Eight of the 127 responsive units in these sessions showed a selective response to pictures of an individual as well as his/her name. Six of these were in the hippocampus, one in the entorhinal cortex, and one in the amygdala.

5.4 Discussion

We have shown that neurons in the MTL can respond in an invariant manner to several different representations of a particular individual. Given the diversity of images we used (pencil sketches, caricatures, letter strings, photographs taken from different

viewpoints, with different backgrounds, etc.) it is unlikely that this degree of invariance is based on common low-level features in the images. Instead, our results here strongly suggest that these neurons encode abstract representations of individuals, landmarks, and so on.

5.4.1 The MTL and high-level visual representations

What is the role of the MTL in storing such high-level visual representations? As we have mentioned previously, the MTL is the site at which information from different sensory areas converges. Given this anatomical basis, these structures could thus be prime candidates for combining information maintained at distinct cortical sites. Thus, these structures could link information that differs significantly in its format but with the same content (e.g. the name and the face of a particular individual). Neurons such as the one that responded to both color photographs of the Sydney Opera House and a letter string with its name, or Halle Berry and her name, clearly establish links between pictorial and semantic sources of information. Indeed, as we discussed in Chapter 4, cortical sites that respond to letter strings and words appear to be distinct from those where faces or buildings are represented, and our results suggest that information from these sites can be pooled together in these particular MTL cells.

On the other hand, the hippocampus and related structures can also form conjunctions between events and stimuli that should ordinarily have no relationship with each other (e.g. different contents and potentially the same format). This is compatible with the role of the MTL proposed by Squire and colleagues (Squire, Shimamura, & Amaral, 1989; Squire & Zola-Morgan, 1991). Examples of these responses are cells that fire to both Jennifer Aniston and Brad Pitt (Figure 5.9), or to Halle Berry and Catwoman (Figure 5.6). These are distinct concepts, and the fact that these cells can link them

together based on a high-level knowledge of the current affairs of Hollywood illustrates the role of the MTL in associative memory.

This role, however, is probably only temporary as is evidenced by the fact that very remote memories are unaffected by damage to the medial temporal lobe (Squire, 1992). As we discussed previously, the infero-temporal (IT) cortex is believed to be the site where more permanent storage occurs (Miyashita, 1993; Miyashita & Hayashi, 2000; Miyashita, 2004) and indeed it is now well known that complex stimuli such as ours can be represented in IT. The specific contribution of the MTL could be to maintain online but temporary representations for currently important stimuli. Indeed most of our cells were recorded around the time when Jennifer Aniston was on the cover of most popular magazines, when the movie Catwoman was released, or when the patient had recently visited the Sydney Opera House or the Tower of Pisa. Given that these stimuli had been recently encountered by the patients, it is possible, in accordance with the previously developed theories, that the MTL was engaged in establishing novel associations involving those persons and places to facilitate their later storage in IT. This is not to say that persons or places that have been familiar to the patient for several years will not be represented in the MTL. As long as they are currently relevant to the patient's life, current theories would predict that the MTL would need to encode them. However, they would also predict that items no longer relevant (such as a once-famous actor now forgotten by the media) might be stored in other areas but not in the MTL.[4]

[4] Thus the MTL structures might serve as a "temporary buffer" for information before it gets consolidated into long-term memory. Reports of patients with lesions to MTL structures could shed light on how long "temporary" could be. Patients with lesions restricted to the hippocampus suffer from retrograde amnesia that extends up to a couple of years (Kapur & Brooks, 1999). Lesions in other MTL structures such as the perirhinal and parahippocampal cortices are possibly involved in establishing memories over longer time scales since patients with lesions in these areas suffer from more severe retrograde amnesia.

5.4.2 Coding schemes

How neurons represent information is a hotly debated topic in neuroscience. One proposed scheme for neuronal representation, known as population coding, relies on the distributed activity of a large number of neurons. According to this hypothesis, individual neurons do not represent a particular object or concept. Rather, the relevant information is broadly distributed over a population of neurons and becomes available through their concerted activity (Georgopoulos, Kalaska, Caminiti, & Massey, 1982; Georgopoulos, Schwartz, & Kettner, 1986; Georgopoulos, 1987). Evidence for this coding scheme comes from neurons such as those found in the motor cortex. In this area, individual neurons are not finely tuned for a preferred direction of movement. Rather they have broad and overlapping tuning curves in three-dimensional space, which makes it impossible to accurately predict the direction of movement from the activity of any one neuron. However, by combining information from a population of neurons, movement directions can be predicted with much higher precision. Specifically, Georgopoulos and colleagues found that movement directions could be accurately predicted by a population code in which each neuron contributed its preferred direction, and this contribution was weighted by the strength of the neuron's response. Similar distributed codes have also been proposed for encoding continuous stimulus variables such as orientation.

In contrast to this distributed coding format, another strategy could rely on a local or sparser encoding (Konorski, 1967; Barlow, 1972; Thorpe, 1998; Gross, 2002; Olshausen & Field, 2004). According to this scheme, individual neurons play a more decisive role in the representation of an object or concept. In the extreme case, one individual neuron could signal a particular notion: for instance, the representation of my grandmother could rely on a single neuron in my brain. This neuron would be activated

whenever I see my grandmother, independent of the viewpoint, the distance, or her emotional expression, etc. Additionally, this neuron would not fire in response to other concepts even if they were related to her (e.g. my grandfather or another elderly woman). This extreme version of the sparse coding scheme has been sarcastically termed the 'grandmother scheme.'

Nevertheless, cells in higher-level visual areas such as IT cortex do provide support for the notion of highly specific representations. In the 1970s, Gross and colleagues first reported the existence of cells in IT that fired selectively to hands and faces (Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda et al., 1972). Since then, a number of different studies have confirmed and extended these findings and provided further evidence for the fact that individual neurons can specifically represent particular objects (Perrett, Rolls et al., 1982; Rolls, 1984; Perrett, Smith et al., 1985; Rolls & Baylis, 1986; Perrett, 1987; Yamane, Kaji, & Kawano, 1988). More recently, it has been shown that cells can be trained to show a high degree of specificity for the representation of complex objects, such as paperclips or computer-generated fractal patterns (Logothetis & Sheinberg, 1996; Tanaka, 1996).

The coding strategy of the cells that we have presented here seems to be more in line with the latter scheme. We have observed neurons that maintain an invariant response to several views of an object, thereby explicitly signaling the presence of that object. Further, in contrast to population coding, which predicts that during any percept a large number of cells must be active, we find that most neurons do not respond to the majority of images seen by the patients.

This is not to say, however, that a strict, sparse, explicit and invariant coding strategy is implemented by these neurons such that single neurons must code for one and only one percept. Indeed we do find instances of cells that respond to more than one object. Additionally, given the limited time we have for these experiments, we cannot

explore a very large portion of the stimulus space. It could very well be that with slightly longer recording sessions, we would find that many of our cells represented several different concepts. Furthermore, on purely theoretical grounds, a pure grandmother coding strategy would be neither feasible nor robust. If every person, object, place, and animal were to be represented by a dedicated neuron, the brain would soon reach the limits of its capacity to encode information. And if the neuron representing the concept of my grandmother disappeared, would I become unable to recognize my grandmother?

So what coding scheme do we observe in the MTL? In our results, we have observed that individual neurons can carry information in a sparse, invariant, and explicit, although somewhat redundant, manner. However, if only for theoretical reasons, this strategy cannot apply to an infinite number of concepts. Rather, it is more likely that sparse, explicit representations like the ones we have observed here are only formed, through repeated exposure, for highly familiar concepts. The processing of less frequently encountered objects must rely on a more distributed coding scheme.

5.4.3 Explicit representations and the Neuronal Correlates of Consciousness

Zeki first suggested the concept of “essential nodes,” the idea that specific regions in the brain were responsible for supporting awareness of particular aspects of vision (Zeki & Bartels, 1999; Zeki, 2001). This suggestion was based on studies with neurological patients with whom it was observed that when particular areas were lesioned, consciousness of certain features was lost. Thus for example, a lesion in area V5/MT would result in a loss of motion perception, or a lesion in the FFA would result in prosopagnosia (the inability to recognize faces). Neurons in these areas would be expected to explicitly encode the relevant feature in their firing rates. That is, it should be possible based on the firing rates of these neurons to decide if that particular feature

was present in the stimulus. This, in turn, implies that these neurons “should be invariant to those aspects of the input that do not convey any specific information about the feature symbolized” (p. 27 (Koch, 2004)). It has been proposed that such explicit representations could form the basis for the neuronal correlates of consciousness (NCC) (Crick & Koch, 2003). More specifically, these representations could be a necessary condition for the NCC in that if such representations do not exist, the corresponding percept cannot reach consciousness.

The MTL neurons we have presented in this chapter seem to meet the requirements for explicit representations: they specifically respond to particular concepts (such as “Halle Berry” or the “Sydney Opera House”), and because of their selectivity, the presence of these stimuli can be deduced from the firing rates of these neurons (see for example the ROC curves). Moreover, these responses are invariant to most changes made to the pictures of the individuals, landmarks, etc. In particular, these neurons also respond to the written names of these objects. Given these high-level, abstract, and explicit representations, these cells are candidates of choice for participating in the NCC of these notions.

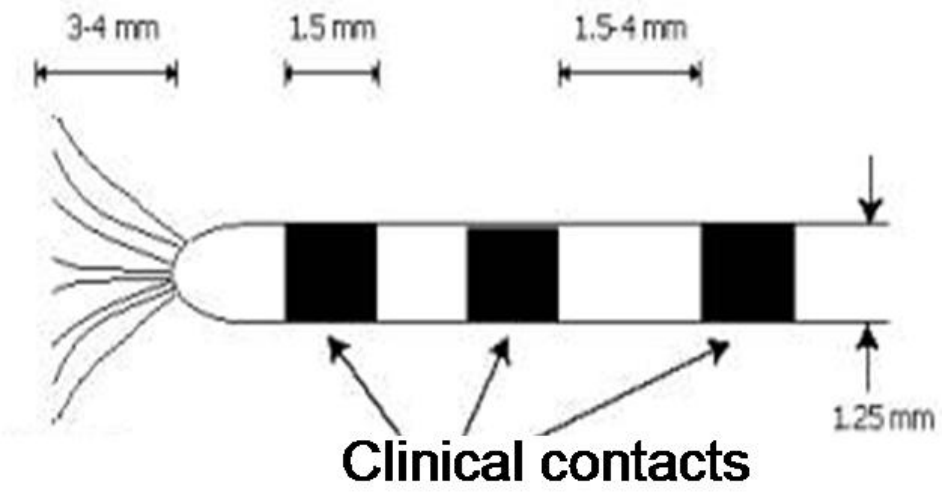

Figure 5.1 An example of the electrodes used. The electrodes were 1.25 mm in diameter. Platinum contacts (1.5 mm in length) along the electrodes were used to collect clinical data. Through each electrode 9 microwires (including a reference) were inserted. The microwires extended 3–4 mm from the tip of the electrode, and lay within a cone subtending 45º. From (Kreiman, 2002).

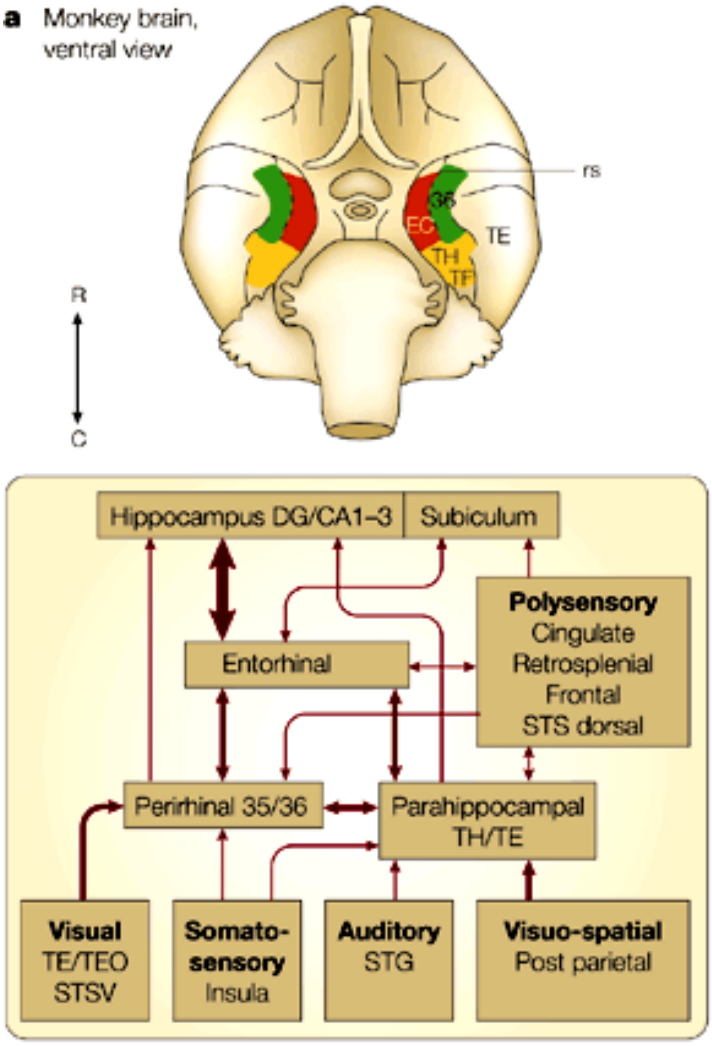

Figure 5.2 Schematic of the anatomy of the MTL in the monkey brain. a) The relative locations of the perirhinal cortex (areas 35 and 36), parahippocampal cortex (TF/TH), entorhinal cortex, and TE in a monkey brain. b) The connection diagram shows the routes by which sensory information projects to the cortical regions and then enters the hippocampus. The thickness of the arrows indicates the size of the projection. Note that several projections are reciprocal. The amygdala is now shown in this figure. (Adapted with permission from Brown & Aggleton, 2001). Copyright, Nature Publishing Group.

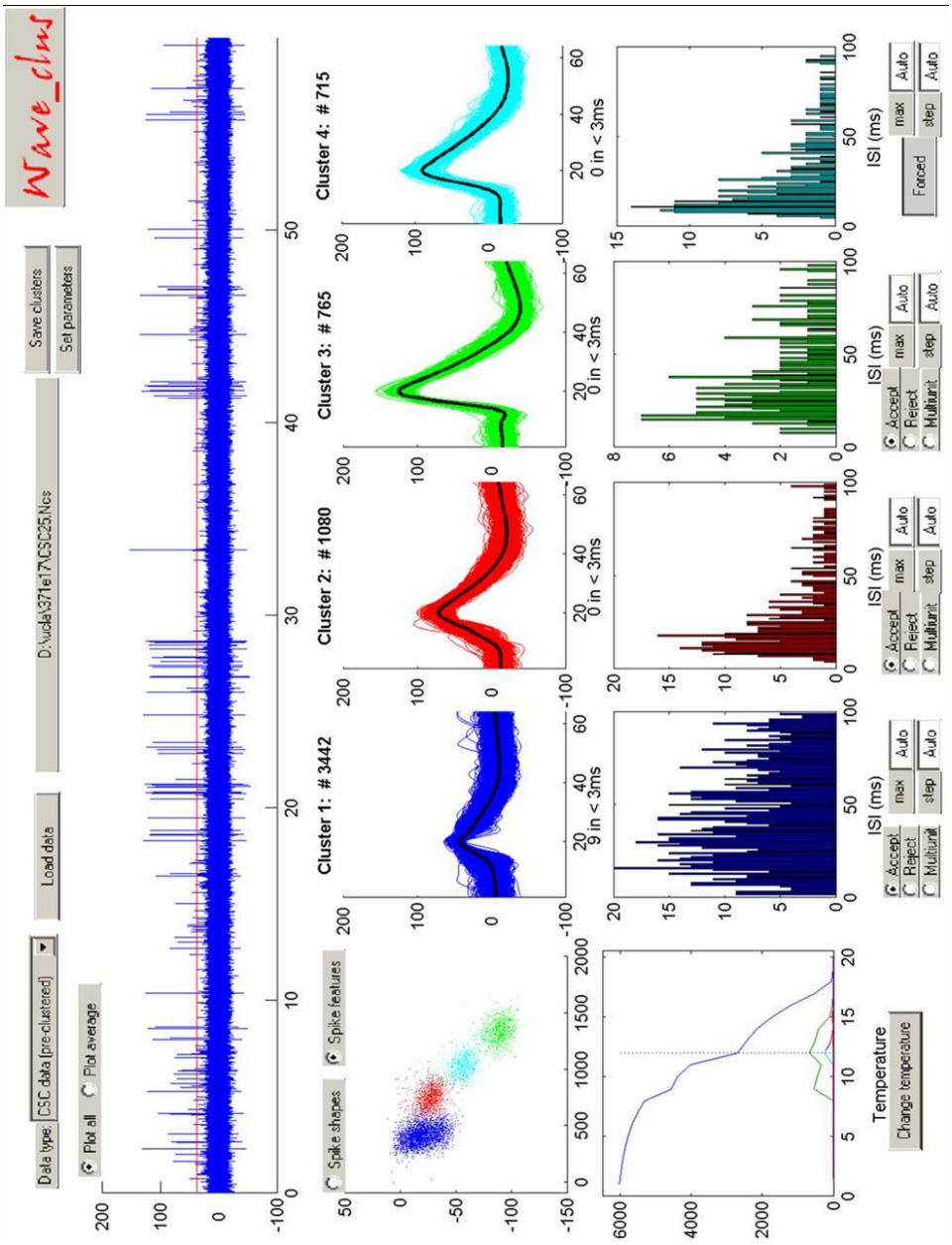

Figure 5.3. The output of the spike sorting algorithm. Upper plot: 60 seconds of continuous data, band-pass filtered between 300 and 3000 Hz. The amplitude of the signal is in μV. The red line represents the amplitude threshold used for spike detection. Middle plots: Distribution of the wavelet coefficients in a 2-D space for four different clusters and the spike shapes corresponding to these clusters. Lower plots: The leftmost plot shows the number of spikes in each cluster as a function of the temperature. The optimal temperature chosen by the algorithm in an unsupervised manner is represented by the dotted line. The other 4 plots correspond to the ISI distribution for the 4 clusters.

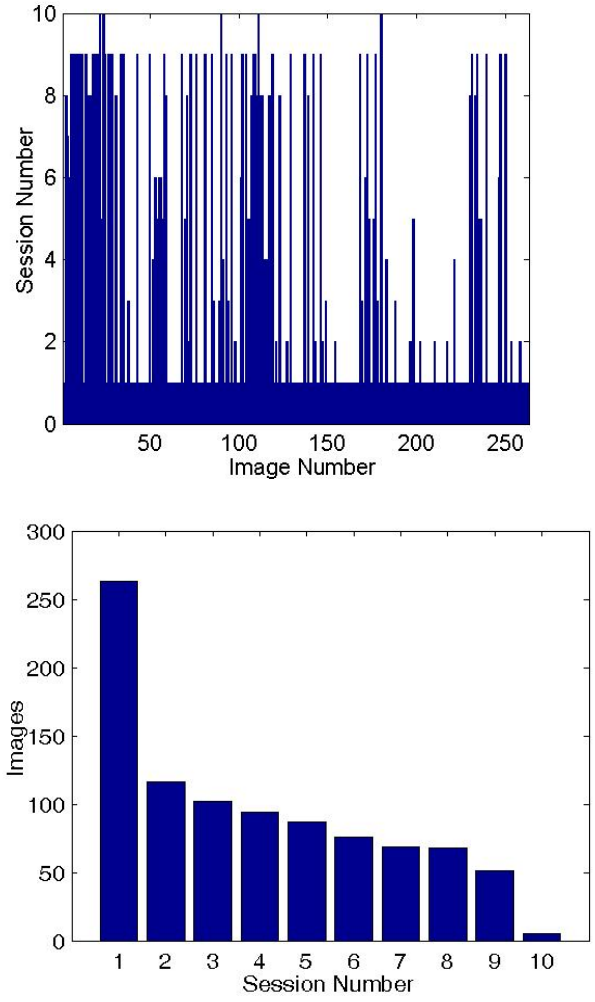

Figure 5.4 Distribution and overlap of images shown in the screening sessions. a) There were 261 unique images (x-axis) presented in 10 Screening Sessions (y-axis). Each bar in the figure represents which session a particular image was shown in. b) The number of images that were shown repeatedly in *at least* X sessions where X = [1 10]. Thus for instance, only 5 images were shown in all 10 sessions, and approximately 55 were shown in at least 9, etc.

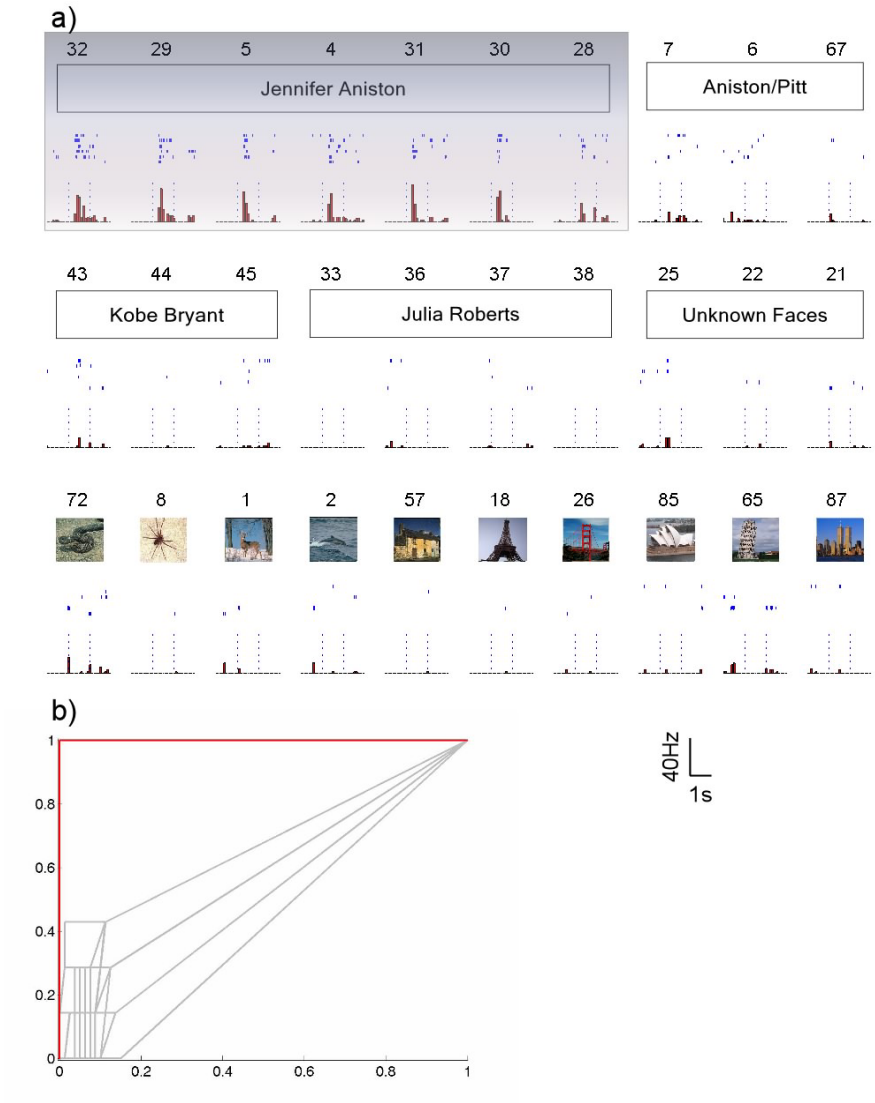

Figure 5.5 Responses of a single unit in the left posterior hippocampus that responds to Jennifer Aniston. a) Responses to 30 out of 87 presented pictures. For each picture presented, the corresponding raster plots (on each of the 6 trials) and post-stimulus time histograms are shown. The vertical dashed lines represent stimulus onset and offset. Stimuli were presented for 1 s. This cell fired exclusively to 7 different pictures of the actress Jennifer Aniston. Note that the cell does not fire to pictures of Aniston with Brad Pitt (images 7, 6, 67). b) The associated ROC curve (red line) and the 99 surrogate curves (grey lines; due to overlap not all 99 curves are visible). The area under the curve is 1.0. Note that in this and other figures in this chapter, we cannot reproduce the original images of celebrities shown to patients due to copyright issues. The white boxes with text in black represent different images of a particular individual that were shown to the patients.

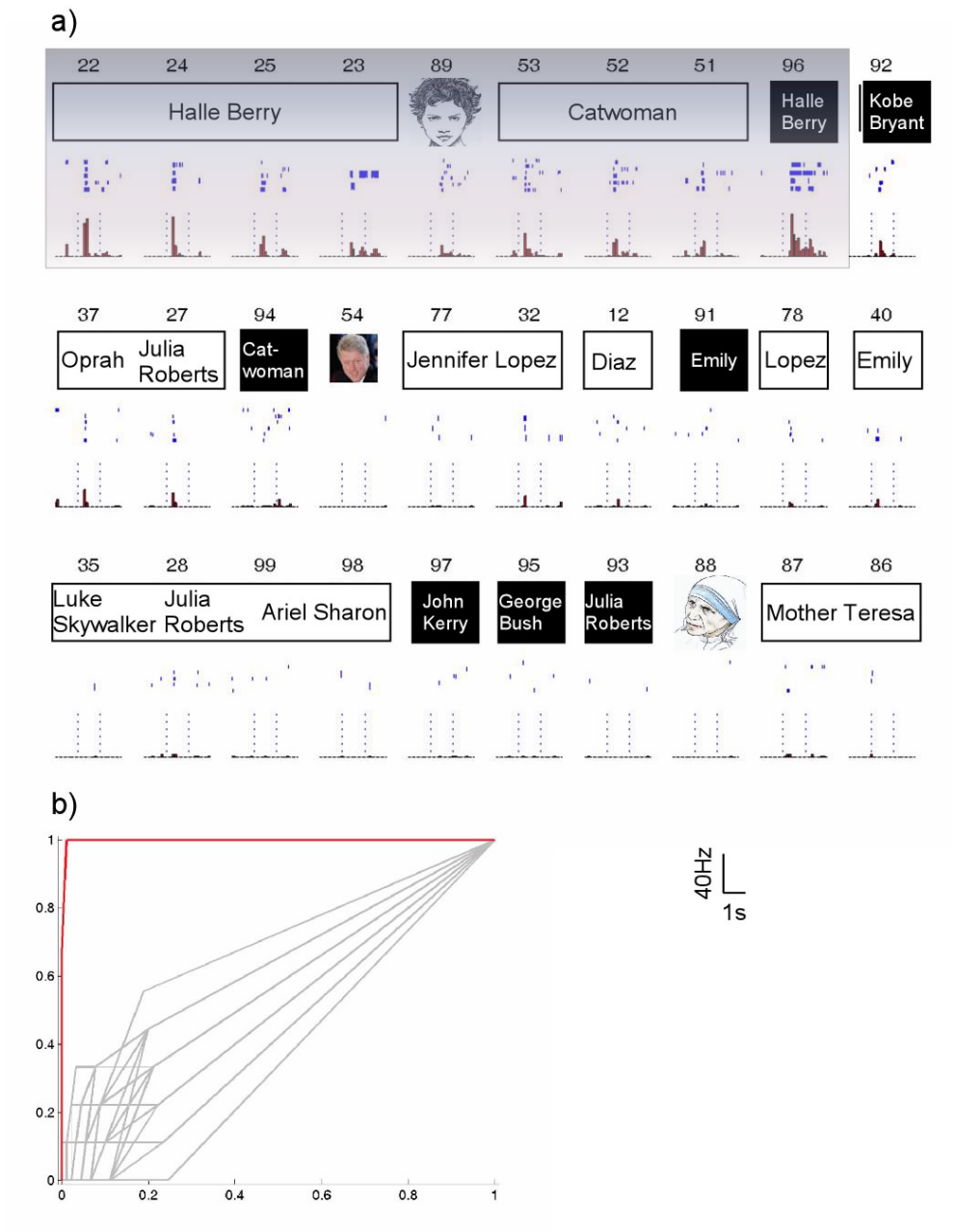

Figure 5.6 Responses of a single unit in the right anterior hippocampus that responds to pictures of Halle Berry. a) This cell responds not only to images of Halle Berry, but also to a caricature of her, images of her dressed as Catwoman, and to a letter string of her name (image 96). All black squares with text in white represent image strings shown to the patient. The format of this figure is the same as above. b) The associated ROC curve (red line) and the 99 surrogate curves (grey lines). The area under the curve is 0.99.

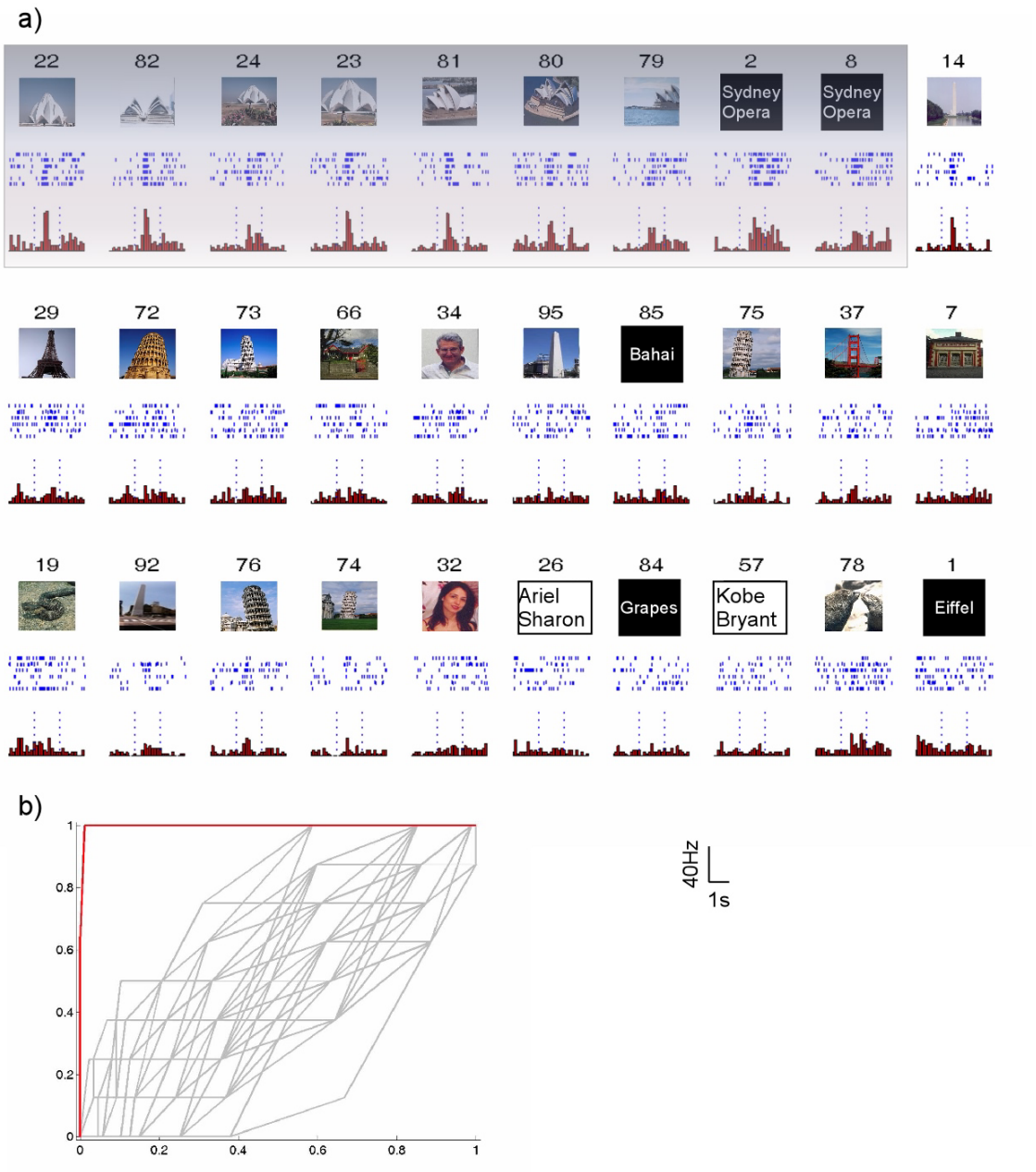

Figure 5.7 Responses of a multi-unit in the left anterior hippocampus that responds to pictures of the Sydney Opera House and the Bahai Temple. a) The format of this figure is the same as above. (The patient identified the Bahai Temple as the Sydney Opera House.) This cell also responded to the text “Sydney Opera” (images 2 and 8). b) The associated ROC curve (red line) and the 99 surrogate curves (grey lines). The area under the curve is 0.97.

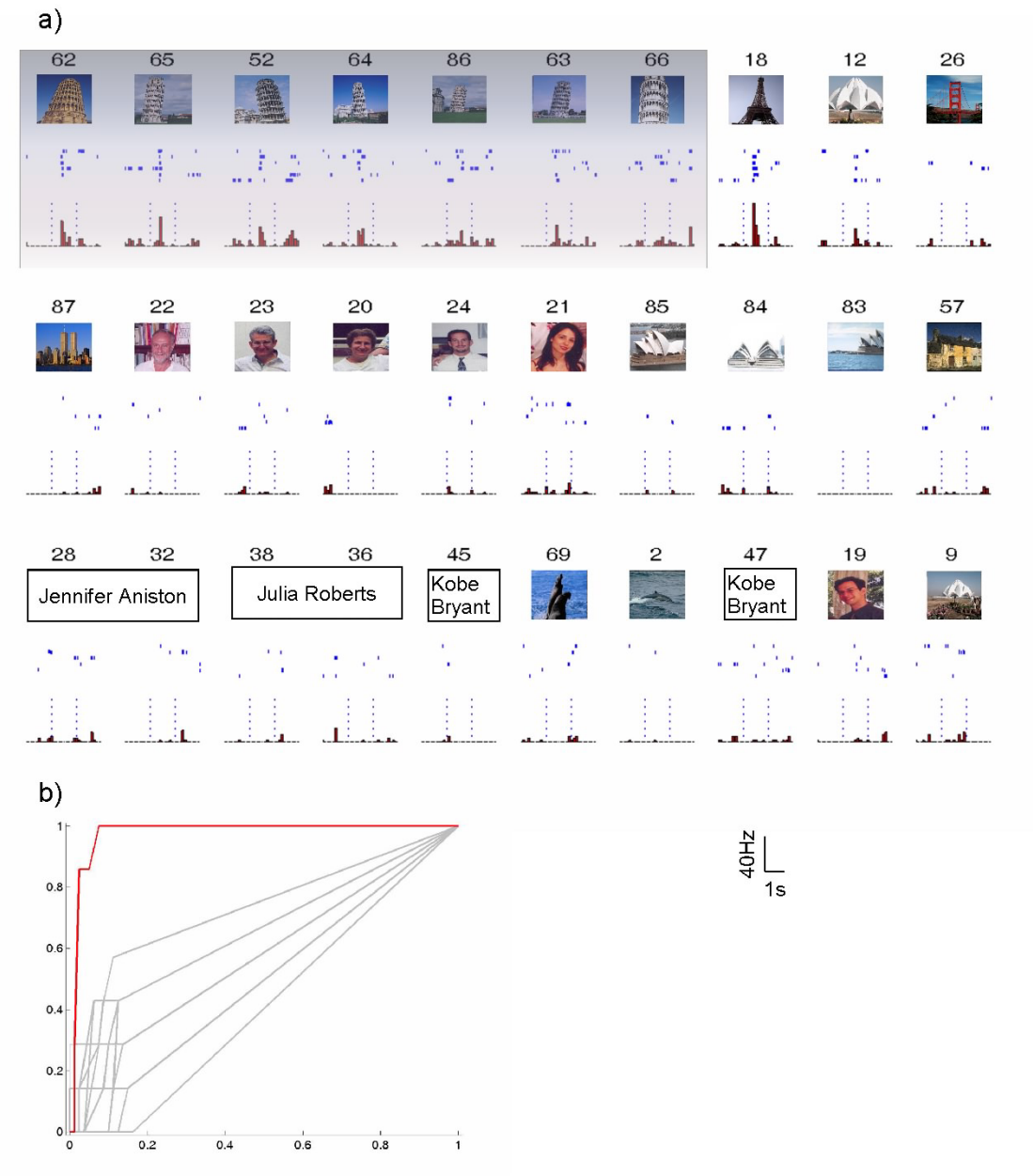

Figure 5.8 Responses of a multi-unit in the right posterior hippocampus that responds to pictures of the tower of Pisa. a) This unit responded to 7 different images of the Tower of Pisa, and also to one picture of the Eiffel Tower. The format of the figure is the same as above. b)The area under the ROC curve is 0.98.

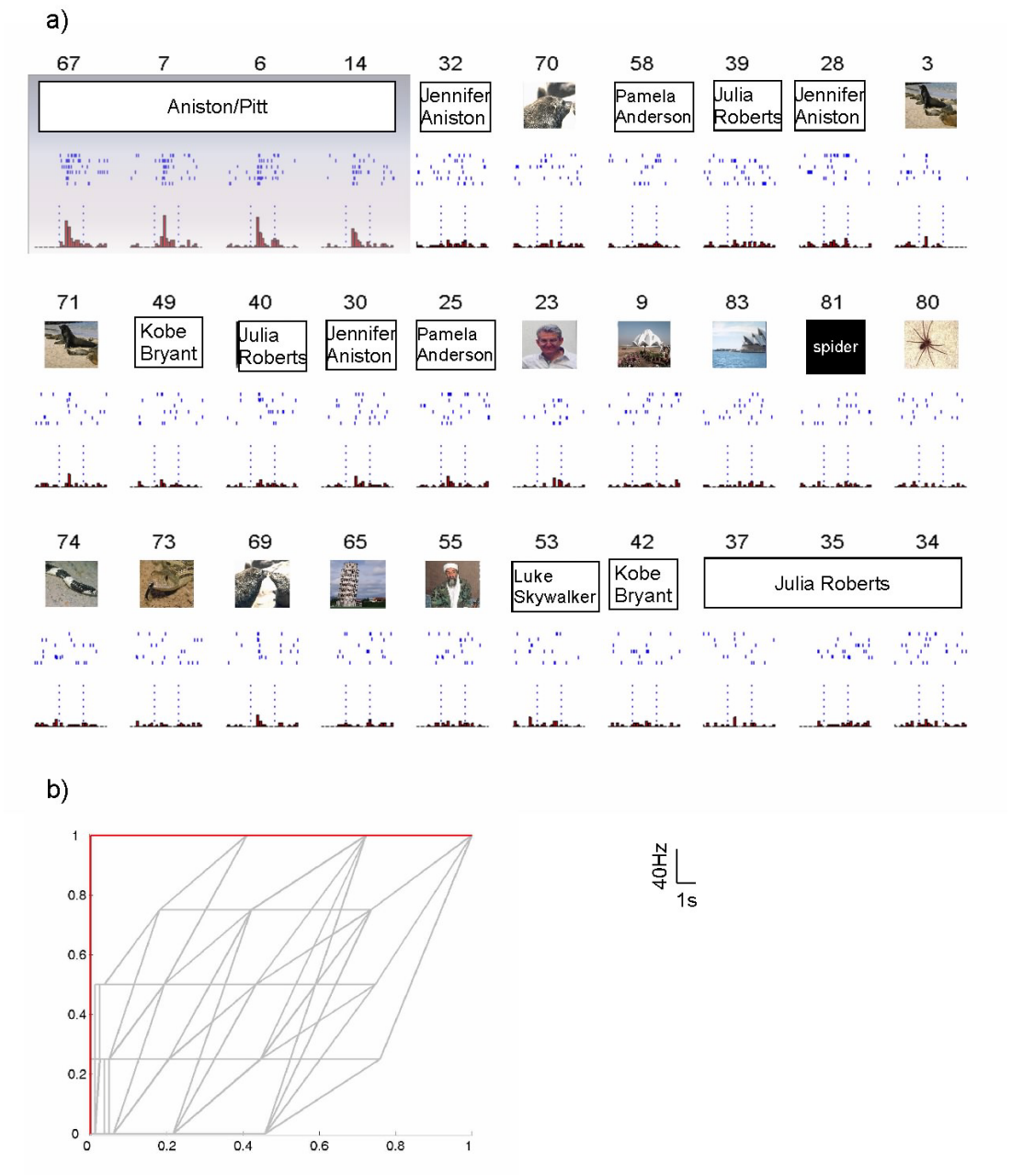

Figure 5.9 Responses of a multi-unit in the right posterior hippocampus that responds to pictures of Brad Pitt and Jennifer Aniston together, but not to images of either of them alone. a) This unit was recorded in the same patient and session as the cell shown in Figure 5.5. The format of this figure is the same as above. b) The area under the ROC curve is 1.0.

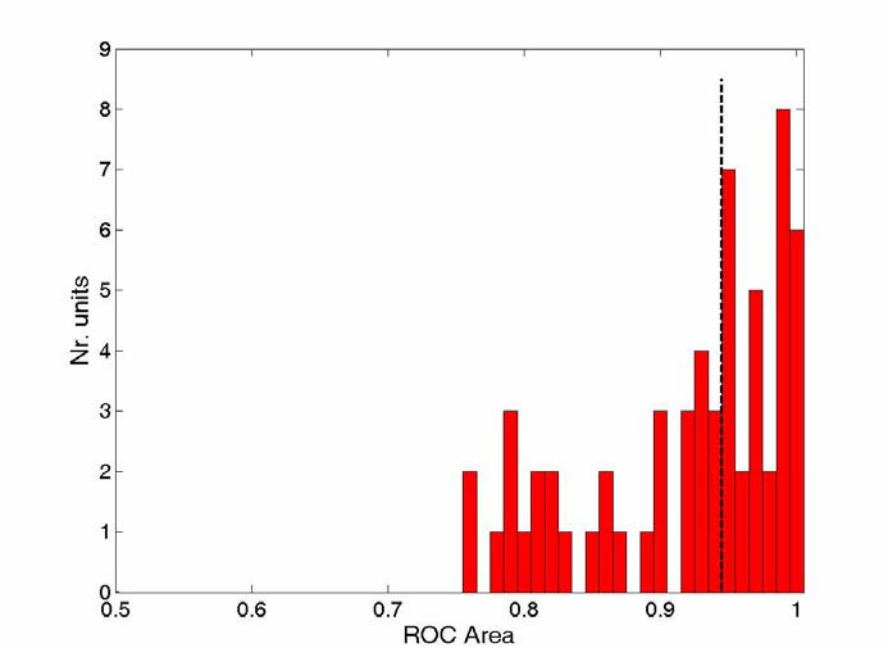

Figure 5.10 Distribution of areas under the ROC curves for the 52 units showing invariant representations. Of these cells, 44 responded to a single individual or object and 8 to 2 individual or objects. The dashed vertical line marks the median of the distribution (0.94).

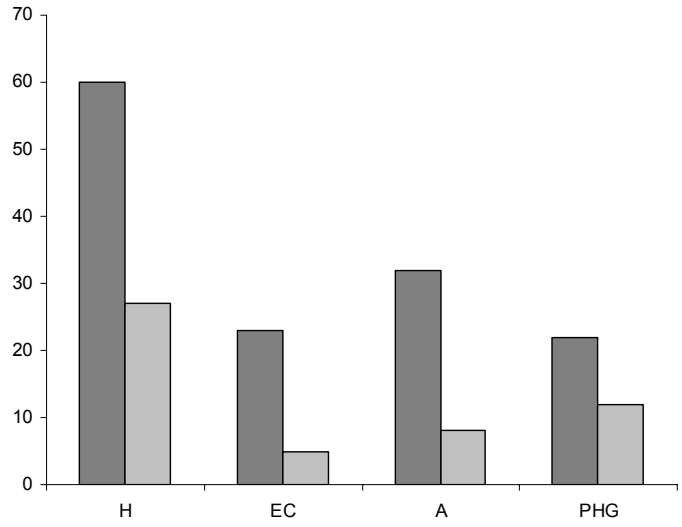

Figure 5.11 Location of the responsive and invariant cells. The dark grey bars correspond to the visually responsive cells found in each brain area. The light gray bars correspond to the number of invariant units. H is the hippocampus, EC entorhinal cortex, A amygdala, and PHG the parahippocampal gyrus. Thus for instance, 60 of our visually responsive cells were located in the hippocampus, and of these 27 were invariant to particular individuals or buildings, etc.

Table 5. 1 Location of electrodes during screening and testing sessions

a) Location of Electrodes during Screening Sessions

Amygdala	17
Entorhinal Cortex	12
Hippocampus	18
Parahippocampal Gyrus	6
Orbito-frontal Cortex	5
Occipital Cortex	4

b) Location of Electrodes during Testing Sessions (Note: there were more Testing sessions, hence the comparatively higher number of electrodes in each location).

Amygdala	38
Entorhinal Cortex	29
Hippocampus	45
Parahippocampal Gyrus	15
Occipital Cortex	6
Transverse Gyrus	8
Anterior Cingulate	2
Supplementary Motor Area	1