Behavioral/Systems/Cognitive

Plasticity of Temporal Pattern Codes for Vocalization Stimuli in Primary Auditory Cortex

Jan W. H. Schnupp, Thomas M. Hall, Rory F. Kokelaar, and Bashir Ahmed
University Laboratory of Physiology, University of Oxford, Oxford OX1 3PT, United Kingdom

It has been suggested that "call-selective" neurons may play an important role in the encoding of vocalizations in primary auditory cortex (A1). For example, marmoset A1 neurons often respond more vigorously to natural than to time-reversed twitter calls, although the spectral energy distribution in the natural and time-reversed signals is the same. Neurons recorded in cat A1, in contrast, showed no such selectivity for natural marmoset calls. To investigate whether call selectivity in A1 can arise purely as a result of auditory experience, we recorded responses to marmoset calls in A1 of naive ferrets, as well as in ferrets that had been trained to recognize these natural marmoset calls. We found that training did not induce call selectivity for the trained vocalizations in A1. However, although ferret A1 neurons were not call selective, they efficiently represented the vocalizations through temporal pattern codes, and trained animals recognized marmoset twitters with a high degree of accuracy. These temporal patterns needed to be analyzed at timescales of 10–50 ms to ensure efficient decoding. Training led to a substantial increase in the amount of information transmitted by these temporal discharge patterns, but the fundamental nature of the temporal pattern code remained unaltered. These results emphasize the importance of temporal discharge patterns and cast doubt on the functional significance of call-selective neurons in the processing of animal communication sounds at the level of A1.

Key words: **sound; training; behavior; auditory cortex; vocalization; temporal coding**

Introduction

Recent years have seen a renewed interest in the processing of complex sounds and animal vocalizations in the central auditory system (Rauschecker, 1998; Nelken et al., 1999; Wang, 2000; Sen et al., 2001; Romanski et al., 2005), and much research has been aimed at trying to identify specific mechanisms within the mammalian auditory pathway that might facilitate the processing of vocalization calls. Lesion studies indicate that the primary auditory cortex (A1) is essential for the recognition of communication sounds (Heffner and Heffner, 1986). Early studies into the representation of vocalization sounds were optimistic about finding specific "call detectors" (Winter and Funkenstein, 1973) within the auditory cortex, but, although many neurons in A1 respond vigorously to species-specific vocalizations, previous research suggests that these responses may often be explained in terms of sensitivity to relatively simple acoustic features of the sounds (Pelleg-Toiba and Wollberg, 1991; Gehr et al., 2000). Nevertheless, a recent study by Wang et al. (1995), which compared responses of neurons in A1 of the common marmoset (*Callithrix jacchus*) to natural and time-reversed marmoset "twitter" vocalizations, found that marmoset A1 neurons often responded much more vigorously to natural than to time-reversed vocalizations. When neurons in A1 of the cat were tested with the same marmoset vocalization stimuli (Wang and Kadia, 2001), no response selectivity for natural over time-reversed marmoset calls was seen. It is tempting to interpret the selectivity for the natural marmoset twitter calls in marmoset A1 as a specialization that facilitates a "categorical" response that might aid in the "recognition" of an ecologically important stimulus. Such a specialization might be innate or acquired through sensory experience. However, if response specificity is a prerequisite for the recognition or discrimination of natural marmoset calls, one would have to conclude that A1 neurons of carnivores, such as the cat, which respond just as vigorously to forward as to time-reversed marmoset calls, would not represent these unfamiliar vocalizations in an effective manner. Furthermore, one might expect call selectivity for marmoset calls to be induced in A1 of carnivores if they were trained to recognize these calls. Alternatively, higher-level features of the neural response, such as temporal discharge patterns, might carry sufficient stimulus-related information to support efficient processing and recognition of these vocalizations, although the neurons appear "nonselective" in their overall firing rates. To investigate these possibilities, we recorded responses to the same natural and time-reversed marmoset twitters used by Wang and colleagues (Wang et al., 1995; Wang and Kadia, 2001) from A1 of five "naive" adult ferrets who had never been exposed to marmoset vocalizations before the recording experiments, and we analyzed these responses for stimulus-related information carried in the neural discharge patterns. Furthermore, to assess whether response specificity to natural marmoset calls in A1 can

Received Oct. 11, 2005; revised March 5, 2006; accepted March 10, 2006.

This work was supported by Biotechnology and Biological Sciences Research Council Project Grant Nr 43/S19595 (J.W.H.S.) and a Defeating Deafness vacation scholarship (T.M.H.). We give particular thanks to Dr. X. Wang (Johns Hopkins University School of Medicine, Baltimore, MD) for providing recordings of marmoset vocalization and to Robert Campbell and Jenny Bizley (both from University Laboratory of Physiology, University of Oxford, Oxford, UK) for their help with electrophysiological recordings.

Correspondence should be addressed to Jan W. H. Schnupp, University Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK. E-mail: jan.schnupp@physiol.ox.ac.uk.

DOI:10.1523/JNEUROSCI.4330-05.2006

 0270-6474/06/264785-11$15.00/0

be induced through experience or training, we also recorded electrophysiological data from two ferrets that had been trained to discriminate the natural marmoset twitter calls from a host of other natural sounds in a positive reinforcement Go/No-go paradigm.

Materials and Methods

All experiments were approved by the local ethical review committee and were performed under license from the United Kingdom Home Office in accordance with the Animal (Scientific Procedures) Act of 1986.

Electrophysiological recording. All electrophysiological data were obtained in a sound-insulated chamber (Industrial Acoustics Company, Winchester, UK). The animals were anesthetized by a 2 ml/kg intramuscular injection of alphaxalone/alphadolone acetate (Saffan; Schering-Plough Animal Health, Welwyn Garden City, UK). The left radial vein was cannulated, and anesthesia was maintained by continuous infusion of medetomidine (Domitor; Pfizer, Walton Oaks, Surrey, UK) and ketamine (Ketaset; Fort Dodge Animal Health, Overland Park, KS) at a typical rate of 0.022 and 5.0 $mg \cdot kg^{-1} \cdot h^{-1}$ respectively, along with 5 ml/h saline supplemented by 5% glucose. Expired CO_2, electrocardiogram, and muscle tone were carefully monitored, and anesthetic infusion rates were adjusted as required to ensure stable anesthesia throughout. The animals were also artificially ventilated with oxygen-enriched air.

The parietal and left temporal aspects of the skull were exposed, and a stainless steel head holder was fixed to the skull with stainless steel screws and dental acrylic above the midsagittal ridge. The auditory cortex was exposed by craniotomy and removal of the dura. Mineral oil was applied to the exposed pial surface to prevent dehydration. Acoustic stimuli were presented diotically at a sound level of ~68 dB sound pressure level (SPL) using Tucker Davis Technologies (TDT) (Alachua, FL) System3 digital signal processors and custom headphones based on Panasonic (Bracknell, UK) RPHV297 drivers. The three natural marmoset twitter stimuli used here were identical to those used previously by Wang and colleagues (Wang et al., 1995; Wang and Kadia, 2001).

To characterize neural responses, each twitter call and its time-reversed counterpart were presented 20 times in a pseudorandom order. The responses to these stimuli were recorded using 2 MΩ 4 × 4 silicon array "Michigan probes" (Center for Neural Communication Technology, University of Michigan, Ann Arbor, MI), bandpass filtered (300 Hz to 3 kHz), and digitized at 25 kHz using a TDT Pentusa multichannel recording system. BrainWare software (TDT) was used to control stimulus presentation and data collection and to extract single-unit "clusters" of action potentials from the electrode signal. Only units responding with a mean rate of at least 2 Hz were included in the additional analysis. Spike-timing data from acoustically responsive units were then exported to Matlab (MathWorks, Natick, MA) for additional analysis.

Behavioral training. To study the effect of experience, two male adult ferrets were trained to distinguish the three marmoset twitter calls from a set of eight other sounds (downloaded from the internet), which included the bark of a coyote (*Canis latrans*), the song of a northern cardinal (*Cardinalis cardinalis*), the chirping of a katydid (*Orchelium vulgare*) and of a cricket (*Euscyrtus concinnus*), and the vocalizations of a bottlenose dolphin (*Tursiops truncatus*), a guinea pig (*Cavia porcellus*), and a killer whale (*Orcinus orca*), as well as a 1 s broadband noise burst. Figure 1 shows the spectrograms of the natural sound stimuli used. The choice of the non-marmoset training stimuli was to some extent arbitrary, except that the set of these sounds was deliberately chosen to be an acoustically very diverse sample of natural sounds. For a human observer, each of these stimuli is relatively easy to distinguish from the marmoset twitter calls, but the stimulus attributes that most obviously distinguish each from the twitters are quite different in each case. The coyote and guinea pig sounds, for example, all have a syllabic structure that imposes amplitude modulations not too unlike those of the marmoset calls, but they do not extend quite as high in frequency range and differ in pitch from the twitters to greater or lesser extent. The dolphin and cricket sound, in contrast, have a very different syllabic structure but overlap with practically the entire frequency range of the twitter calls and are more similar to the twitters in pitch. Thus, the "perceptual attributes" that most clearly distinguish the No-go stimuli from the Go stimuli vary from case to case. Thus, the task of discriminating these sounds from a marmoset twitter is likely to be not too dissimilar from the acoustic discriminations a marmoset has to perform when identifying a conspecific call against a background of communication sounds from other animals.

Training was performed in a custom-built behavioral training chamber that was made from a plastic pet carrier fitted with a Visaton FRS 8 loudspeaker (Visaton, Haan, Germany) and two custom water spouts mounted on mechanical switches, one "start spout" and one "reward spout." The behavioral paradigm was controlled by a computer running custom-written Matlab code and connected to a TDT RM1 laboratory interface. During training periods, the animals were put on a water-restricted diet, i.e., they received only dry food in their home cages and had access to drinking water only during the twice-daily behavioral training sessions. Animals initiated a trial by pushing the start spout with their snout or their forepaw. This triggered the release of a small drop of water (~0.025 ml), followed immediately by the presentation of one of the sound stimuli chosen at random. Sound levels in the testing chamber were ~68 dB SPL. If the stimulus was one of the three marmoset twitter calls (a Go stimulus), the animals were expected to respond by pushing the reward spout, and doing so within 8 s from the onset of the Go stimulus triggered the delivery of a water reward (~0.2 ml). In an initial period lasting ~2 weeks, during which the animals had to learn the significance of the start and reward spouts, the animals were presented only with Go stimuli during their training sessions. Thereafter, initially a set of three to four No-go stimuli was added for an additional 3–4 weeks training before the final full set of eight No-go stimuli was introduced. At each trial, stimuli were chosen randomly, but, to keep the animals motivated, we found it advantageous to adjust the frequency of Go stimuli according to the individual animal's temperament, to lie between 50 and 66% of all trials. If the stimulus was a No-go stimulus (i.e., any sound other than one of the three marmoset twitter calls), then the animals were expected to ignore the stimulus and were free to initiate the next trial immediately by pressing the start spout again. To encourage the animal to listen to the stimuli rather than just performing rapid but arbitrary random choices, responses made <0.9 s after the onset of the stimulus were ignored. Incorrect responses (i.e., pressing the reward spout after a No-go stimulus or failing to press the reward spout after a Go stimulus) triggered a brief timeout of 4 s to provide negative feedback. The number of trials obtained from the animals in each training session was variable, but ~100–150 trials per training session were typical. To minimize the risk of animals becoming dehydrated during training periods, small water supplements were given at the end of a day of training if an animal had collected less than a minimum of 35–50 ml (depending on body weight) of water rewards during the training of that day. Also, after a training period of at most 14 d, animals were given rest periods of at least 2 d (but sometimes considerably longer), during which they had access to *ad libitum* water in their home cages. In this manner, animals were trained over a period of >4 months before electrophysiological recording.

Results

Electrophysiology in naive animals

We characterized responses from 142 single units from A1 of five adult ferrets. As a first step, we analyzed these units in the manner identical to that used by Wang and colleagues (Wang et al., 1995; Wang and Kadia, 2001). The mean spike rates of each unit in response to each natural (R_{Nat}) and time-reversed (R_{Rev}) twitter call were calculated over a 1200 ms period from stimulus onset, and the corresponding "selectivity index" (d) was calculated using the formula $d = (R_{Nat} - R_{Rev})/(R_{Nat} + R_{Rev})$.

Figure 2 shows scatter plots of R_{Nat} against R_{Rev} for all of the cells in our study, with responses to the first, second, and third exemplar of the twitter stimuli. The spectograms of these stimuli are shown at the top left of Figure 1. If we assume that ferret A1 units are not selective for natural marmoset calls, then we expect that the data in these scatter plots should clusters more or less tightly around the main diagonal ($R_{Nat} = R_{Rev}$), and the distribu-

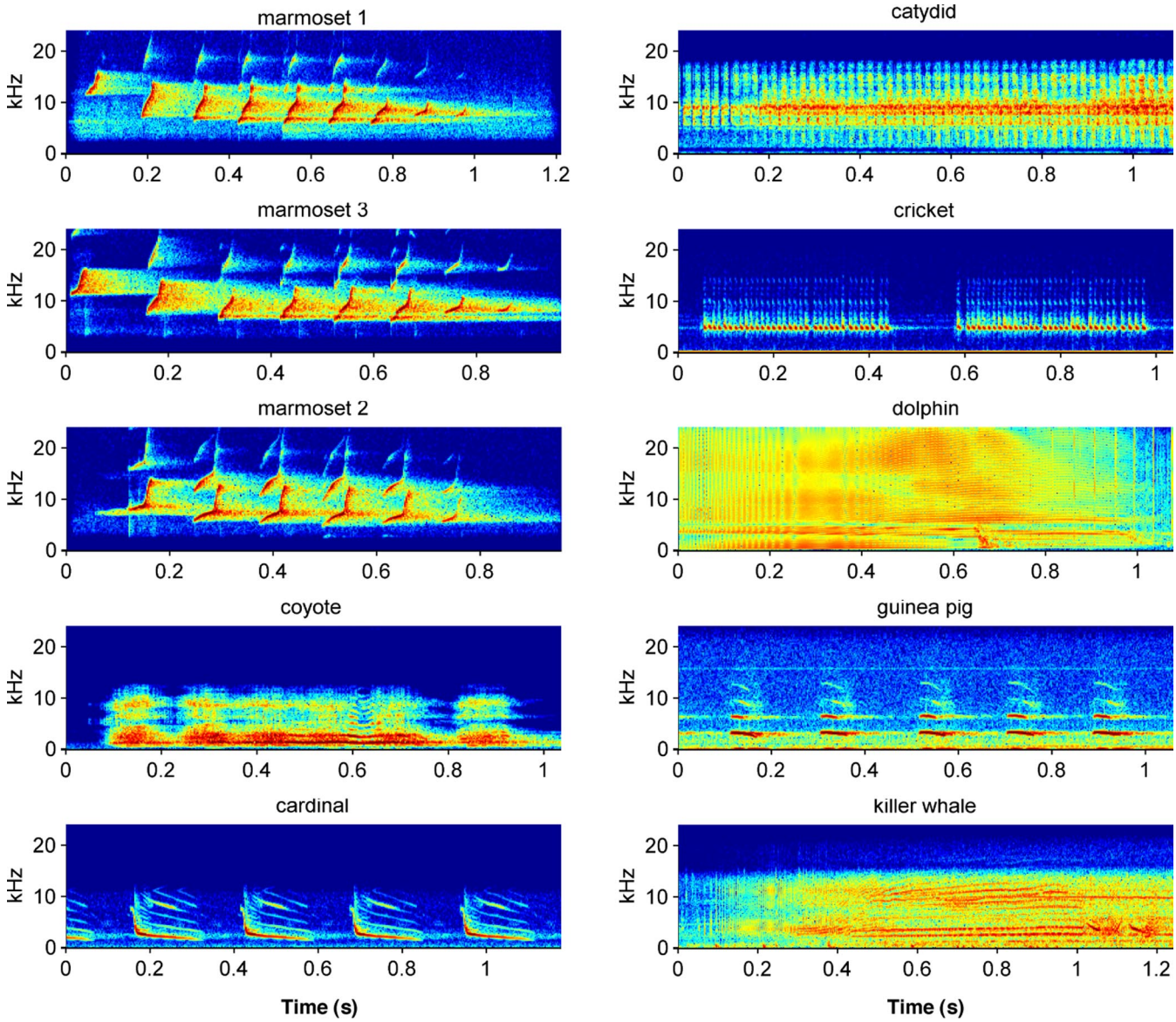


Figure 1. Spectrograms of the natural sound recordings used as stimuli for this study. The color scale saturates over 70 dB.

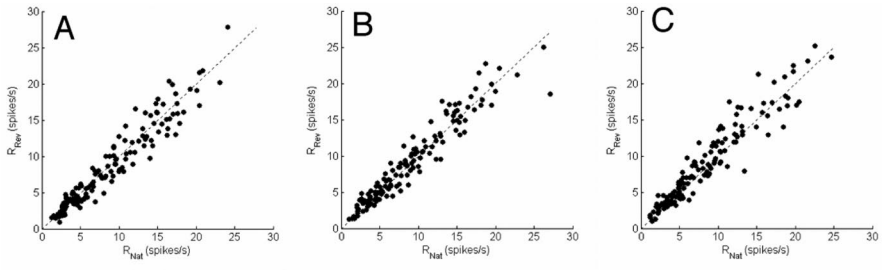


Figure 2. Mean response rate evoked by natural (R_{Nat}) plotted against response to time-reversed (R_{Rev}) twitter call stimulus. Responses for first, second, and third exemplar of the twitter stimuli are shown in ***A***, ***B***, and ***C*** respectively.

tion of the corresponding values of d should have a mean very close to 0. This expectation is very much borne out. The means ± SD of the values for d obtained for each of the three stimuli were 0.030 ± 0.101, −0.003 ± 0.093, and −0.020 ± 0.099, respectively. The overall median value for d (0.002) was not significantly different from 0 (sign test, $p = 0.74$). The selectivity index values we obtained in the ferret are comparable with those reported by Wang and Kadia (2001) for cat A1 (0.047 ± 0.265, 0.086 ± 0.238, and 0.068 ± 0.372, respectively) and are smaller by an order of magnitude or more than those reported for the marmoset (0.479 ± 0.361, 0.335 ± 0.302, and 0.385 ± 0.340). Preferences for natural over time-reversed marmoset calls in ferret A1 are therefore tiny or 0, and a coarse rate-code-based scheme would have difficulty distinguishing natural from time-reversed calls on the basis of ferret A1 responses.

Encoding of twitter calls in temporal spike patterns

When analyzed in terms of mean spike rate over the entire stimulus duration, units in naive ferret A1, like those in cat, typically lack "selectivity" for natural marmoset twitter calls. Like previous authors (Wang et al., 1995; Gehr et al., 2000; Wang and Kadia,

2001), we studied call selectivity in an anesthetized preparation, and, although we have at present no reason to assume that rate-based call selectivity in A1 is radically altered by anesthesia, future studies in awake preparations will be needed to confirm this experimentally. To human observers, the natural and time-reversed twitter calls sound quite different. We do not know whether ferrets also perceive these sounds differently, but, if they do, then any information that might underpin any perceptual differences between the stimuli would most likely be either carried in a temporal pattern code or processed in a neural pathway that bypasses A1. When we plotted the neural responses in a standard raster plot format, we noticed that ferret A1 units varied considerably in their ability to respond to the marmoset twitter calls with distinctive, reproducible discharge patterns. Considerable unit-to-unit variability in the properties of A1 responses to vocalization stimuli has been noted by other authors (Wallace et al., 2005) and is as such not unexpected. Figure 3 gives illustrative examples from two different units. The spike patterns shown in Figure 3*A* appear to be highly "informative," in that the unit responds to the twitter calls with a fairly reproducible series of short bursts. The latency of these bursts and the number of spikes in each burst varies somewhat from trial to trial and these burst patterns are subjected to some degree of "noise" from spontaneous activity, but one nevertheless notices that the burst patterns evoked by the natural call exhibit systematic differences from those evoked by the time-reversed call. (Note also that the response pattern elicited by the time-reversed call does not appear to be a time-reversed copy of the response to the natural stimulus.) The spike patterns exhibited by the unit illustrated in Figure 3*B*, in comparison, appear much less informative with respect to stimulus identity than those seen in Figure 3*A*, in that the spike patterns are more variable from trial to trial and the patterns evoked by the normal and the reverse stimulus are less distinctive.

To be able to quantify these apparent unit-to-unit differences in the amount of information that appears to be carried in the temporal response patterns, we devised a simple pattern recognition algorithm that attempts to "guess" which stimulus evoked a particular response pattern. The first step in this algorithm is to represent each individual response as a poststimulus time histogram (PSTH). The choice of bin size used to construct these PSTHs is of some importance, which is why a range of different bin sizes were considered, as will be discussed further below. Figure 4, *A* and *B*, shows the data for the units whose response rasters were shown in Figure 3, when the responses to each of the 20 presentations of each stimulus are binned in 20-ms-wide bins. Once individual responses are represented in this way, each 1200-ms-long spike train becomes a list of 60 spike count values and can be thought of as a point, or vector, in a 60-dimensional space. When represented in this manner, the set of responses to a particular stimulus effectively form a "cloud" in an abstract, high-dimensional response space, and one can quantify how similar two responses are by calculating the Euclidian distance between the two responses in this space. If we now draw one of the response patterns from the dataset (a "test pattern") and we calculate the Euclidean distance of this test pattern to each of the "cloud centers" (mean response patterns) formed by the remaining responses (the "training sets"), then we can try to guess which stimulus might have evoked this test response pattern simply by assigning the test pattern to the stimulus associated with the closest training set. If the response patterns are reproducibly similar for repeated presentations of the same stimulus and reproducibly different from patterns evoked by other stimuli, then the response patterns will form distinct clusters in the response space and most test patterns will be correctly assigned, but, if the responses lack reproducible and distinctive patterns, then the assignments will be essentially random. By picking each response in turn as the test pattern and noting the proportion of correct assignments, we can therefore quantify how informative individual response patterns are with respect to stimulus class.

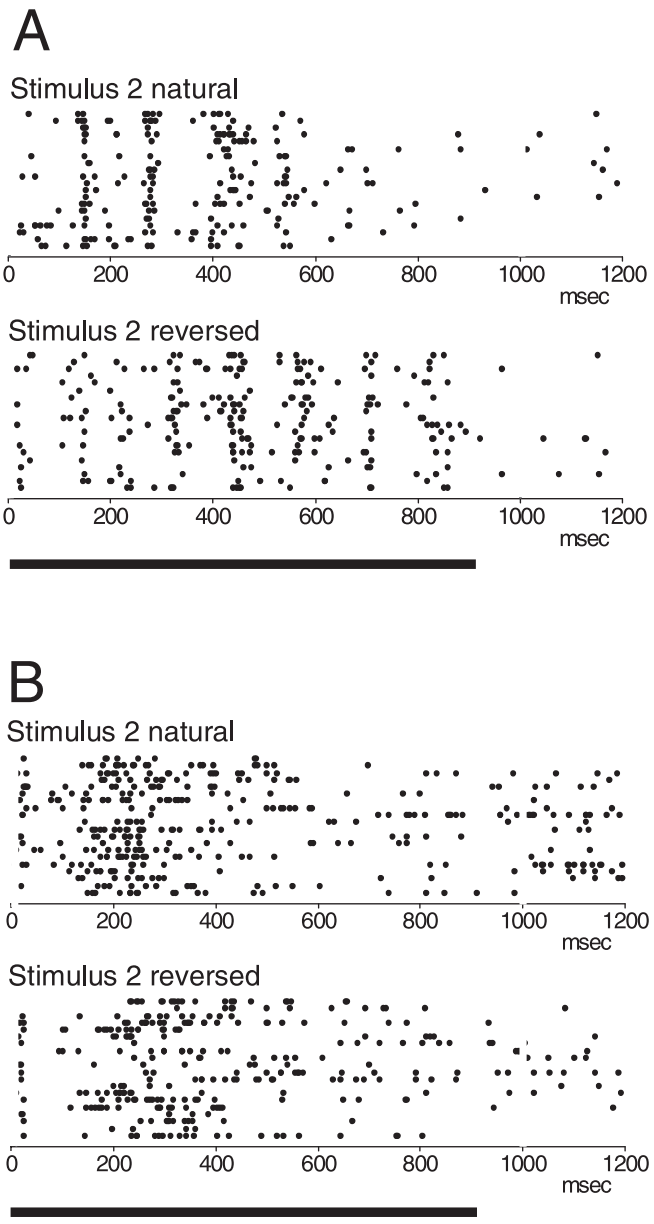


Figure 3. ***A***, Raster plot display of responses from one unit to the second twitter call and its time-reversed counterpart. Each dot indicates the timing of one action potential, and each row of dots gives the response to a single stimulus presentation. The thick black line underneath the raster plots indicates the duration of the stimulus. ***B***, Responses from a different unit to the same stimuli.

In practice, we found that performing a principle component analysis (PCA) on the response pattern of each unit before running this assignment algorithm improves the performance of this decoding algorithm somewhat in a number of cases. PCA transforms the coordinates of the response pattern vectors in a manner that exploits the correlation structure of the dataset and allows us to reduce the dimensionality of the response space by disregarding dimensions that capture only minimal amounts of the variance of the data. The small amount of variance in the dataset

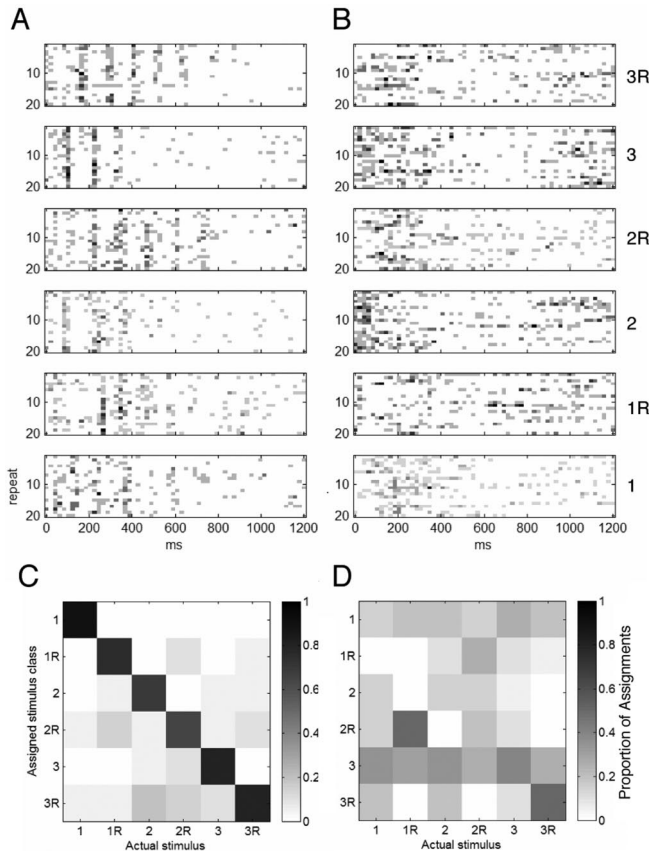


Figure 4. ***A***, ***B***, Responses of two different units. Each individual response is plotted as a grayscale histogram, i.e., the number of spikes in each 20 ms bin is given in a grayscale from white (0 spikes/bin) to black (5 spikes/bin). Individual responses to each of the 20 repeats of each stimulus are shown. Responses are grouped by stimulus class, and the corresponding stimulus is indicated by the label to the right (1, 2, 3 indicates first, second, and third twitter call stimulus from Fig. 1, whereas 1 R, 2R, 3R indicates the corresponding time-reversed counterpart). ***C***, ***D***, "Assignment" or "confusion" matrices illustrating the performance of our pattern classifier algorithm in decoding the spike patterns shown in ***A*** and ***B***, respectively. The grayscale indicates the proportion of the 20 responses to the stimulus class indicated on the ordinate that was attributed by the algorithm to the stimulus class indicated on the abscissa.

attributed to these dimensions is most likely attributable to noise in the response pattern, e.g., from spontaneous activity, whereas dimensions that account for much of the variance in the data can be thought of as "features" of the data. By running our assignment algorithm in a "principal component space" in which we retained a sufficient number of principal components to capture 90% of the variance of the data (between 16 and 40 components, depending on the unit), we were able to increase the performance of the classifier in some cases, presumably because it allowed us to avoid problems of "overfitting," but results obtained by running the classification on the raw spike count vectors are essentially similar.

Figure 4, *C* and *D*, shows "confusion matrices" that illustrate how well the classifier algorithm was able to "decode" individual spike trains. The grayscale indicates the proportion of test spike patterns from the set of responses to the stimulus given on the *x*-axis were "assigned" (i.e., judged to be "closest," or "most similar to" the set of responses evoked by) the stimulus indicated on the *y*-axis. If the algorithm was to identify all spike trains correctly, then the confusion matrix would feature a black main diagonal on a white background. However, if the algorithm fails to find useful information in the spike patterns, then the assignments are essentially random. Qualitatively, it certainly appears that the responses of the unit shown in Figure 4*A* are often correctly identified by the classifier, whereas those of the unit shown in Figure 4*B* are misclassified so frequently that it is not immediately clear whether the decoding process performed significantly above chance.

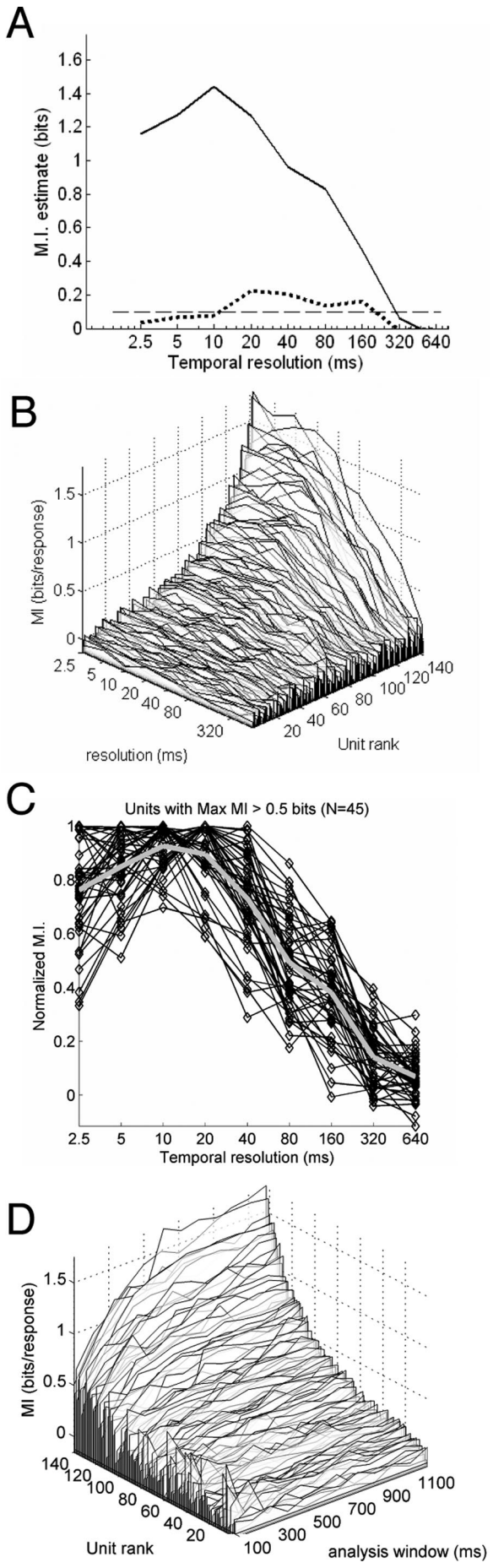
Confusion matrices like those shown in Figure 4, *C* and *D*, allow us to estimate the "information content" of the response patterns, or, more accurately, the mutual information (MI) between response and stimulus class. The MI (in bits) is given by Shannon's formula:

$$\mathrm{MI} = \sum_{x,y} p(x,y) \cdot \log_2\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right)$$

where x and y are the values taken by the random variables "presented stimulus class" and "assigned stimulus class" ($x, y \in \{1, 1R, 2, 2R, 3, 3R\}$), and one adopts the convention that $0 \cdot \log(0)$ evaluates to 0. The a priori probability $p(x)$ of any one stimulus having evoked any one particular response is ⅙ because we used six different stimuli in the experiment and each stimulus was presented with the same frequency. The probability of a response being assigned to any one stimulus class $p(y)$ and the joint probability of observing a particular combination of stimulus and response assignment $p(x, y)$ are not known a priori but can be estimated from the observed frequency distributions in the confusion matrix. However, it is important to remember that MI estimates calculated in this manner can be subject to non-negligible positive sampling biases, because using the observed frequency distributions as a necessarily rough estimator for the true underlying probabilities can easily lead to somewhat inflated MI estimates (Rolls and Treves, 1998; Trappenberg, 2002; Nelken et al., 2005). Here we estimated the expected size of this bias by calculating MI values for "shuffled" data, in which the response patterns had been randomly reassigned to stimulus classes. The shuffling was repeated 10 times, and the mean MI estimate for the 10 shuffled datasets was used as estimator for the bias. All MI values reported below were "bias corrected," i.e., the bias estimate obtained for each unit was subtracted from the original MI estimate. Bias estimates varied little from unit to unit, the median bias was 0.18 bits/response, and <5% of bias estimates exceeded 0.28 bits/response, so that bias corrected MI values >0.1 bits/response might, as a rough approximation, be deemed "statistically significant."

When analyzed in this manner, we estimate that the responses of the unit shown in Figure 4*A* transmit on average 1.28 bits of information about which of the twitter stimuli were transmitted, whereas the MI estimate for the unit shown in Figure 4*B* is only 0.21 bits (a much more modest value but one that exceeds our rough significance criterion of 0.1 bits, and, indeed, if we treat the confusion matrix shown in Fig. 5*D* as a χ^2 contingency table, then we can reject the null hypothesis that the actual and assigned stimuli are statistically independent at $p < 0.001$, confirming that the firing patterns of this unit does transmit very modest but statistically significant amounts of information).

As noted above, it is to be expected that the performance of our pattern classifier algorithm, and hence the MI estimates obtained, will depend on the bin width chosen at the first step, when individual response patterns are expressed as PSTHs. One could attempt to devise a more sophisticated algorithm that might be able to estimate the information content of a response pattern without temporal binning (Victor, 2002). However, we felt that it



would be of considerable interest to explore the dependency of the MI estimate on the initial bin size, or "temporal resolution," of the decoding step by systematically varying the bin size. The result of reanalyzing the data from the two units from Figure 4 at various bin sizes from 2.5 to 640 ms is shown in Figure 5*A*. One observes that, for the responses of the first unit, the pattern classifier performs well, extracting in excess of 1 bit of information per response, as long as it operates at a temporal resolution of 40 ms or finer. For coarser temporal resolutions, the amount of information that can be extracted from the temporal response patterns drops dramatically. For the responses from the unit shown in Figure 4*B*, the amount of information that can be extracted remains low and barely exceeds the critical value for statistical significance of 0.1 bits/response regardless of the temporal resolution used in the decoding step.

Figures 3–5*A* may give the impression that there may be two very different types of cells in auditory cortex, one transmitting large amounts of information about the identity of complex, natural stimuli through fairly precise temporal discharge patterns and the other conveying very little information about stimulus identity regardless of the temporal resolution at which its responses are analyzed. An obvious question is whether these units are representative examples of two more or less discrete classes of cells or rather two samples along a continuous spectrum of neural response properties. We address this question in Figure 5*B*, which shows a "waterfall" plot of estimated MI (on the *x*-axis) as a function of temporal resolution (on the *z*-axis) for all 142 A1 units recorded in the naive animals. Units are ordered by their maximal MI value along the *y*-axis of that plot. From this plot, it becomes apparent that units appear to form a continuum, with many cells transmitting practically no information about stimulus identity (although they responded vigorously to the stimuli), whereas others transmitted various amounts of information, up to ~1.7 bits/response.

Figure 5*B* suggests that the units in our sample form a continuum rather than clearly distinct classes, but, to facilitate additional discussion, we shall nevertheless draw a somewhat arbitrary distinction between "highly informative" and "poorly informative" units simply on the basis whether they transmit >0.5 bits/response. For the 47 highly informative units in our dataset, it is clear that the amount of information extracted from the responses tends to decline when the temporal resolution of the decoding becomes much coarser than 40 ms. Figure 5*C* serves to visualize the dependency of MI on temporal resolution for these units more clearly. It shows the normalized MI estimates (relative to the units own maximal MI value) plotted against temporal resolution. Data from all 47 highly informative units are shown superimposed. The mean normalized MI values across these 47 units is also shown as a light gray line. Although there is some unit-to-unit variability, one nevertheless observes a clear trend for MI values to reach a maximum at resolutions between

←

Figure 5. ***A***, MI between stimulus and response as estimated from the performance of the classifier algorithm at different temporal resolutions for the two units shown in Figure 3, *A* (solid line) and *B* (dotted line). MI values above the hatched horizontal line at $y = 0.1$ are deemed statistically significant at $\alpha = 0.05$. ***B***, Waterfall plot of MI as a function of temporal resolution for all 142 units from the untrained ferrets in this study. The units are ranked by maximum MI. ***C***, Normalized MI as a function of temporal resolution for units with maximal MI values >0.5 bits/response are plotted in black. The light gray line shows the mean of these normalized MI functions. ***D***, Waterfall plot of MI at a temporal resolution of 10 ms as a function of the length of the response period analyzed. The data are from the same units, ranked in the same order, as in ***B***.

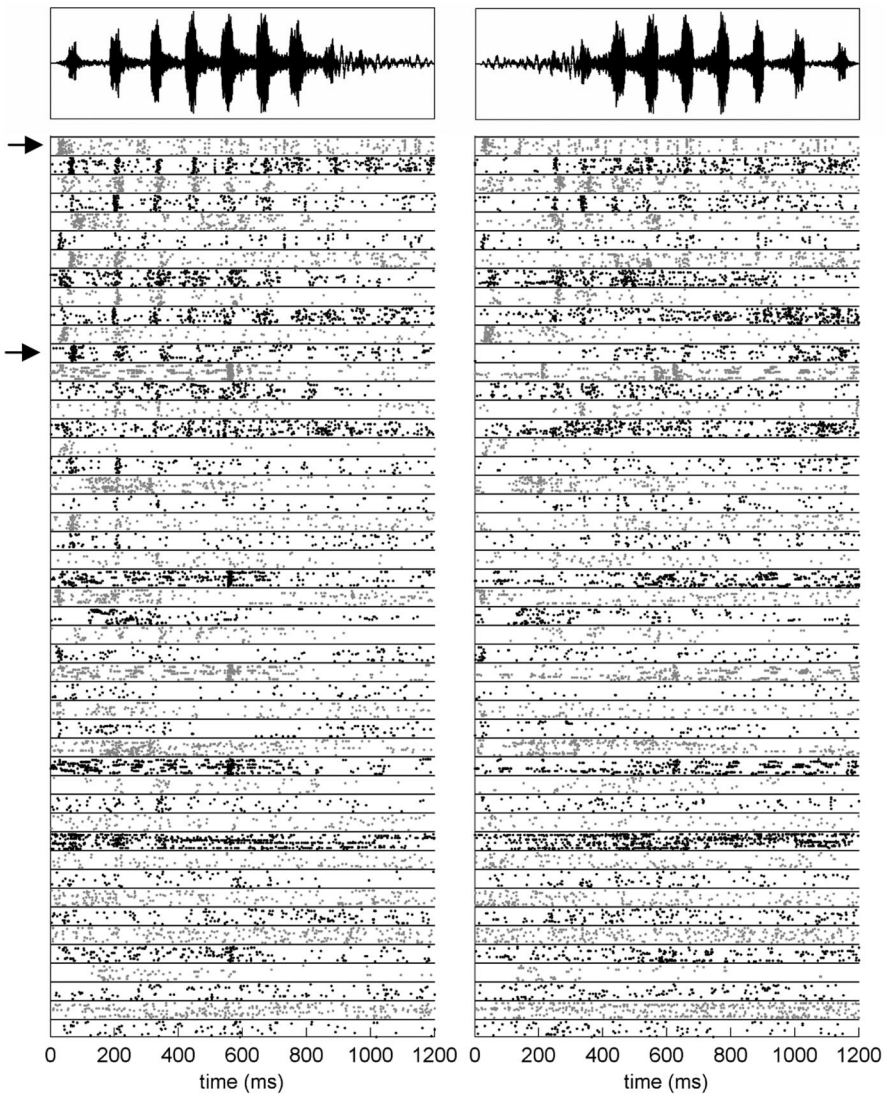


Figure 6. Diversity of response patterns. Responses to twitter 1 and twitter 1 reversed for 47 different A1 units shown in raster plot format. Responses are shown for every third unit in our dataset, arranged in ascending order of maximum MI, starting with the responses of the least informative unit at the bottom. Alternating dark and light gray dots and dividing lines are used to visually offset responses from different units. Above the raster plots, the temporal waveforms of the corresponding acoustic stimuli are shown. The arrows on the left mark the responses of the units ranked 142 of 142 (top) and 109 of 142 (12 rows lower down) in maximal MI.

10 and 20 ms. For most units, MI values start to decline sharply at resolutions >40 ms. MI estimates sometimes also decline somewhat if the temporal resolution used by the decoder becomes too fine, i.e., <5 ms.

An additional question one might ask is whether the entire response pattern is necessary to distinguish the various stimuli in this set. To investigate this question, we repeated the analysis of the response patterns at a 10 ms time resolution but fed only the first 100, 200, 300, . . . ms of the response patterns into the classifier algorithm. The results of this analysis with restricted time windows is shown in Figure 5*D*. It shows that the amount of information extracted by the classifier from the spike patterns increases monotonically with the length of the spike trains available for analysis, as one might expect, but, for most units, the extracted MI appears to asymptote so that there are much more rapid gains in MI with an increased analysis period during the first 600 ms compared with the second 600 ms of the response.

Comparing Figures 3, 4, and 5*A*, one might get the impression that perhaps those units that exhibit responses that are clearly strongly time locked to individual syllables or other transients in the stimulus envelope should be highly informative, whereas those that do not carry little information about stimulus identity. However, that is probably an oversimplification. In Figure 6, we show dot raster displays for the responses of every third unit in this sample to the first twitter call played forward and reversed. The dot rasters are arranged in order of increasing MI (i.e., the same rank order used in Fig. 5*B*) from the bottom up, and the waveforms of the stimuli are reproduced at the top of the figure to facilitate a visual comparison between stimulus envelope and timing of response bursts. Although there is a general trend for units with a clear tendency to time lock to individual syllables to occur closer to the top of the figure, that alone would not explain the ranking. Compare, for example, the responses plotted at the very top with the 12th unit from the top in this plot. These responses are marked by little arrows to the left of the raster plots. The rank order of these units in Figure 5*B* is 142 and 109, respectively, but, whereas the second appears to give more vigorously time-locked responses, the first one carries almost twice as much information in each response, presumably because its discharge pattern, although not easily attributable to particular peaks in the envelope of the stimulus, are nevertheless quite reproducible for representations of the same stimulus but distinct for different stimuli.

In summary, the responses recorded in naive ferrets are similar to those reported by Wang and Kadia (2001) for the cat and very different from those reported for the marmoset, in that overall mean spike counts do not discriminate natural from time-reversed twitter calls. Nevertheless, our analysis clearly demonstrates that many ferret A1 neurons carry significant amounts of information about the marmoset twitter stimuli in their temporal pattern, information that could be used to distinguish natural from time-reversed calls. Furthermore, although the decline in MI at resolutions <5 ms seen for some units in Figure 5*C* is clearly a limitation of our decoder (a more sophisticated decoder might be able to combine bins appropriately to recover the information available at coarser resolution), we believe that the decline at resolutions coarser than 40 ms reflects physiological properties of the representation of this particular stimulus set at the level of A1. Of course, it would be of considerable interest to know whether this representation changes and whether it perhaps becomes more "marmoset-like," if ferrets are familiarized and trained to recognize marmoset twitter stimuli. To explore this question, two ferrets were trained to associate marmoset twitters with water rewards in a Go/No-go paradigm.

Behavioral training

Using the paradigm described in Materials and Methods, two ferrets were trained to respond to marmoset twitter stimuli by pushing a reward spout to receive a water reward (a Go response). Sound stimuli other than marmoset twitters were to be ignored, and the next trial was to be initiated by pushing a start spout (a No-go response). Figure 7*A* shows a typical example of the performance of the animals during a behavioral testing session after ~2 months of training. The animal has clearly learned an association between marmoset twitter stimuli and water reward, making Go responses to almost every presentation of the twitter stimuli but only very rarely making Go responses to sounds other than the marmoset twitters. Figure 7*B* shows the "reaction times," i.e., times from the onset of the sound stimuli to the triggering of the reward spout (Go response) or the start spout (No-go response), respectively, for the same training session. Bearing in mind that the sound stimuli are typically just >1 s in duration (compare with Fig. 1) and that the behavioral training program was set up not to register any responses made <0.9 s after stimulus onset to encourage the animals to listen carefully before making their choice, it is apparent that the trained animals typically make their choices very rapidly after the end of the stimuli. Figure 7*C* shows the "learning curves" for the two trained animals. The percentage of correct responses in each training session is plotted against time (in days) elapsed since the first training session. In an initial training period, lasting 9 d for the first animal and 16 d for the second, during which the animal familiarized itself with the setup and the mechanics of the task, only Go stimuli were used, and, because this early period of procedural training involved no acoustic discrimination, no performance data are shown for this early stage of training. Although there are slight differences in the shapes of the learning curves for these two animals, both animals typically perform at 80% correct or better after 25 d of training and routinely exceed 90% correct after ~2 months of training. As mentioned in Materials and Methods, intense training periods during which an animal would frequently run two sessions per day were interspersed with "training holidays," which could last anywhere from 2 d to >3 months. The timing of these training holidays was not governed by scientific considerations but was dictated partly by animal welfare considerations (prolonged uninterrupted training periods with their accompanying restrictions of access to drinking water could otherwise lead to weight loss and adverse health effects) and partly by other constraints on the experimenters' time. Figure 7*C* illustrates that the performance in the first session immediately after a long training holiday can be noticeably reduced, particularly during early stages of the training, but the animals typically return to >90% performance after just a single "refresher" session.

Electrophysiological responses in the trained animals

After the training periods charted in Figure 7*C*, the animals were prepared for physiological recording in a manner identical to that used for the naive animals. In total, we obtained responses to the marmoset twitter calls from 501 units in the two trained animals.

The data from these animals were analyzed in a manner identical to that used for the naive animals. Figure 8 plots the response rate evoked by natural twitter calls against rates evoked by their time-reversed counterparts for the trained animals. The layout of the figure is identical to that of Figure 2. As in the untrained ferrets, the data cluster tightly around the main diagonal ($R_{Nat} \approx R_{Rev}$), and, for all three twitter stimulus exemplars, the mean values of the selectivity index d were close to 0 (means ± SD of −0.030 ± 0.106, −0.0036 ± 0.0978, and −0.028 ± 0.1069, respectively). The overall median selectivity index was, in fact, slightly but statistically significantly negative (−0.021; $p < 10^{-5}$, sign test), indicating that the responses to the natural twitter calls, which had been part of the training set, were if anything ever so

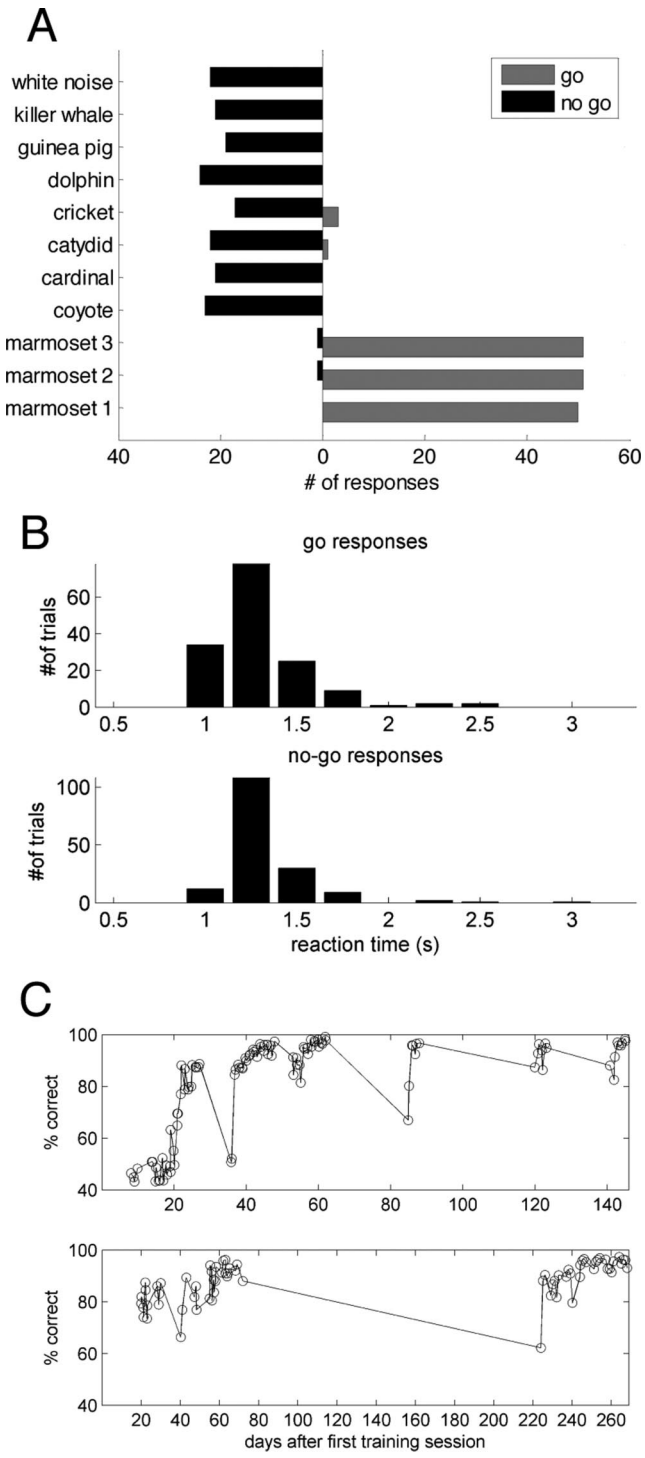


Figure 7. ***A***, Histogram showing a typical example of the behavioral performance after 2 months of training in a Go/No-go paradigm in which marmoset twitter calls served as a Go stimulus. The length of the black and gray horizontal bars gives the number of No-go and Go responses, respectively, to each of the stimuli indicated along the *y*-axis. In this run, the animal made only two inappropriate No-go responses (1 each to twitters 2 and 3) and only four inappropriate Go responses (1 to the katydid and 3 to the cricket call). ***B***, Reaction times, relative to stimulus onset, for Go and No-go trials, respectively, from the training session shown in ***A***. ***C***, Learning curves plotting performance (percentage of correct responses) against days from the start of training.

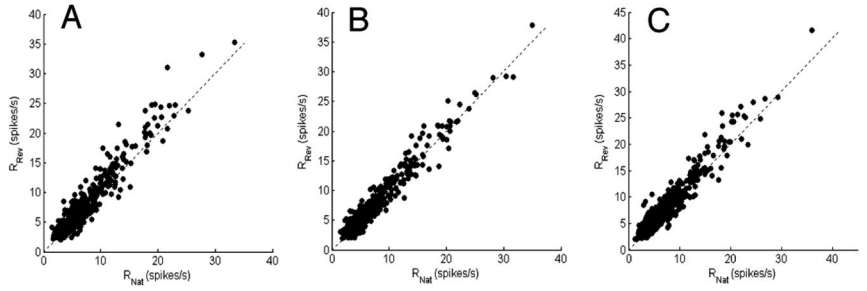


Figure 8. Mean response rate evoked by natural (R_{Nat}) plotted against response to time-reversed (R_{Rev}) twitter call stimulus in ferrets trained to recognize marmoset twitter calls. Responses for first, second, and third exemplar of the twitter stimuli are shown in ***A***, ***B***, and ***C***, respectively.

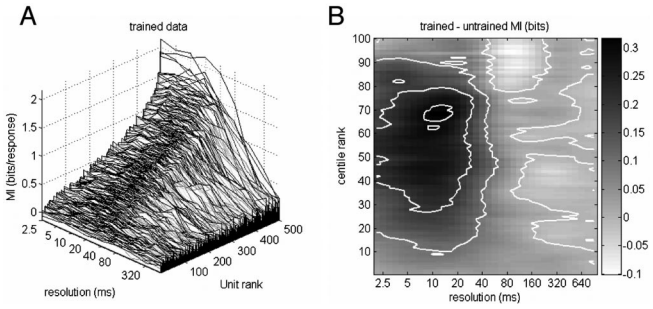


Figure 9. ***A***, Waterfall plot of MI as a function of temporal resolution for all 501 units from the trained ferrets in this study. ***B***, Grayscale map showing the difference between trained (***A***) and untrained (Fig. 5*B*) animals. White contour lines are drawn at $z = 0$, 0.1, 0.2, and 0.3 bits.

slightly weaker on average than those to the time-reversed sounds that the animals had never heard before the electrophysiological recording. However, this effect is so small that, statistical significance notwithstanding, one would be hard pushed to accord it much "physiological significance." In any case, the fact that the trained animals have learned to recognize and associate a reward with the natural marmoset twitter vocalizations has clearly not induced the sort of preference or selectivity for these natural marmoset twitters that has been reported for neurons in marmoset A1 (Wang et al., 1995).

In Figure 9*A*, we show the estimates of the mutual information between response pattern and stimulus for the trained data obtained with the decoding algorithm described above and plotted in a waterfall format identical to that used in Figure 5*A*. Like in the untrained data, we notice that the amount of information extracted by the pattern classification algorithm varies from unit to unit and depends on the temporal resolution, with the highest MI values typically seen at resolutions finer than 40 ms. When comparing Figures 9*A* and 5*A*, one also notices that the maximum MI values are in general higher in the trained than in the untrained animals. In Figure 5*A*, only approximately one-third of units reached MI values exceeding 0.5 bits/response. In the trained animals, this proportion has grown to approximately two-thirds, and the most informative cells reach MI values in excess of 2 bits/response, close to the theoretical maximum given by the entropy of the stimulus set of $\log_2(6) = 2.58$ bits/response. Figure 9*B* shows the differences between the MI distributions in the trained and naive animals. It effectively plots the difference between the waterfall plots shown in Figures 9*A* and 5*B*, respectively, but, given that the number of units in the trained and untrained samples was not the same, the "MI surfaces" shown in the waterfall plots first had to be resampled along the unit rank dimension (*x*-axis) using the resample function of the Matlab signal processing toolbox to give interpolated MI surfaces with 100 "unit centile rank" steps for each dataset. The difference between the resampled surfaces is shown in Figure 9*B* as a grayscale map. It shows that, although the amount of information recoverable from the spike trains at coarse temporal resolutions >80 ms hardly changed as a consequence of training, for shorter temporal resolutions, one sees fairly substantial increases in information throughout a large proportion of the sample. A Wilcoxon's rank-sum test comparing the maximum MI values across units for the trained and untrained animals was used to confirm that these marked increases in transmitted information are statistically highly significant ($p < 10^{-8}$).

Discussion

Together, the results of our study emphasize the importance of temporal pattern codes operating at a resolution in the order of tens of milliseconds in the representation of vocalization stimuli in primary auditory cortex and argue against an important role of call-specific neurons. Training an animal to recognize previously unfamiliar vocalizations does not lead to the formation of call-selective neurons in the trained animal's A1, but the amount of information carried in temporal pattern codes maintained by nonselective neurons increases markedly as a consequence of training.

These results are therefore in general agreement with a recent study by Gehr et al. (2000), who found that a rate-based representation of vocalization stimuli in cat A1 would be highly inefficient.

Our simple decoding algorithm extracted the highest amounts of information from the neural spike patterns when it operated at temporal resolutions between 10 and 20 milliseconds, and, for resolutions of 80 ms or greater, performance decreased markedly. If we compare this result with findings from an *in vivo* intracellular recording study of visual cortical neurons by Azouz and Gray (1999), who found that the energy of membrane voltage fluctuations in the gamma frequency band (20–70 Hz, i.e., at periods of 14–50 ms) were highly predictive of spike rates, then we are led to suspect that the relationship we observed between information extracted and the temporal resolution of the decoder may reflect fundamental physiological properties of neocortex.

It is interesting to compare these timescales with the timescale of other known auditory perceptual or physiological phenomena. For example, accurate temporal order judgments on the click trains with varying amplitude require interclick intervals of >20 ms (Hirsh, 1959). Neurons in the primary auditory cortex respond to brief stimuli presented repeatedly in rapid succession only if repetition rates remain below ~20–35 Hz (Schreiner et al., 1997; Lu et al., 2001a; Eggermont, 2002). Furthermore, when samples of recorded speech are cut into segments that are locally time reversed, then this speech remains highly intelligible provided the fragments are no more than 50 ms long (Saberi and Perrott, 1999). These observations suggest that some aspects of auditory perception may, in a sense, occur at a "frame rate" of 20–50 Hz, and our analysis suggests that this would also be an appropriate frame rate for reading out the temporal pattern representations of complex sounds at the level of the primary auditory cortex, although additional studies with more extensive sets

of complex stimuli will be required to examine whether this finding generalizes.

Previous studies have shown that temporal discharge properties in auditory cortex can be subject to plastic changes. For example, Kilgard and colleagues (Kilgard and Merzenich, 1998; Kilgard et al., 2001) have shown that pairing rapid pure-tone sequences with electrical stimulation of the nucleus basalis can induce increased tone following rates in A1, and Bao et al. (2004) reported enhanced responses to rapid click trains in the A1 of rats that had been trained to use such click trains as auditory cues, which helped them localize a food source. Our study extends these findings and demonstrates explicitly that training-induced plastic changes in the temporal response properties of A1 neurons lead to an increase in task-relevant acoustic information carried in the neural discharge patterns.

However, teaching an animal a few new "items of vocabulary" does not seem to induce call selectivity at the level of A1. Wang and colleagues (Wang et al., 1995; Wang and Kadia, 2000) described what appeared to be call selectivity in a subset of marmoset A1 neurons and hypothesized that this subpopulation might be specialized for the quick and accurate detection of frequently heard vocalizations. However, examples of firing patterns recorded from marmoset A1 neurons reproduced by Wang et al. (1995) suggests that at least some of these neurons do also exhibit intricate, reproducible discharge patterns, in addition to modulating their overall firing rate. These firing patterns are likely to be highly information bearing, and the relative importance of putative temporal pattern codes versus rate coding through putative call selectivity in marmoset A1 remains uncertain. In fact, in marmoset A1, apparent call selectivity in overall response rates and temporal pattern coding may both operate within the same population of neurons.

By referring to the response asymmetry observed in marmoset cortex as call selectivity, Wang and Kadia (2001) encourage their readers to interpret the "preference" of normal to time-reversed vocalizations within A1 as a necessary or at least useful step in the processing of vocalizations, but our result that ferrets successfully "process" the same sounds in a behavioral task without exhibiting any obvious response asymmetry or selectivity in A1 does suggest that we have to keep our minds open to alternative interpretations.

Thus, the preference for natural over time-reversed stimuli in marmoset A1 might simply reflect asymmetries in the response to temporal features of sounds in general (Lu et al., 2001b) rather than a specificity for vocalization stimuli in particular. At the level of A1, "genuine" call selectivity might be rare or nonexistent. How such asymmetries in the response properties of marmoset A1 neurons arise also remains unclear. It is conceivable that they could be the result of auditory experience in infancy (Nakahara et al., 2004) or they might be innate.

Studies on the encoding of conspecific vocalizations in primates (Rauschecker and Tian, 2000) suggest that call specificity that is manifest in coarse overall changes in firing rate may be more common in higher-order cortical fields (lateral belt areas), and it would be of considerable interest to know to what extent this apparent call specificity is innate or shaped by experience. Ultimately, good performance at a Go/No-go task like the one used here requires the animal to make a categorical decision as to whether a complex sound belongs to a particular class of vocalizations or not, and this categorical decision is likely to manifest itself in distributed rate representations involving prefrontal and even premotor cortical areas (Romo and Salinas, 2003). Interestingly, Romanski et al. (2005) recently described neurons in primate prefrontal cortical areas that appear to exhibit a high degree of call selectivity when tested with vocalization stimuli, but the role of these neurons in auditory processing or recognition tasks and the transformation of acoustic information from A1 to these high-order areas is only beginning to be explored.

The trained animals in our study were only exposed to the marmoset vocalizations in adulthood, and one might ask whether we might have observed the emergence of call selectivity if exposure to these calls had commenced in infancy. Some authors argue that exposure during a "critical period" early in life may exercise a particularly powerful influence on neural response properties (Nakahara et al., 2004), and, in humans, auditory experience in infancy can greatly affect which acoustic or phonetic distinctions an individual will perform with ease later in life and which ones are difficult (Kuhl et al., 1992). However, within the parameters laid down during the critical period, mature mammals are clearly able to learn to recognize new complex sounds and vocalizations, and there is at present no reason to assume that adult A1 would use a fundamentally different coding scheme for the encoding of vocalizations learned in adulthood from those learned in infancy. Furthermore, one might note that Gehr et al. (2000) report that, in cat A1, there appears to be no strong preference for natural over time-reversed meows, although one would expect kittens to be exposed to conspecific calls throughout any putative critical period. Instead, these authors consider it "more likely that the temporal structure in the firing patterns, reflecting that of major peaks in the vocalization envelope, is at the basis of a cortical representation of cat meows" in cat A1 (Gehr et al., 2000).

In the visual system, one can observe neurons with responses that are highly selective for certain classes of natural stimuli [e.g., faces or hands (for review, see Rolls and Treves, 1998; Tsao et al., 2006)]. This type of specificity for natural object classes is usually characterized by a degree of "response invariance" as well as selectivity, i.e., face cells respond to faces regardless of the position within very large receptive fields, whether shown in frontal view or in profile, large or small, but they respond very poorly or not at all to objects that are not faces. This type of selectivity for a class of natural stimuli has hitherto only been described in higher-order visual cortical areas. Interestingly, a recent report suggests that "auditory object-specific" responses can be induced in nonprimary forebrain areas of songbirds by operant conditioning (Gentner and Margoliash, 2003). However, auditory object specificity was clearly not induced in ferret A1 in our operant conditioning study. Thus, it appears that the role of A1 in the processing of vocalization stimuli is predominantly that of representing acoustic features of stimuli through temporal pattern codes in a nonselective manner, whereas a more auditory object-selective representation of the acoustic environment may well emerge in higher-order areas of auditory cortex.

References

Azouz R, Gray CM (1999) Cellular mechanisms contributing to response variability of cortical neurons *in vivo*. J Neurosci 19:2209–2223.

Bao S, Chang EF, Woods J, Merzenich MM (2004) Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. Nat Neurosci 7:974–981.

Eggermont JJ (2002) Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. J Neurophysiol 87:305–321.

Gehr DD, Komiya H, Eggermont JJ (2000) Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. Hear Res 150:27–42.

Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. Nature 424:669–674.

Heffner HE, Heffner RS (1986) Effect of unilateral and bilateral auditory

cortex lesions on the discrimination of vocalizations by Japanese macaques. J Neurophysiol 56:683–701.

Hirsh IJ (1959) Auditory perception of temporal order. J Acoust Soc Am 31:759–767.

Kilgard MP, Merzenich MM (1998) Plasticity of temporal information processing in the primary auditory cortex. Nat Neurosci 1:727–731.

Kilgard MP, Pandya PK, Vazquez J, Gehi A, Schreiner CE, Merzenich MM (2001) Sensory input directs spatial and temporal plasticity in primary auditory cortex. J Neurophysiol 86:326–338.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255:606–608.

Lu T, Liang L, Wang X (2001a) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131–1138.

Lu T, Liang L, Wang X (2001b) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. J Neurophysiol 85:2364–2380.

Nakahara H, Zhang LI, Merzenich MM (2004) Specialization of primary auditory cortex processing by sound exposure in the "critical period." Proc Natl Acad Sci USA 101:7170–7174.

Nelken I, Rotman Y, Bar Yosef O (1999) Responses of auditory-cortex neurons to structural features of natural sounds. Nature 397:154–157.

Nelken I, Chechik G, Mrsic-Flogel TD, King AJ, Schnupp JW (2005) Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. J Comput Neurosci 19:199–221.

Pelleg-Toiba R, Wollberg Z (1991) Discrimination of communication calls in the squirrel monkey: "call detectors" or "cell ensembles"? J Basic Clin Physiol Pharmacol 2:257–272.

Rauschecker JP (1998) Cortical processing of complex sounds. Curr Opin Neurobiol 8:516–521.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci USA 97:11800–11806.

Rolls ET, Treves A (1998) Neural networks and brain function. Oxford: Oxford UP.

Romanski LM, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734–747.

Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4:203–218.

Saberi K, Perrott DR (1999) Cognitive restoration of reversed speech. Nature 398:760.

Schreiner CE, Mendelson J, Raggio MW, Brosch M, Krueger K (1997) Temporal processing in cat primary auditory cortex. Acta Otolaryngol Suppl 532:54–60.

Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. J Neurophysiol 86:1445–1458.

Trappenberg TP (2002) Fundamentals of computational neuroscience. Oxford: Oxford UP.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674.

Victor JD (2002) Binless strategies for estimation of information from neural data. Phys Rev E Stat Nonlin Soft Matter Phys 66:051903.

Wallace MN, Shackleton TM, Anderson LA, Palmer AR (2005) Representation of the purr call in the guinea pig primary auditory cortex. Hear Res 204:115–126.

Wang X (2000) On cortical coding of vocal communication sounds in primates. Proc Natl Acad Sci USA 97:11843–11849.

Wang X, Kadia SC (2001) Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J Neurophysiol 86:2616–2620.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685–2706.

Winter P, Funkenstein HH (1973) The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. (*Saimiri sciureus*). Exp Brain Res 18:489–504.