

# Neural Ensemble Codes for Stimulus Periodicity in Auditory Cortex

Jennifer K. Bizley,<sup>1</sup> Kerry M. M. Walker,<sup>1</sup> Andrew J. King,<sup>1</sup> and Jan W. H. Schnupp<sup>1,2</sup>

<sup>1</sup>Department of Physiology, Anatomy, and Genetics, University of Oxford, Oxford OX1 3PT, United Kingdom, and <sup>2</sup>Robotics, Brain, and Cognitive Sciences Department, Italian Institute of Technology, 16163 Genova, Italy

We measured the responses of neurons in auditory cortex of male and female ferrets to artificial vowels of varying fundamental frequency ( $f_0$ ), or periodicity, and compared these with the performance of animals trained to discriminate the periodicity of these sounds. Sensitivity to  $f_0$  was found in all five auditory cortical fields examined, with most of those neurons exhibiting either low-pass or high-pass response functions. Only rarely was the stimulus dependence of individual neuron discharges sufficient to account for the discrimination performance of the ferrets. In contrast, when analyzed with a simple classifier, responses of small ensembles, comprising 3–61 simultaneously recorded neurons, often discriminated periodicity changes as well as the animals did. We examined four potential strategies for decoding ensemble responses: spike counts, relative first-spike latencies, a binary “spike or no-spike” code, and a spike-order code. All four codes represented stimulus periodicity effectively, and, surprisingly, the spike count and relative latency codes enabled an equally rapid readout, within 75 ms of stimulus onset. Thus, relative latency codes do not necessarily facilitate faster discrimination judgments. A joint spike count plus relative latency code was more informative than either code alone, indicating that the information captured by each measure was not wholly redundant. The responses of neural ensembles, but not of single neurons, reliably encoded  $f_0$  changes even when stimulus intensity was varied randomly over a 20 dB range. Because trained animals can discriminate stimulus periodicity across different sound levels, this implies that ensemble codes are better suited to account for behavioral performance.

## Introduction

Any sound will activate neurons throughout auditory cortex, but we know little about how the activity of neural populations is “read out.” In addition to spike rate, response latencies can encode sensory events (Jenison, 2000; Brugge et al., 2001; Nelken et al., 2005), but it is unclear what would serve as a temporal reference, given that the brain has no independent measure of stimulus onset. It has been suggested that single-unit spike latencies might be referenced to the population onset response (Chase and Young, 2007) or that the order in which neurons fire might be important (Gautrais and Thorpe, 1998; Van Rullen and Thorpe, 2001). Relative latency codes are particularly attractive because they could potentially facilitate rapid behavioral judgments (Van Rullen and Thorpe, 2002; Johansson and Birznieks, 2004).

There have so far been few successful attempts to record from sufficiently large ensembles of neurons to assess the potential of different population codes. One criterion against which to judge a neural code is to ask whether it provides sufficient stimulus-related information to account for behavioral performance. Neu-

rometric techniques have therefore been used to model sensory discrimination abilities from the firing behavior of single neurons (Parker and Newsome, 1998). Extending this approach to codes carried by activity patterns distributed across neuron ensembles may provide novel insights into the link between neural activity and perception (Nishikawa et al., 2008; Walker et al., 2008).

Natural sounds are often periodic, and this periodicity can evoke the percept of pitch. Therefore, examining how neurons encode the fundamental frequency ( $f_0$ ) of sounds may contribute to our understanding of the neural basis of pitch perception. The periodicity of the waveform of a sound is also reflected in the harmonicity of its spectrum, and both temporal and spectral mechanisms appear to contribute to pitch extraction (Cedolin and Delgutte, 2005; De Cheveigne, 2005; McDermott and Oxenham, 2008).

Sound periodicity is represented by the time-locked discharges of auditory nerve fibers (Javel, 1980; Winter et al., 1993; Cariani and Delgutte, 1996) and cochlear nucleus neurons (Sayles and Winter, 2008). In the midbrain, periodicities up to a few 100 Hz are represented as temporal patterns of spikes, and faster periodicities are represented with rate codes (Langner and Schreiner, 1988; Schreiner and Langner, 1988; Rees and Palmer, 1989). Sensitivity to  $f_0$  is found throughout ferret auditory cortex (Bizley et al., 2009), while a specialized region of marmoset auditory cortex responds to the missing fundamental of sounds (Bendor and Wang, 2005). However, none of the proposed coding mechanisms have been shown to account for pitch judgments independent of changes in other parameters, such as sound intensity or spectral timbre. Thus, it is unclear whether, or how, the responses of cortical neurons, which are

Received Nov. 4, 2009; revised Feb. 25, 2010; accepted March 4, 2010.

This work was supported by Biotechnology and Biological Sciences Research Council Grant BB/D009758/1 (J.W.H.S., A.J.K., J.K.B.), Engineering and Physical Sciences Research Council Grant EP/C010841/1 (J.W.H.S.), a Rothermere Fellowship and Hector Pilling Scholarship (K.M.M.W.), and a Wellcome Trust Principal Research Fellowship (A.J.K.). We are grateful to Johannes Dahmen, Fernando Nodal, and Andreas Schultz for assistance with data collection.

Correspondence should be addressed to Dr. J. K. Bizley, Department of Physiology, Anatomy, and Genetics, Sherrington Building, University of Oxford, Parks Road, Oxford OX1 3PT, UK. E-mail: jennifer.bizley@dpag.ox.ac.uk.  
DOI:10.1523/JNEUROSCI.5475-09.2010

Copyright © 2010 the authors 0270-6474/10/305078-14\$15.00/0

**Table 1. Numbers of units (comprising small multiunit clusters and single units) and neural ensembles recorded in each animal**

Animal	Units	Ensembles
F0534	178	8
F0614	84	11
F0615	95	12
F0626	78	10
F0627	61	6
F0634	77	6
F0709	171	5

Animal F0615 was awake and passively listening to sounds. All other recordings were performed in anesthetized animals.

typically broadly sensitive to multiple stimulus dimensions, could be decoded to explain behavioral performance.

We compared the responses of single auditory cortical neurons and ensembles of simultaneously recorded neurons to the behavioral performance of ferrets trained on an  $f_0$  discrimination task. We examined different putative decoding strategies and found that codes based on spike timing or count can discriminate stimulus periodicity approximately equally well.

## Materials and Methods

**Animal preparation.** All animal procedures were approved by the local ethical review committee and performed under license from the United Kingdom Home Office. Twelve pigmented ferrets were used in this study. Five of these animals participated in behavioral testing, and electrophysiological recordings were obtained from the seven others.

**Electrophysiological recordings.** Recordings were performed in one awake, passively listening animal and six anesthetized ferrets (Table 1). For the acute recordings, anesthesia was induced with medetomidine hydrochloride (Domitor;  $0.022 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ ) and ketamine (Ketaset;  $5 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ ; Fort Dodge Animal Health) and maintained with an intravenous infusion ( $5 \text{ ml/h}$ ) of this mixture in physiological saline containing 5% glucose. The ferret also received a single subcutaneous dose of  $0.06 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$  atropine sulfate (C-Vet Veterinary Products) and subcutaneous doses of  $0.5 \text{ mg/kg}$  dexamethasone (Dexadreson; Intervet UK Ltd.) every 12 h to reduce bronchial secretions and cerebral edema, respectively. The animal was intubated, placed on a ventilator (7025 respirator; Ugo Basile), and supplemented with oxygen. Body temperature, end-tidal  $\text{CO}_2$ , and the electrocardiogram were monitored throughout the experiment.

The animal was placed in a stereotaxic frame, and the temporal muscles on both sides were retracted. A metal bar was attached to the right side of the skull, holding the head without additional need of a stereotaxic frame. The left temporal muscle was mostly removed, and the auditory cortex was exposed by a craniotomy. The dura was removed, and the cortex was covered with silicon oil. The animal was then transferred to a small table in an anechoic chamber (IAC Ltd.).

Recordings were made with silicon probe electrodes (Neuronexus Technologies). In three animals, we used electrodes with eight active sites on four parallel probes, with a vertical spacing of  $150 \mu\text{m}$ . In a few recordings in one of these animals and for all recordings in an additional three animals, we used electrodes with 16 active sites spaced at  $100 \mu\text{m}$  intervals on each of two probes. The electrodes were positioned so that they entered the cortex approximately orthogonal to the surface. A photographic record was made of each electrode penetration to document their location relative to anatomical landmarks (surface blood vessels and sulcal patterns).

Extracellular recordings were also performed in one ferret while it was awake and passively listening to stimuli. A cranial mount was surgically implanted 1 month before the first recording session. During this procedure, surgical anesthesia was induced with an intramuscular injection of medetomidine and ketamine. The animal was intubated, and anesthesia was maintained with 1–2% isoflurane in oxygen-enriched air. It was placed in a stereotaxic frame, and the temporal muscles were retracted and partially removed. The auditory cortex was exposed by a craniotomy,

**Table 2. Numbers of units (single units) and neural ensembles recorded in each auditory cortical field**

Cortical field	A1	AAF	PPF	PSF	ADF
Units (single)	333 (180)	82 (20)	152 (111)	128 (78)	40 (19)
Ensembles	29	5	12	7	5

and a cranial mount of bone cement with a resealable metal well was attached to the skull above the craniotomy. The cranial mount also contained a metal fitting that allowed the head to be fixed to a solid recording frame. During the month after implant surgery, the animal was trained with positive food reinforcement to accept head restraint. Recordings were performed with the head restrained using up to 5 quartz/platinum–tungsten electrodes (Thomas Recording) lowered through the dura using a Mini Matrix System microdrive (Thomas Recording).

Five cortical fields were investigated: two tonotopic primary fields, the primary auditory cortex (A1) and anterior auditory field (AAF); two tonotopic secondary fields on the posterior ectosylvian gyrus, the posterior suprasylvian (PSF) and posterior pseudosylvian (PPF) fields; and one non-tonotopic secondary area on the anterior ectosylvian gyrus, the anterior dorsal field (ADF) (Bizley et al., 2005). The number of units recorded in each cortical field are listed in Table 2.

**Stimuli.** Artificial vowel sounds were generated in MATLAB (MathWorks), by using an algorithm adapted from Malcolm Slaney's Auditory Toolbox (<http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/>) to bandpass filter click trains. The click rate determined the value of  $f_0$  and therefore the evoked pitch. The bandpass filters determined the stimulus timbre and, in both behavioral and neurophysiological testing, were kept constant at values corresponding to the vowel /i/ (formant frequencies centered at 430, 2132, 3070, and 4100 Hz). The vowel sounds (150 ms in duration, 5 ms onset/offset ramps) were normalized to have equal root-mean-square amplitudes, and calibrations were performed to ensure that changes in periodicity did not influence the overall sound pressure level (SPL). At each recording site, a fixed range of  $f_0$  values was presented, from either 150 or 200 Hz to  $\sim 1900$  Hz.

Pure-tone stimuli were used to obtain frequency response areas, to both characterize individual units and determine tonotopic gradients. The latter were used to confirm the cortical field in which the recordings were made.

Stimuli were delivered using Tucker Davis Technologies System 3 digital signal processors systems and were presented to anesthetized animals through customized pairs of Panasonic RPHV297 earphones attached to plastic otoscope speculae inserted into the ear canals. The earphones were closed-field calibrated using an  $\frac{1}{8}$ -inch condenser microphone (Brüel and Kjær). In the awake ferrets, stimuli were presented in the free field in the same anechoic room via an Audax TWO26M0 speaker (Audax Industries) located  $\sim 80$  cm from the animal's head at  $30^\circ$  contralateral to the recording chamber. The speaker was calibrated using a  $\frac{1}{4}$ -inch condenser microphone (Brüel and Kjær) placed near the position of the ferret's head within the recording setup. The speaker and headphone calibrations were used to create inverse filters to ensure that a flat ( $\pm 5$  dB) output was produced from 100 to 24,000 Hz.

In the electrophysiological studies, vowel sounds were presented in isolation, at a rate of  $1 \text{ s}^{-1}$ . In the behavioral experiments, vowels were presented in pairs, with a fixed "reference"  $f_0$  followed by a 50 ms silent interval and then a "target"  $f_0$ . Presenting single vowels in the recording experiments allowed us to vary which  $f_0$  we chose as a reference sound in later analysis. However, for 152 neural units, we instead collected responses to stimuli presented as reference–target pairs, matching precisely the stimulus configuration of the psychoacoustic experiments (see below). The responses recorded with vowel pairs were very similar to those recorded with vowels presented in isolation, i.e., response functions calculated for the second vowel in the pair (the target vowel) also commonly had either high-pass or low-pass  $f_0$  tuning characteristics. Two example responses are shown in supplemental Figure 1 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

**Data acquisition and analysis.** The electrophysiological recordings were bandpass filtered (500–5000 Hz), amplified, and digitized at 25 kHz. Data acquisition and stimulus generation were performed using

BrainWare software (Tucker Davis Technologies). Spike sorting was performed offline. Single units were isolated from the digitized signal by either manually clustering data according to spike features, such as amplitude, width, and area, or using an automated *k*-means clustering algorithm in which the voltage potential at seven points across the duration of the spike window served as variables. We also inspected autocorrelation histograms, and only responses with a clear refractory period in their interspike interval histograms were classed as single neurons. In instances in which we could not observe a clear refractory period, we classed the responses as multi-neuron clusters (i.e., spikes were assumed to be produced by more than one neuron).

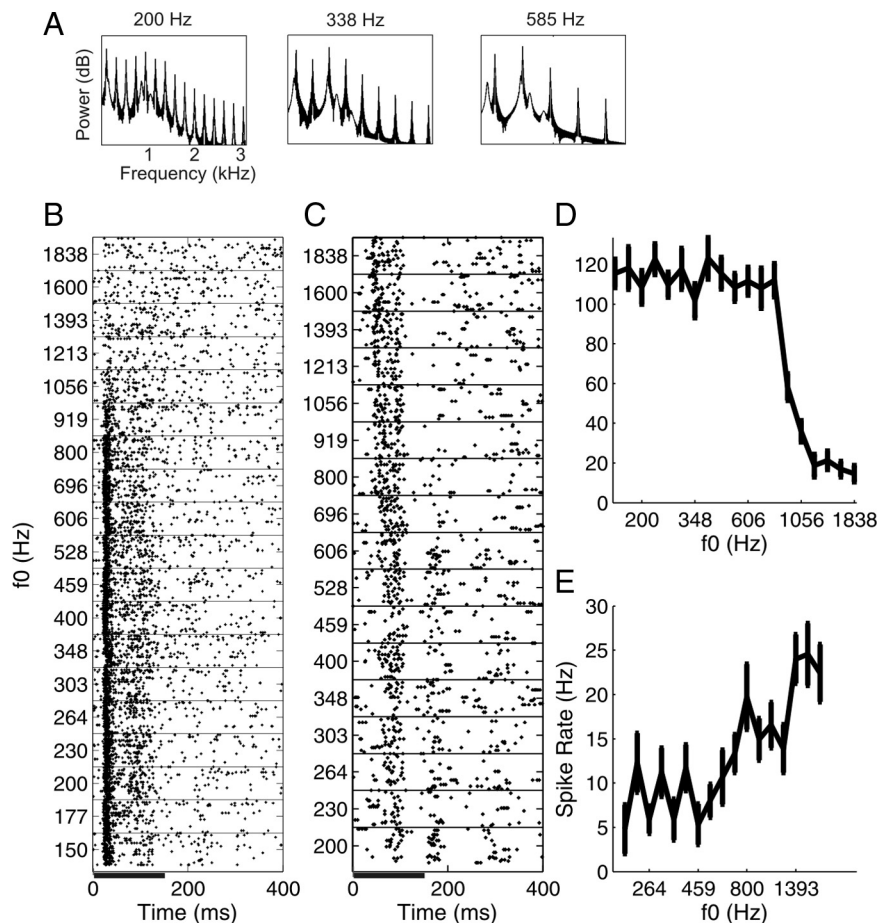
The responses of 744 isolated single units ( $n = 409$ , 53.5%) and small multiunit clusters ( $n = 355$ , 46.5%) were analyzed in MATLAB. We found  $f_0$ -sensitive units in all five cortical fields that were sampled. All 744 units included in our analysis were significantly driven by the vowel sounds (paired *t* test,  $p < 0.05$ ). Multiunits were included in our population analysis because we did not implement decoding methods that relied on the precise temporal response structure of single neurons.

**Behavioral methods.** Five animals participated in psychophysical experiments to determine pitch discrimination. Behavioral data from four of these animals have been reported previously (Walker et al., 2009). The stimuli consisted of the same artificial vowels used for electrophysiological recordings. The training and testing methods are described fully by Walker et al. (2009). Briefly, ferrets were trained to lick a central “start” spout. This triggered the presentation of a pair of artificial vowel sounds. The first vowel was a reference sound and was held at a fixed  $f_0$  for each weekly testing run. The second vowel was a target and was either higher or lower in  $f_0$  than the reference. The animal’s task was to respond to a spout located to the right of the central spout when the target vowel was higher in pitch and to a spout located to the left of the start spout when the target had a lower pitch than the reference. Animals were water restricted during the testing period and received a water reward when they responded correctly. If the animal made an incorrect choice, then this was signaled by a brief burst of broadband noise and a 14 s timeout. After an incorrect response, the ferret received a correction trial in which an easier target in the same pitch direction was presented. Correction trials served to prevent an animal from developing a bias toward a preferred response direction and were repeated until an animal made the correct response but were excluded from subsequent analyses. Plots of the proportion of trials in which the animal responded at the right spout as a function of the log of the target  $f_0$  were sigmoidal in shape and approximated a cumulative Gaussian function. Therefore, psychometric curves were estimated by fitting cumulative Gaussian distributions to the data using probit generalized linear models. The maximum slope of the psychometric function was used to quantify performance.

## Results

### Single-neuron periodicity sensitivity

We recorded responses of ferret auditory cortical neurons to artificial vowel sounds that varied in fundamental frequency ( $f_0$ ). The power spectra of three example vowels are shown in Figure 1A. We recorded from five anesthetized animals and one awake,



**Figure 1.** Stimuli and neural responses. **A**, Amplitude spectra of three example vowel sounds with their fundamental frequency,  $f_0$ , which determines the perceived pitch, indicated. **B**, **C**, Spike raster plots showing the responses of two units to 150-ms-long artificial vowel sounds with different  $f_0$  values. **D**, **E**, Spike rate functions (mean  $\pm$  SEM, calculated over 150 ms beginning at stimulus onset) for these two units. Typically,  $f_0$ -sensitive units responded to increasing  $f_0$  with a decrease (as in **B** and **D**) or increase (**C** and **E**) in spike rate rather than being tuned to a specific value. The unit in **B** was a small multiunit recorded in field PSF, and the unit in **C** was a single unit recorded in field PSF. Both units were recorded in anesthetized animals.

passively listening animal (see Table 1). The responses of 744 single units and small multiunit clusters were analyzed from two core and three belt areas of auditory cortex. Recordings were assigned to cortical fields on the basis of their pure-tone tuning characteristics and location on the ectosylvian gyrus (Bizley et al., 2005). All the data from the awake ferret were obtained from A1, whereas recordings were performed from all five fields under anesthesia (Table 2). Fifty-five percent of recordings were from single units, but, because we observed no clear differences between the periodicity sensitivity of neurometric functions for the single units and multi-neuron clusters (Wilcoxon’s rank sum test for single neuron vs multiunit neurometric slope,  $p = 0.26$ , across all units; identical tests separating units into their respective cortical fields yielded  $p$  values of  $>0.1$ ) or between awake and anesthetized recordings (Wilcoxon’s rank sum test,  $p = 0.34$ ), data were combined across these groups.

Sensitivity to  $f_0$  was tested using both regression analysis and ANOVA. A regression analysis of stimulus  $f_0$  on spike count (calculated over 150 ms beginning at stimulus onset) was performed for each unit, and, if significant ( $p < 0.05$ ), the sign of the regression slope was used to classify each as high-pass or low-pass for  $f_0$ . Thirty-seven percent of recordings had significant slopes, of which half (49.7%) were classified as low pass and half (50.3%) as high pass. To determine whether other neural units might exhibit

a non-monotonic sensitivity to  $f_0$ , we also tested the relation between  $f_0$  and spike rates with a one-way ANOVA. Another 12% of neural units were found to have their spike rate modulated by  $f_0$  in this manner, and two examples can be seen in supplemental Figure 2 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). This class included nonlinear high-pass and low-pass units, as well as a small number whose responses were band-pass tuned or varied in a more complicated manner with  $f_0$ . All driven units were included in subsequent analysis, regardless of whether they were classified as being significantly modulated by stimulus  $f_0$ .

We assessed the sensitivity of our units to the periodicity of vowels and then compared this sensitivity with the performance of ferrets in discriminating the direction of  $f_0$  changes in the same sounds. Periodicity-sensitive neurons were located in all five sampled cortical fields and spanned the whole range of possible pure-tone characteristic frequencies (supplemental Fig. 3, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Representative examples from two such units are shown in the raster plots in Figure 1, *B* and *C*. The first fired robustly in response to sounds with an  $f_0$  lower than  $\sim 900$  Hz, but its response declined rapidly and monotonically for  $f_0$  values  $> 1000$  Hz (Fig. 1*B*). In contrast, the unit in Figure 1*C* increased its firing rate when  $f_0$  was  $> 1200$  Hz. The spike rates ( $\pm 1$  SEM) for these two units are shown in Figure 1, *D* and *E*.

### Psychophysics

Five ferrets were trained in a task (described above) in which they heard two vowel sounds and were required to report whether the second (target) sound was higher or lower in pitch than the first (reference) sound (Walker et al., 2009). Two observations indicate that the animals solved this task by judging the pitch of the target sound rather than using other cues such as spectral or harmonic density. First, these animals immediately transferred pitch discrimination learning from a task using artificial vowel sounds to one using pure tones, which have very different spectral density (Walker et al., 2009). Second, two animals were also tested with artificial vowels generated from temporally “jittered” (i.e., randomized) click trains. Increasing temporal jitter disrupts the periodicity of the sound, thus reducing pitch salience but leaving the spectral density of the stimulus unchanged. Temporal jitter disrupted the ferrets’  $f_0$  discrimination performance, suggesting that they indeed responded to the (increasingly less salient) pitch rather than the (essentially unchanged) spectral density of the stimuli (supplemental Fig. 4, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). A typical psychometric function from one animal is shown in Figure 2*A*. Data from five animals are pooled in Figure 2*B* to illustrate performance across all reference pitches, and the slopes from each animal’s individual testing runs are plotted in Figure 2*C*. The slopes do not vary significantly as a function of the reference  $f_0$  within the tested range (Kruskal–Wallis test,  $p = 0.077$ ). For the purpose of comparing behavioral and neural discrimination measurements, we defined the “normal” range of behavioral performance as the 2.5th to 97.5th percentiles of the range we observed in our trained ferrets. The median value was 34%/octave (Weber fraction of 0.24), with lower and upper bounds of 18%/octave (Weber fraction of 0.52) and 64%/octave (Weber fraction of 0.12), respectively.

### Neurometric analysis

Having established that auditory cortical neurons are modulated by the  $f_0$  of artificial vowels, we sought to examine whether

these responses might provide the physiological signal on which ferrets’ pitch discrimination decisions are based. A simple algorithm was used to estimate the  $f_0$  discrimination “performance” supported by the responses of neural units (“neurometric” performance), in a manner that could be compared with the psychometric estimates of  $f_0$  discrimination measured for ferrets in our behavioral task. During behavioral testing, animals were presented with a range of  $f_0$  values spanning 2 octaves ( $\pm 1$  octave from the reference frequency). Neurometric curves were generated in a similar manner from the responses of neural units obtained over an  $f_0$  range of  $\pm 1$  octave, around a particular reference  $f_0$ . Because the  $f_0$  sensitivity of each unit was measured in  $1/5$ -octave steps over an  $\sim 3$  octave range, we were able to calculate neurometrics over a 2-octave range for each of at least six different  $f_0$  reference values. For each reference  $f_0$ , we calculated the median spike rate of the responses to 20–40 presentations of the corresponding artificial vowel stimulus.

We then asked whether the number of spikes evoked by each individual response to all other vowels in the  $f_0$  range was greater than the median response to the reference. Thus, for each target  $f_0$ , we obtained a proportion of trials wherein the spike rate response exceeded the median response to the reference  $f_0$ , and this relative spike count was used by a classifier to “guess” whether the target was higher or lower than the reference. Responses with spike counts that equaled the median reference response were randomly assigned to the “target higher” or “target lower” class with equal probability. For high-pass units, such as the one shown in Figure 1*C*, when the response to the target was stronger than the median reference response, the classifier guessed that the target  $f_0$  was higher than the reference. Conversely, for low-pass units (Fig. 1*B*), a greater target firing rate indicated that the target was lower than the reference  $f_0$ . Using this very simple firing rate comparison to decode each of the observed responses in turn, we obtained the percentage of “higher” responses our classifier made for each target sound, which can be directly compared with the percentage correct scores obtained in our behavioral pitch discrimination experiments.

As with the psychometric curves, neurometric functions were obtained by fitting cumulative Gaussian curves to the observed proportions of “higher” responses for each target  $f_0$ . In this manner, we calculated neurometric functions for each of the 634 neural units at all possible reference periodicities within the tested range. Our analysis did not assume a priori that units were either high pass or low pass. Instead, we computed neurometric functions for each unit twice, once assuming that a greater firing rate indicated an increase in  $f_0$  and once with the opposite assumption. The neurometric function with a positive slope was then selected for additional analysis. Single-neuron neurometric functions derived from spike count values were calculated over 30, 75, 150, and 300 ms response windows. A response duration of 150 ms (beginning at stimulus onset) was found to provide the best neurometric slopes for most neurons (supplemental Fig. 5, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material), and therefore all neurometric functions reported here calculate spike counts over this time period.

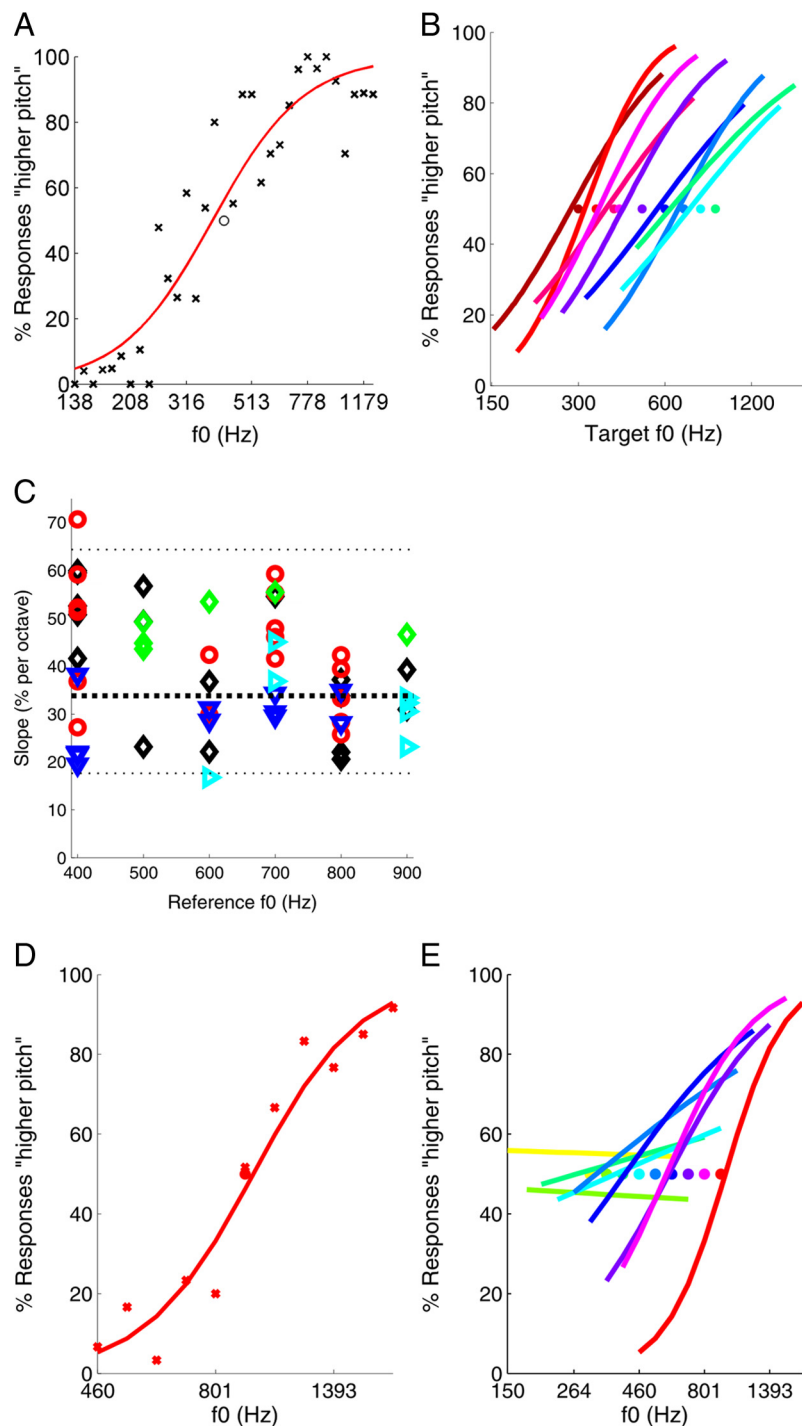
Figure 2*D* shows the neurometric function obtained from the unit in Figure 1, *B* and *D*, for a reference  $f_0$  of 919 Hz and a response window of 150 ms after stimulus onset. The neurometric function obtained for this unit at this reference  $f_0$  had a slope of 64%/octave, which is within the normal range of ferrets’ behavioral pitch discrimination performance. However, such high neurometric slopes were only rarely obtained from individual units. Also, in this instance, a reference  $f_0$  of 919 Hz lies close to

the steepest, most informative part of the  $f_0$  sensitivity curve of the unit (Fig. 1D). Selecting reference periodicities increasingly farther away from this value resulted in a pronounced and systematic decline in neurometric performance, as illustrated in Figure 2E.

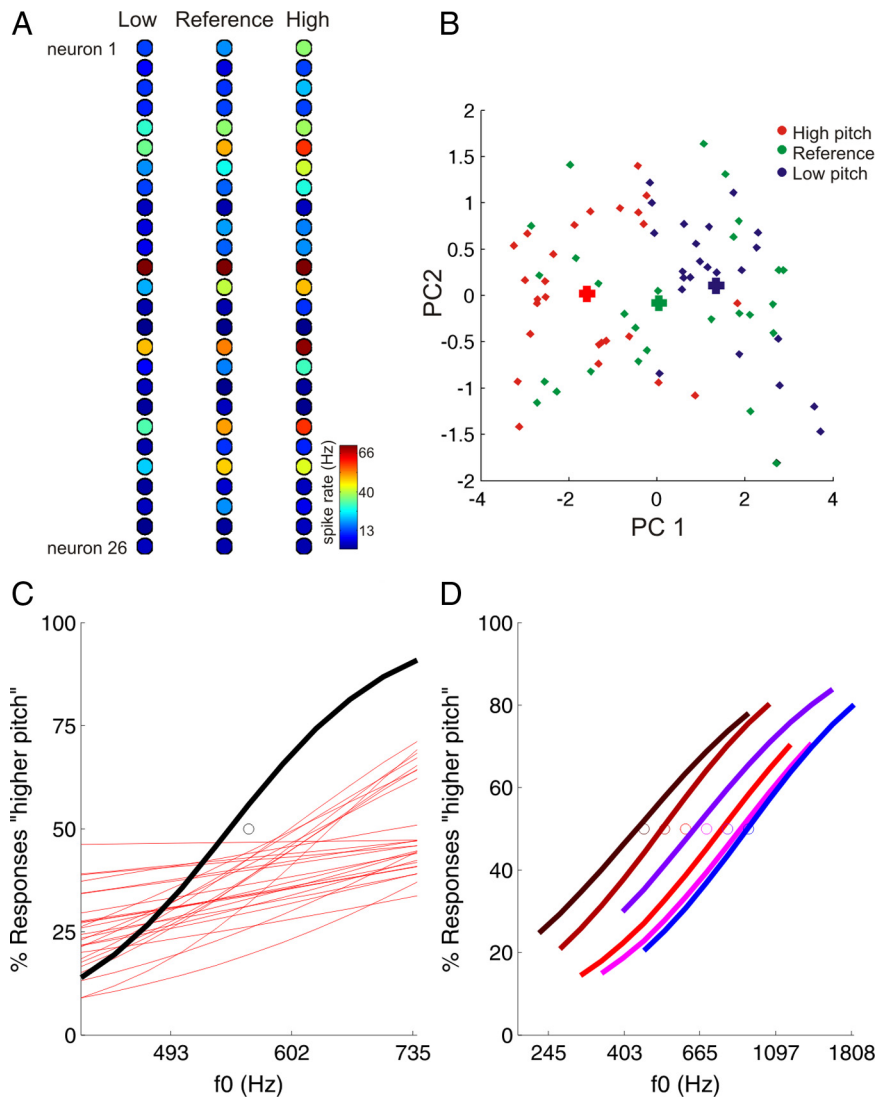
Only a small fraction (13 of 634) of the units in our sample had neurometric functions whose slopes fell within the trained ferrets' normal behavioral sensitivity range (i.e., slope between 18 and 65%/octave), and none of our neural units could match behavioral sensitivity across the full range of reference  $f_0$  values tested. Thus, to distinguish upward from downward periodicity changes on the basis of the firing rate of individual cortical neurons, the ferret's brain would either need to select the "best neuron" for the appropriate periodicity reference or, more likely, combine information across multiple neurons in the population. Neurons throughout the CNS typically form divergent and convergent connections with many other neurons, an anatomical arrangement probably more conducive to integrating responses across sizeable ensembles of neurons than the targeted selection of a few. We therefore developed neurometric algorithms that combine information across ensembles of simultaneously recorded units to test their ability to inform sufficiently accurate periodicity judgments robustly over a range of reference periodicities.

### Ensemble neurometrics

Simultaneous recordings from multiple neural units ("ensembles") were obtained by recording with multisite electrodes (typically 32 channels, configured as  $4 \times 8$  site linear arrays or  $2 \times 16$  channel linear arrays; for details, see Materials and Methods). Our sample of 744 units could therefore be grouped into 58 ensembles, each of which reflects the activity recorded with one 32-channel array. In some fortuitous cases, we obtained acoustically driven activity at all 32 recording sites and it was possible to discriminate more than one spiking unit from many of the recording sites, whereas at other penetrations only a few of the recording sites yielded neural responses. Consequently, the number of simultaneously recorded units in these ensembles varied from one penetration to the next (maximum of 61; mean  $\pm$  SD,  $15.3 \pm 13.2$ ). Thirty-nine of the 58 recorded ensembles were tested across a wide enough range of stimulus periodicities to analyze performance across multiple reference values, and results from



**Figure 2.** Psychometric and neurometric periodicity discrimination performance. **A**, Psychometric function from one ferret. The raw data (crosses) and fitted function (red line) are shown, with the circle indicating the  $f_0$  of the reference sound. This comprised  $\sim 900$  trials (30 per data point) collected over 10 twice-daily testing sessions. The percentage correct scores for each frequency were fitted with a cumulative Gaussian, resulting in the plotted psychometric function. **B**, Psychometric data pooled from five animals across a range of reference  $f_0$  values, indicated by the circle plotted in a matching color. **C**, The slope of the psychometric function can be used as a measure of performance. Slopes of psychometric functions for five ferrets across a range of reference  $f_0$  values. Heavy dashed line indicates the median level of performance with 5% confidence intervals (2.5th and 97.5th percentiles) indicated from the stippled lines. **D**, Neurometric function for one cortical unit. Classifications at each target  $f_0$  are indicated by the crosses, and the line shows the fitted neurometric function. The slope of this neurometric is 65%/octave. Because this was a low-pass unit, a trial in which the spike count was higher than the reference signaled that the target  $f_0$  was lower than the reference. The "% Responses higher pitch" on the y-axis therefore reflects the proportion of trials in which the spike count was lower than the median spike count in response to the reference  $f_0$ . **E**, Neurometric functions for the same unit across a range of reference  $f_0$  values. Each function is colored separately with a matching circle to indicate the value of the reference  $f_0$ . Note that, as the reference  $f_0$  decreases, the capacity of the neural response to discriminate between the target sounds decreases to chance levels.



**Figure 3.** Deriving neurometrics from ensembles of neurons. **A**, Schematic showing the mean spike rates of 26 simultaneously recorded units in response to a reference  $f_0$  of 919 Hz (middle column) and to  $f_0$  values one octave below (low) or above (high) the reference sound. The color scale indicates the spike rate calculated over a 75 ms window. **B**, The colored diamonds show individual responses of the same ensemble of 26 units to 30 repetitions of each of the high (red), low (blue), and reference (green)  $f_0$  sounds. To facilitate visualization, the spike count vectors for these units were projected onto the first two principal components (PC). The average response to each is represented by the appropriately colored cross. **C**, Spike rate neurometrics are shown in red for each of the 26 simultaneously recorded units in the same ensemble. Overlaid in black is the ensemble neurometric computed using our pattern recognition algorithm. The slope of this neurometric function lies within the range of psychometric performance thresholds for ferrets in this task. The performance of the ensemble also exceeds that of any of the individual units. **D**, The ensemble neurometric of the same 26 units tested over six different reference  $f_0$  values.

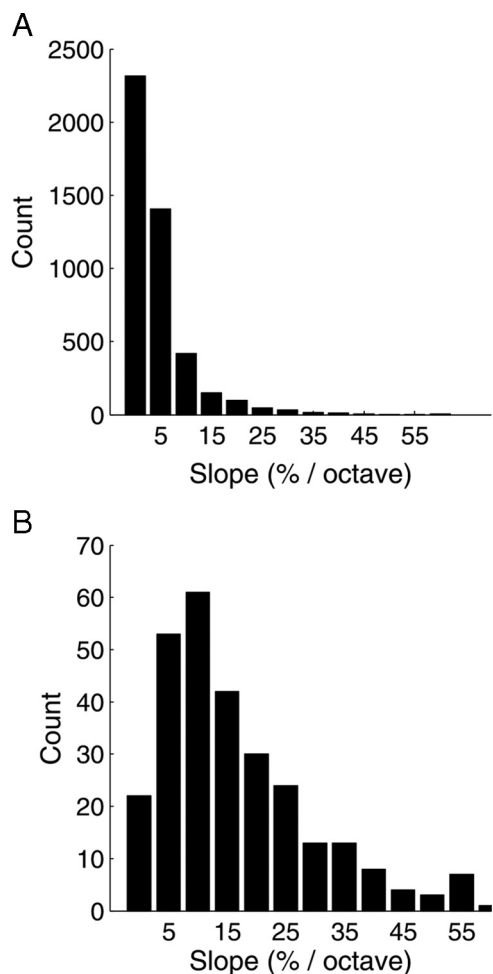
these 39 are reported here. Although the simultaneously recorded units could come from sites separated on the cortical surface by up to 600  $\mu\text{m}$ , we ensured that these were always within a single cortical field.

We implemented a novel decoding method in which responses were classified based on the pattern of activity across each ensemble. We first calculated the number of spikes in the response of each units on each trial over a specified time window (30, 75, 150, or 300 ms). The array of spike counts across the ensemble on a single trial could then be described as a point in a high-dimensional space, in which each dimension corresponded to the spiking response of each individual unit. Thought of in this way, the ensemble responses to repeated presentations of the same stimulus form a “cloud” in this high-dimensional “spike rate

space,” and the average response to a particular sound can be thought of as the center of that cloud. If neural responses to repeated presentations of the same stimulus are highly variable, then they scatter widely throughout the space (the cloud is diffuse), but if responses are more reproducible, then the region occupied by the responses will be more compact. The regions of spike rate space occupied by the responses to stimuli of different  $f_0$  might be quite distinct, making classification and stimulus decoding relatively easy, or they could be partly or wholly overlapping, making classification more difficult.

We decoded the response of each ensemble in a two-stage process. First, we asked whether the ensemble activity could distinguish between the reference and target sounds. For each presentation of the target, we asked whether the response on that particular trial was closer (i.e., had a smaller Euclidean distance from) the average response to the reference or to the average response to target sounds of the same  $f_0$ . If the response to the target was closer to that of the reference, the two sounds were deemed to be indistinguishable, and such trials were randomly classified as either higher or lower with equal probability. More commonly, however, the response on an individual trial was closer to other responses to the same target than to the average reference response, and the decoding algorithm then made a high/low periodicity judgment by determining whether this target response was more similar to the average response to the highest target in the  $f_0$  range or the lowest  $f_0$  response (i.e., vowels 1 octave above or below the reference). The proportions of trials classified in this manner as higher than the reference  $f_0$  were then fitted with a cumulative Gaussian function to obtain the ensemble neurometric function.

Figure 3A illustrates schematically the mean spike rates from one ensemble, comprising 26 neural units, in response to a single presentation of the reference sound and sounds with  $f_0$  values 1 octave above and below the reference. It is clear that some units are rather uninformative, given that their spike rate changes little over the three conditions shown, whereas others modulate their spike rate appreciably as a function of  $f_0$ , as indicated by the change in color. The individual responses of these same 26 units to 30 presentations of the reference and the highest and lowest targets are shown in Figure 3B where, for visualization purposes only, the spike rate vectors are plotted in principal components space. These data suggest that the responses should be relatively easily classifiable: the responses to the high (red) and the low (blue) periodicity sounds are well separated, and the responses to the reference  $f_0$  (green) lie between the two. Figure 3C shows, in black, the neurometric curve for this ensemble of 26 units, overlaid on the neurometrics ob-



**Figure 4.** Distribution of slopes of the neurometric functions. *A*, Distribution of individual unit neurometric slopes (all units, all reference  $f_0$  values). *B*, Distribution of ensemble neurometric slopes (all ensembles, all reference  $f_0$  values).

tained for each individual unit in the dataset (shown in red). The slope of the ensemble neurometric exceeds that of all of the individual neurometric functions (ensemble neurometric, 50.5%/octave; individual unit neurometrics, mean of 14%/octave, maximum of 37%/octave) and lies within the normal range of psychometric function slopes for this task.

#### Ensemble neurometrics are more sensitive and robust across a wider range of stimulus parameters

Figure 4 shows the distribution of slopes computed using individual unit neurometrics (*A*) and population neurometrics (*B*). The mean individual unit neurometric performance is 4.6%/octave, with only 3.8% of units having slope values that exceeded the lower limit of performance we observed in trained ferrets (slopes >18%/octave). In contrast, the mean ensemble neurometric slope is 16.3%/octave, and approximately one-third (35.2%) of all groups of simultaneously recorded units yielded neurometric curves that fell within the behavioral range.

We also examined how well our ensembles performed relative to individual units across a range of reference  $f_0$  values. Figure 3*D* plots the ensemble neurometric functions across reference  $f_0$  for the same set of 26 units shown in Figure 3*A*. This sample ensemble gave a particularly stable neurometric performance. In general, the ensemble neurometrics were less affected by changes in the reference  $f_0$  than individual unit neurometrics (compare with

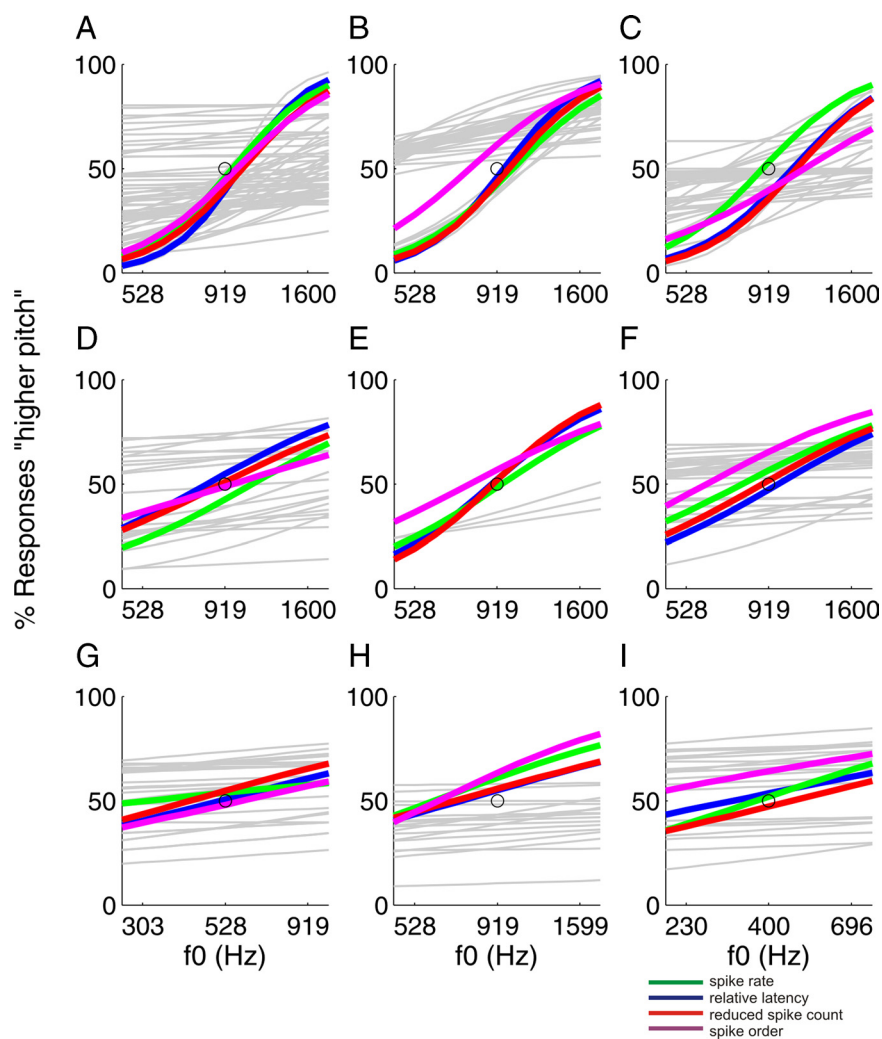
Fig. 2*E*). To assess this, we took a robust encoding of stimulus  $f_0$  to mean that the neural sensitivity, as assessed by the slope of the neurometric curve, should reach or exceed the “minimal psychophysical performance criterion” for a wide range of reference  $f_0$  values. The minimum behavioral criterion here corresponded to a neurometric slope of 18%/octave, corresponding to the 2.5th percentile of the slopes of the observed psychoacoustic functions. Twenty-five of our neural ensembles yielded neurometrics that reached the behavioral criterion for at least some reference  $f_0$  values. For these 25 ensembles, we simply counted the number of reference  $f_0$  values for which either the ensemble neurometric or the neurometrics of any of the individual units exceeded the 18%/octave criterion. Ensemble neurometrics yielded exceeded the behavioral criterion for a larger number of reference  $f_0$  values than single neuron neurometrics in 22 of these 25 ensembles. In the remaining three ensembles, the count was equal. In no case did single units yield supra-criterion neurometrics over a wider range of  $f_0$  reference values than ensembles.

To further demonstrate that the neurometric discrimination indeed used the pattern of activity across units and was not simply dominated by one or very few highly selective units, we recomputed the ensemble neurometrics repeatedly, on each occasion excluding a different unit. This had a negligible effect on the resulting neurometrics (i.e., the observed neurometric based on the whole population always fell within the 95th percentile of the bootstrapped values based on ensembles made by excluding one neuron). Two examples are shown in supplemental Figure 6 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). As an additional control, we also tested ensembles in which the best unit was excluded from the ensemble, then the top two units, the top three, and so on (supplemental Fig. 7, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). This revealed a gradual decrease in performance consistent with ensemble neurometric performance being based on the distribution of activity across a number of units rather than merely reflecting the performance of just the best unit.

#### Reading the code: alternative neural response measures

Potentially, several aspects of the neural discharge patterns could carry information. So far, we have only considered spike rate. Neural responses to different  $f_0$  values tended to show similar discharge patterns, but sometimes differed in their onset latencies (Fig. 1*B, C*). We therefore investigated whether the relative timing of spikes across the ensemble carried information about stimulus  $f_0$ . For each ensemble, we extracted the first time a spike occurred, across all units, after the stimulus onset. We then computed the first spike latency of all other units relative to this first ensemble spike. In this manner, we obtained vectors of relative spike latencies for the ensemble, which were then decoded using the same pattern-matching algorithm described for spike counts. Spikes were considered across different response durations (30, 75, 150, or 300 ms), beginning at stimulus onset. Trials in which a unit failed to fire at any point during the response window were assigned a latency value that was 1 ms greater than the maximum response duration under consideration. We also considered a reduced spike count code, which simply asked whether or not a spike occurred during the response window. This was essentially a binary “spike or no-spike” code. A final, fourth code represented ensemble responses by the order in which the units fired. All of the spike latencies occurring within a trial were ranked, and these ranks were used as the input to the decoder.

The choice of code made very little difference to the slope of the neurometric functions. Figure 5 shows the four alternative



**Figure 5.** Exploring alternative neural codes. **A–I**, Nine examples of ensemble neurometric functions obtained using different decoding strategies. **A–C** represent three of the best performing ensembles, **D–F** lie in the midrange of performance, and **G–I** are some of the more poorly performing examples. In each case, individual unit spike-rate neurometrics are plotted in gray, with the colored lines illustrating the ensemble neurometric derived from the spike rate, relative latency, reduced count, and spike order codes (shown in green, blue, red, and magenta, respectively).

ensemble neurometrics along with their single-unit neurometrics from nine example ensembles. These nine represent three of the best (Fig. 5A–C), three average (Fig. 5D–F), and three of the worst (Fig. 5G–I) ensembles in our dataset. Supplemental Figure 7 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) shows, schematically (as in Fig. 3A) the behavior of three different ensembles (comprising 35, 3, and 26 units, respectively) when decoded with each of the four neural codes.

The “best code” for each ensemble was defined as the response code that resulted in the steepest neurometric function. Across all penetrations, response durations, and reference periodicities, the best code was distributed relatively equally between the four candidate codes (spike count, 298; relative latency, 289; binary count, 291; spike order, 246), and there were, overall, no significant differences in the sensitivity (i.e., average neurometric slope) between these alternative codes (Kruskal–Wallis test,  $p = 0.077$ ). However, if only the longer response durations were considered, then there were small but statistically significant differences in the performance of each of the decoding strategies (Kruskal–Wallis test,  $p < 0.0001$ ). For example, with a 150-ms-wide response window, 32% of units performed best with the spike count code, 28%

with relative latency, 24% with reduced spike count, and 16% with the spike-order code. When the slope values are compared at this duration (supplemental Fig. 5B, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material), those units that performed best with the relative latency code had significantly greater slopes than those performing best with the spike count code (*post hoc* Tukey–Kramer test,  $p < 0.05$ ).

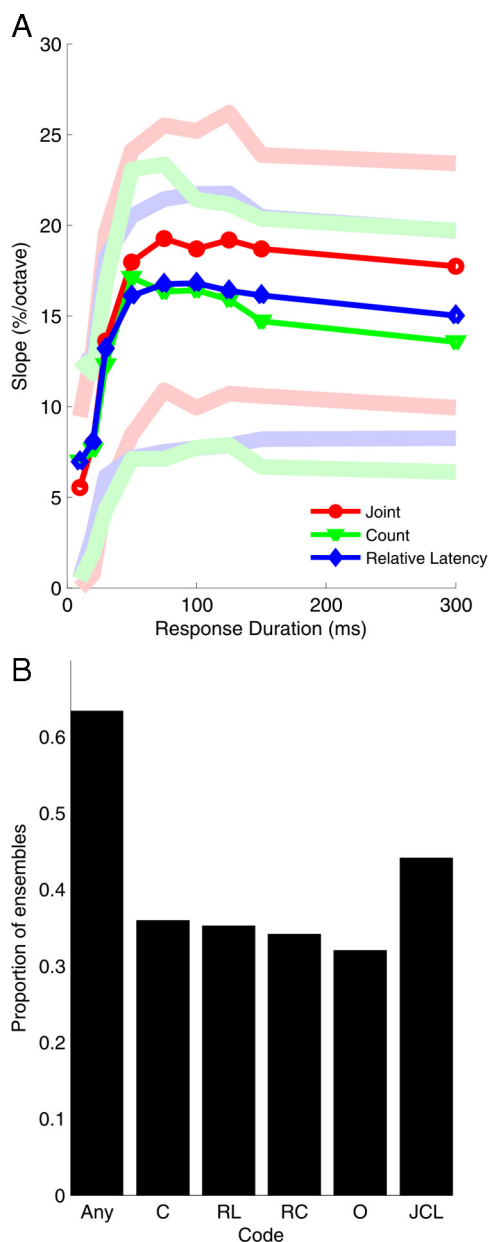
### Combined count and latency coding

To examine the extent to which these coding strategies were redundant, we also considered a code that used both spike count and relative latency information. Joint count/latency ensemble decoding was performed by first normalizing the relative latency and spike count values separately to their maximum and then using these values as inputs to the pattern classifier. Each neuron in the ensemble is therefore represented by a  $2N$ -element vector, containing  $N$  normalized spike count and  $N$  normalized latency values (spike count and latency values were separately normalized to the maximum count or latency observed across all trials and units in the ensemble). This joint code produced a small but significant improvement in performance when compared with either the spike count or relative latency code alone, indicating that both codes carry independent information (Kruskal–Wallis test for slope values at 150 ms,  $p < 0.01$ ; Tukey–Kramer *post hoc* comparisons,  $p < 0.05$  for both joint code-count and joint-latency comparisons). Figure 6A illustrates this by showing the average (and the 25th and 75th percentile) performance for ensemble spike count, relative latency, and the joint count/latency codes. For all three (combined, count, and relative latency) codes, performance initially increases equally rapidly as the response window is increased from 10 ms but then reaches a maximum near 75 ms. The joint code provides ~10% better periodicity discrimination than either the spike count or relative latency codes by themselves. A two-way ANOVA showed significant effects of both response duration and code type (response window duration,  $F = 104$ ,  $p < 0.0001$ ; code choice,  $F = 18.0$ ,  $p < 0.0001$ ). Additional analysis of the effect of response window length on all four codes is included in supplemental Figure 5, C and D (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). The proportion of ensemble–reference combinations that exceeded the behavioral threshold was calculated for each of the four codes and the joint count–latency code (Fig. 6B).

### Correlated firing can limit performance

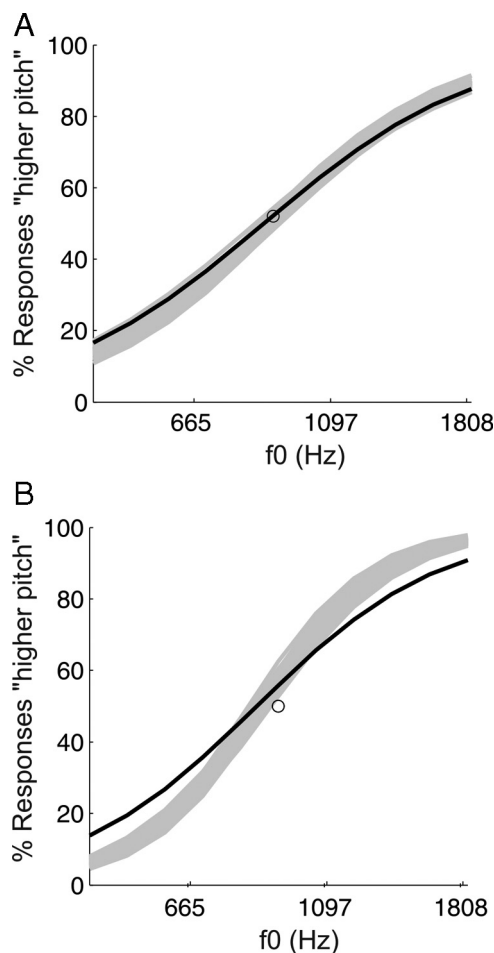
Correlations in neural responses have been shown previously to limit neurometric performance in some cases (Zohary et al., 1994; Walker et al., 2008) and to improve it in others (Wang et al., 2007). To examine the effects of “noise correlations” (i.e., correlated background activity) on performance, we recomputed the





**Figure 6.** *A*, Joint count and relative latency coding. Average neurometric performance as a function of the analysis time window (time after stimulus onset) is shown for all ensembles at all reference periodicities. Additional response time windows were considered for the relative latency, spike count, and joint codes to explore the timescales over which each code became informative. The saturated lines show the mean, and the faint, unsaturated lines show the 25th and 75th percentiles. Values for the spike count code are plotted in green, the relative latency code in blue, and the joint spike count/relative latency code in red. *B*, Proportion of ensemble–reference combinations that exceeded the minimum behavioral threshold of 18%/octave for each code type. The proportion of ensembles that exceeded this level with any code (including the joint count–latency code), spike count (C), relative latency (RL), reduced count (RC), spike order (O), and joint count–latency code (JCL).

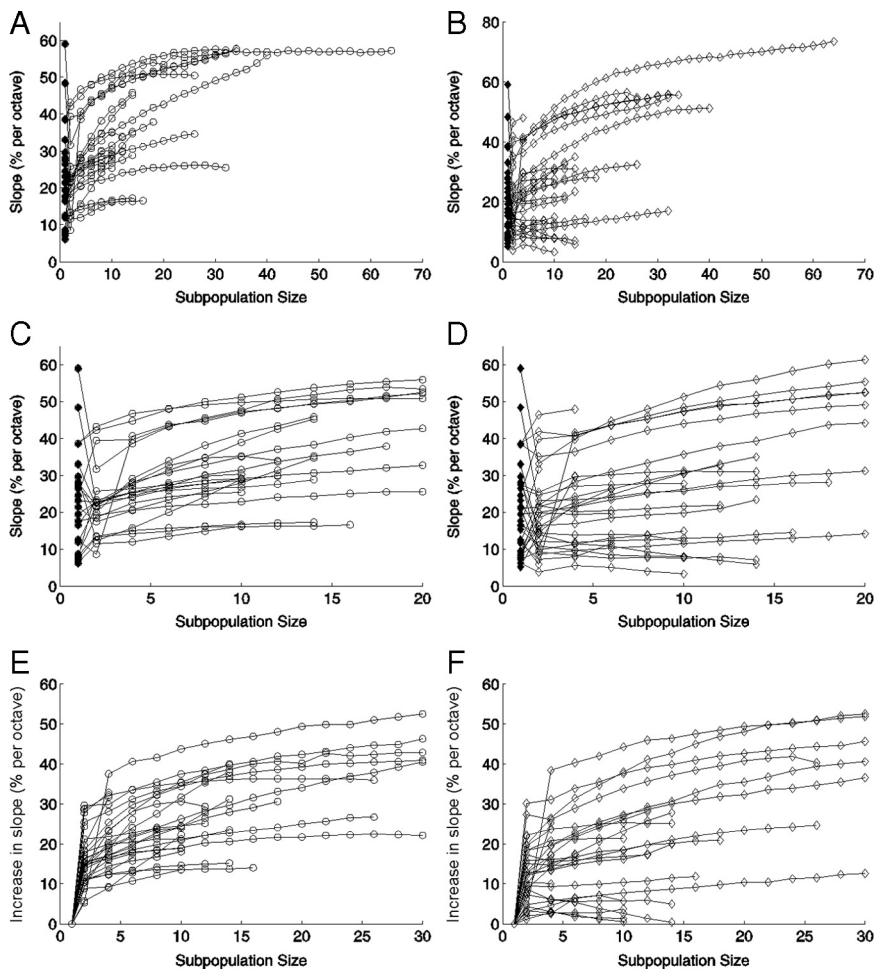
ensemble neurometrics after randomizing (i.e., shuffling) the order of stimulus presentations independently for each unit in an ensemble, so that any correlations in neural responses that might be attributable to common fluctuations in background activity were removed. This procedure was repeated 100 times, and the 5th and 95th percentiles of the shuffled neurometric slopes were used to estimate significance limits. Figure 7, *A* and *B*, illustrates the results of performing this shuffling procedure on two ensem-



**Figure 7.** Effects of correlated noise and response duration on neurometric functions. *A, B*, Neurometric functions from two ensembles, before (black lines) and after shuffling of the responses to desynchronize the ensemble response (gray regions indicate 5th to 95th percentiles of the shuffled responses). *A*, An example in which the ensemble neurometric curves derived from asynchronously recorded responses were slightly steeper (5th percentile bootstrapped slope, 45.5%/octave; 95th percentile, 55.2%/octave) than that obtained from the synchronous recording (43.2%/octave), indicating that correlated noise in the neural ensemble limited performance by a small, but significant, amount. *B*, A more extreme example, in which removing the noise correlations markedly increased performance (simultaneous recording slope, 50.4%/octave; 5th percentile bootstrapped slope, 67%/octave; 95th percentile, 78%/octave).

bles in which the neurometric curves derived from shuffled data were steeper than from the synchronized data.

In total, 20.9% of comparisons (based on 39 penetrations and all reference  $f_0$  values tested) showed a significant limiting effect of correlated noise (i.e., the neurometric computed from the simultaneously recorded ensemble had a slope value that was smaller than the fifth percentile of the shuffled data). These shuffling effects were found in 21 of our 39 ensembles and occurred across several reference periodicities. In contrast, only 5.8% of all shuffled versus synchronized neurometric comparisons yielded a significant decrease in their slopes when the responses were shuffled, and these changes were smaller on average. In comparisons in which shuffling significantly impaired neurometric performance, the average difference between shuffled and synchronized neurometric slopes was 4.8%/octave, but ensembles that provided better neurometrics when shuffled showed an 11.9%/octave slope increase. An identical analysis performed on ensembles decoded with a relative latency code yielded quite different results: whereas 12% of ensemble–reference combinations showed



**Figure 8.** *A–D*, Slopes of neurometric functions derived from subsets of increasing size for all ensembles of >10 units in total. The average performance of all possible subsets of a given size, or 500 randomly chosen ones, are plotted. The black diamond symbols plot the best individual unit neurometric in each ensemble. In *A* and *C*, ensemble activity is decoded with a spike count code and *B* and *D* with a relative latency code. *C* and *D* represent the same data in *A* and *B* but focuses on ensembles of <20 units. In *E* and *F*, the effect of population size is represented by plotting the increase in ensemble slope relative to the average unit neurometric. Ensemble activity in *E* is decoded with the spike count code and *F* with the relative latency code.

a significant limiting effect of noise, a far greater proportion of ensembles (57%) showed a significant decrease in performance after shuffling compared with the synchronous case. Ensemble-reference combinations that had better spike count neurometric performance when shuffled always also showed better performance when their relative latencies were shuffled. Most of the cases that showed a limiting effect of noise correlations for relative latency decoders also showed this effect with spike count. The largest differences between the effects of shuffling on spike count and relative latency decoding were in the number of populations that decreased in slope when responses were decorrelated and the magnitude of this effect, which was greater (−10.8% compared with 4.8%) for spike count. The effect of noise correlations on both codes for smaller sub-ensembles is considered in supplemental Figure 8 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

**Ensemble size**

We considered how neurometric performance was affected by ensemble size. To do this, we recalculated the neurometric slopes from increasingly large subsets of simultaneously recorded units.

The number of possible subsets of size *n* from an ensemble of size *p* is given by:

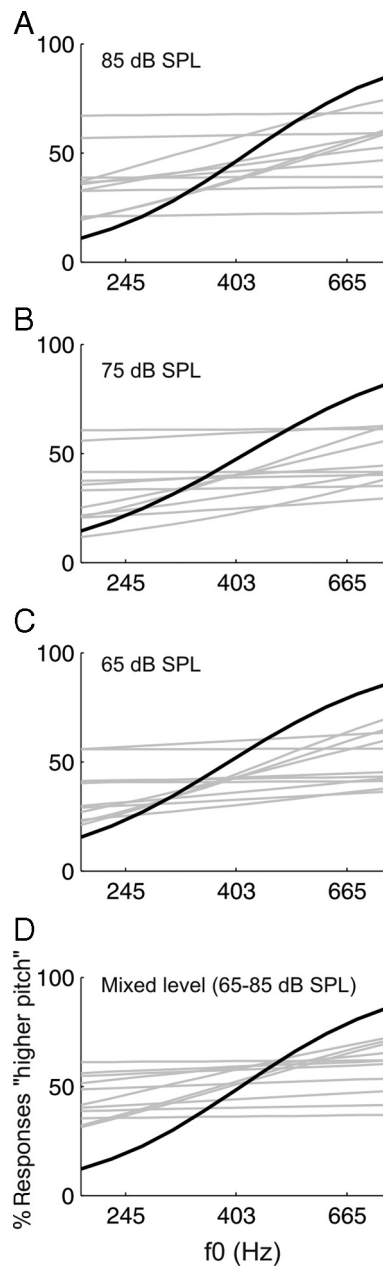
$$\frac{p!}{(p - n)!n!}$$

which can become very large for even modest ensemble and subset sizes. In cases in which there were >500 possible combinations, analyses were limited to 500 randomly selected subsets. We analyzed spike count and relative latency responses at the best reference *f*<sub>0</sub> and response duration of each ensemble. Our sample size was not sufficiently large to make quantitative conclusions about optimal ensemble sizes. However, in keeping with the finding that correlations between neurons mostly limited discrimination, for many ensembles, performance tended to increase with ensemble size and then asymptote (Fig. 8*A,B* for spike count and latency, respectively). Figure 8, *C* and *D*, shows the same data but focuses on the smaller subpopulation (*n* < 20), plotted relative to the best unit in the population and in *E* and *F*, plotted as improvement in slope, relative to the average unit performance. It is clear from these figures that even small subpopulations can substantially improve on the average unit performance and that the improvement with increasing population size is rapid.

**Ensemble activity encodes periodicity across a range of sound intensities**

Perceptual features of sound tend to vary little with sound level, and, during our behavioral testing of periodicity discrimination, sound levels roved randomly over a 15 dB range. Any neural code capable of supporting this discrimination behavior should therefore be capable of operating over a similar range of sound levels. In a subset of recordings (7 ensembles, 81 units), we presented vowel stimuli at three sound levels (65, 75, and 85 dB SPL), allowing us to examine how robust the neural representation of *f*<sub>0</sub> is when sound levels change. For each of these seven ensembles, we constructed neurometric functions at the best reference periodicity and sound level of each ensemble (i.e., the combination of level and *f*<sub>0</sub> that produced the steepest slope value). Six ensembles had slope values above the lower bound of the behavioral range (i.e., >18%/octave) and were therefore analyzed further. For each of these ensembles, we then constructed neurometric functions from all of the responses, pooling across all three sound levels at each *f*<sub>0</sub> value. We compared the slope of this neurometric function with that obtained with the best *f*<sub>0</sub>-level combination.

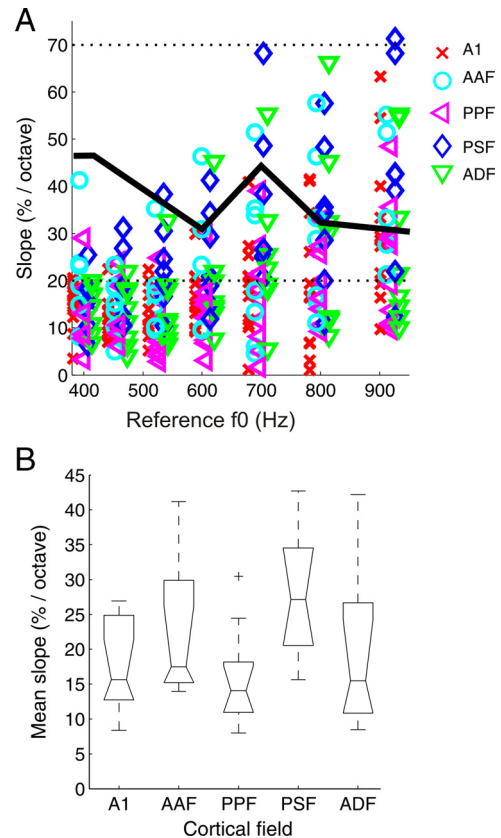
The two ensembles with the steepest neurometric functions (47.2 and 49.0%/octave, respectively) performed similarly when sound levels varied across this 20 dB sound level range (mixed sound level neurometrics were 46.9 and 50.3%/octave, respectively, i.e., performance was 99.4 and 102.7% of the best single level neurometric). Two ensembles showed only very modest de-



**Figure 9.** Periodicity discrimination across varying sound levels. Ensemble (black line) and unit (gray lines) neurometric functions for periodicity discrimination at 85 dB SPL (**A**), 75 dB SPL (**B**), 65 dB SPL (**C**), or mixed (65–85 dB SPL; **D**) sound levels. Ensemble neurometrics robustly encode periodicity information across this range of sound levels.

creases in performance when three sound levels were introduced (from 44.3 to 37.1%/octave and 42.7 to 38.3%/octave, i.e., performance with randomized sound intensity was 83.8 and 89.7% of their best intensity slope value). Of the remaining two ensembles, one dropped from 26 to 16%/octave and the other from 32 to 5%/octave. Very similar results were observed with the relative spike latency code: the mixed-level neurometric slope values were  $86 \pm 10\%$  (mean  $\pm$  SD) of the best periodicity–level combinations, across the six ensembles.

In contrast, single-unit neurometrics were much more heavily affected by changes in sound level. Of the 72 units that comprised these six ensembles, 10 had slopes at their best sound level–periodicity combination that exceeded 20%/octave (mean of  $30 \pm 5\%$ /octave). For these units, the mixed-level neurometric slope



**Figure 10.** Ensemble neurometrics. **A**, Neurometric function slopes for each of the 58 ensembles at all reference periodicities using the relative latency code and a response window of 75 ms. Ensembles from different cortical fields are shown using different symbols. The mean psychometric performance of four ferrets is overlaid in black, whereas the dotted lines mark the 20–70%/octave range within which the psychometric function slopes are normally found. **B**, Box plot illustrating the range of slopes obtained in each cortical field. The average slope (across all reference periodicities) for each penetration was used to examine inter-area differences. Total ensembles in each cortical field: A1, 10; AAF, 6; PPF, 10; PSF, 6; ADF, 7.

value was on average only one-third ( $33.5 \pm 33.2\%$ , mean  $\pm$  SD) of that at the best periodicity–level combination. There were only two neurons whose slope value for the mixed sound level neurometrics was  $>75\%$  of the best-level neurometric, and both of these had a very modest slopes to begin with (23 and 27%/octave, i.e., only just above our lower threshold). Figure 9 illustrates the neurometrics from one ensemble and each of its units, at each of the three sound levels (**A–C**), calculated across the three sound levels (**D**). Overall, the ensemble performance is substantially more robust to changes in sound level than single-unit performance, whether ensembles are decoded using spike counts or relative latencies.

#### Distribution of ensemble performance across cortical fields

Figure 10A plots the slopes of the neurometric functions for all 39 ensembles, at each  $f_0$  tested, using the relative latency code and a response window of 75 ms. Also shown for comparison are the mean and range of psychometric performance from the five trained ferrets. Across the full range of reference  $f_0$  values, neural ensemble neurometrics are often within the range of the behavioral performance of the ferrets. At most reference frequencies, some ensemble neurometric functions were as steep as the best psychometric functions.

Ensembles from different cortical fields are plotted with different symbols in Figure 10A, showing that all five fields contain

ensembles that represent sufficient information about  $f_0$  to account for the behaviorally assessed perceptual decisions. Figure 10*B* illustrates the distribution of neurometric slopes (across all reference  $f_0$  values) for penetrations in each of the cortical fields. Figure 10, *A* and *B*, suggests a trend for field PSF to be the most informative area for periodicity-direction discrimination, but this trend was not statistically significant ( $p = 0.058$ , Kruskal–Wallis test). However, a sample of 39 ensembles may not be large enough to reveal possible subtle inter-area differences. This issue is addressed in more detail in supplemental Figure 10 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

### Distribution of ensemble performance through cortical depth

Because recordings were made with multisite silicon electrodes, we often sampled across multiple cortical depths at the same time. Previously, we demonstrated that units recorded in superficial cortical layers had a higher pitch sensitivity than those recorded in deeper layers (Bizley et al., 2009). To examine whether the ability of neurons to encode the direction of a pitch change varied across cortical depth, recording sites were divided into “superficial” and “deep” locations according to whether they were within 800  $\mu\text{m}$  of the cortical surface. This depth marks the approximate location of the ventral border of layer IV of the ferret auditory cortex (Dahmen et al., 2008). We compared the single-unit neurometric slopes of superficial and deep units. There was no statistically significant difference between the neurometric slopes of deep and superficial layers (Wilcoxon’s rank sum test,  $p = 0.1244$ ). We then divided all ensembles into two subensembles comprising only superficial/deep units and computed ensemble neurometric functions (with spike counts calculated over a 75 ms window). Again, the superficial and deep ensemble neurometrics were not significantly different ( $p = 0.76$ , Wilcoxon’s rank sum test), and there was no difference between the slopes derived from either subgroup and the full ensemble slope ( $p = 0.19$ , Kruskal–Wallis test). We repeated this procedure while decoding responses with the relative timing measure and this time found that the slopes differed significantly across layers (Kruskal–Wallis test for slope values from the full ensemble, superficial, deep,  $p = 0.038$ ). *Post hoc* tests showed that this difference was based on the comparison between the full ensemble slope and the neurometric slopes of the deep units, in which the latter were significantly shallower.

### Alternative decoding methods

We tested a range of methods for decoding individual unit and ensemble responses in addition to those reported above. Unit responses were also decoded with pattern recognition algorithms (using methods similar to those reported by Walker et al., 2008), but these did not differ quantitatively from simple spike count measures. Ensemble responses were also decoded using linear discriminant analysis, a simple perceptron learning model, a dot-product classification (which classified responses along vectors that pointed from the responses to the reference to the target directions), and, finally, using the Euclidean pattern-matching algorithm described but using either transformed (cube-rooted) spike counts or spike counts summarized using principal components analysis. These methods are conceptually similar to the algorithm we used to obtain the results described above, and they generally yielded very similar, although typically slightly worse, results.

We considered one last decoding strategy motivated by our finding that we were unable to distinguish between the possibilities that ferrets compared the target and reference sounds on

each trial or, alternatively, built up an internal representation of either the reference or the highest and lowest targets and then compared individual target sounds with this internal representation (discussed by Walker et al., 2009). We therefore assessed neurometric performance for ensembles while excluding the first stage of the decoding strategy (i.e., we did not first ask whether the response to the target  $f_0$  was differentiable from the response to the reference sound). Omitting this step produced a significant decrease in performance (paired *t* test,  $p < 0.01$ ). However, the magnitude of these differences was very small, with only 5% of ensembles neurometrics having slope values for the single step algorithm that were  $>8\%$ /octave poorer than the two-step algorithm. We reported findings based on the two-step algorithm because it was slightly more effective than omitting this first stage, but our results do not substantially differ when we use the one-step method or any of the other alternative neurometric algorithms mentioned above. For individual neurons and ensembles, there were often occasions in which one of the alternative methods was at least as successful as those we ultimately used. Thus, overall, the simple two-step neurometric approach described produced better results (i.e., steeper neurometric slopes) than the alternatives tested.

### Discussion

We recorded responses to artificial vowel sounds of varying periodicity in five auditory cortical fields and compared them with the performance of ferrets trained to discriminate the pitch of these sounds. Temporal acuity decreases above the cochlear nucleus with a progressive increase in a rate-based representation of temporal information, including stimulus periodicity, at higher levels of the auditory pathway (Wang et al., 2008). The importance of auditory cortex in pitch extraction is illustrated by human imaging studies showing cortical activation by pitch-evoking sounds (Patterson et al., 2002; Penagos et al., 2004). Moreover, cats with bilateral auditory cortex lesions are impaired in a missing-fundamental pitch task (Whitfield, 1980).

We used neurometric techniques to compare the discrimination performance afforded by single units and ensembles of simultaneously recorded units with that of trained ferrets in a periodicity discrimination task. The  $f_0$  spike rate functions of a few individual neurons provided sufficient information to account for the animals’ discrimination judgments but only over a very limited range of periodicity values. However, the response patterns of small ensembles of simultaneously recorded neurons typically discriminated stimulus periodicity with greater accuracy and over a wider range of values than those of individual neurons. Furthermore, ensemble codes for the vowel  $f_0$  were robust to changes in sound intensity and could be read out in at least four alternative ways, based on either spike count or relative latencies.

To decode ensemble responses, we implemented a pattern classification algorithm that took, as its input, one of four different spike summary statistics: spike count, relative spike latency, a reduced “binary” response code, or a first-spike-order code. Similar Euclidean-distance-based metrics have been used previously to decode activity patterns of individual neurons and ensembles recorded either nonsimultaneously (Schnupp et al., 2006; Engineer et al., 2008) or simultaneously (Walker et al., 2008). We also tried other classification algorithms, most of which did not substantially alter the performance of the ensembles. Other studies have emphasized the potential role of spike latency codes (deCharms and Merzenich, 1996; Jenison, 2000; Thorpe et al., 2001; Van Rullen and Thorpe, 2002; Johansson and Birznieks,

2004; Nelken et al., 2005; Gollisch and Meister, 2008). However, attempts to examine ensemble coding have either focused on pairs of simultaneously recorded units or simulated ensemble activity from single-unit recordings. Here we recorded simultaneously from sizeable ensembles of neurons, allowing us to test different population coding strategies in the presence of real neural correlations.

The neurometric performance of ensembles, whether based on spike counts or latencies, was often within the normal range of ferret psychometric performance in our pitch discrimination task. However, spike count and latency codes were not entirely redundant, because a joint count-and-latency code performed ~10% better than the codes based on either spike count or latency alone. Previous studies of auditory cortex have shown that optimal decoding might be achieved with a combined spike count and mean response latency code (Nelken et al., 2005), by patterns of spikes across an ensemble (Furukawa et al., 2000), or by combining information about the precise temporal pattern of spiking relative to the phase of ongoing oscillations (Kayser et al., 2009). We believe this to be the first time, however, that a relative spike latency code has been compared with psychophysical performance in the same species.

The finding that combining spike count and relative latency information provided the most accurate way of inferring stimulus periodicity from neuron ensembles is consistent with recent evidence for “multiplexing” of neural codes in monkey auditory cortex (Kayser et al., 2009). These authors found that the combined information available from pairs of well separated neurons was complementary and essentially independent. Although we did not investigate low-frequency oscillations, it would be interesting to examine whether the relative timing of action potentials and the local field potential might provide additional information. One potential problem for decoding relative latencies across neural ensembles is the possibility that spike times might be erroneously measured relative to spontaneous, rather than stimulus-evoked, events. Information about spike timing relative to the power and phase of the local field potential might allow such occasions to be disambiguated.

Relative timing codes could provide a biologically plausible mechanism for reading out ensemble activity, enabling rapid response times (Van Rullen and Thorpe, 2002; Johansson and Birznieks, 2004; Masquelier et al., 2008), because there is no requirement to integrate or count spikes over a particular time window. However, we found that spike count and latency-based ensemble decoding allowed equally rapid and accurate periodicity judgments, reaching optimal performance within 75 ms of stimulus onset. These population response integration times are also shorter than the optimal response window required for single units, which was found to be 150 ms. Combining information across neurons therefore allowed information to be decoded more rapidly. The finding that the combined count–latency code only exceeds either code alone from 50 to 75 ms suggests that the information in each evolves over a similar timescale. Relative latency and rank-order coding could both be achieved physiologically via shunting inhibition (Thorpe et al., 2001). However, it is interesting to note that our spike-order code performed more poorly than the relative latency code, suggesting that the actual spike times are more informative than simply the order in which the neurons fired.

In both human listeners (Moore, 2004) and our trained ferrets (Walker et al., 2009), pitch discrimination judgments are unaffected by large trial-to-trial fluctuations in sound intensity. Any neural correlate of this discrimination task should therefore show

a similar invariance across sound levels. However, both firing rates and spike latencies tend to change systematically with sound level (Heil, 1997a,b), implying that neural coding of  $f_0$  might not be level invariant. Nevertheless, we found that ensemble responses to sounds that vary in both  $f_0$  and intensity can be accurately decoded using either spike counts or relative latencies. The presence of approximately equal numbers of neurons that monotonically increase or decrease their firing rates with increasing  $f_0$  could help to generate an ensemble periodicity representation that is invariant across sound level. Within a neural population, a change in sound intensity will likely alter the absolute number of spikes but not the pattern of activation across balanced populations of high-pass and low-pass neurons, allowing information about stimulus  $f_0$  to be disambiguated from changes in sound level.

Our finding that correlated noise can impair performance when decoded with a spike count code, even in small ensembles of simultaneously recorded neurons, supports previous work demonstrating that weak correlation between neurons can limit the coding capacity of a neuronal pool (Zohary et al., 1994). In contrast, the predominant effect of decorrelating responses when decoded with a spike latency code was for performance to decline. This suggests that trial-to-trial correlations enhance this relative latency code. It is possible that selecting the optimal combination or differential weighting of neurons within the pool might improve performance even further, but such analyses were not attempted here.

Based on the location of neurons that were tuned to a particular periodicity even when the  $f_0$  was omitted from the spectrum of the sound, Bendor and Wang (2005) described the presence of a pitch center in marmoset auditory cortex. We might therefore expect the optimal ensemble size for representing stimulus periodicity to vary across different fields in ferret auditory cortex. However, we found that the general sensitivity to  $f_0$  observed in all five fields investigated is sufficient to support periodicity discrimination behavior. Stimulus periodicity is one of the key acoustic determinants of the perceptual quality of pitch. Many spectrally different sounds can elicit the same pitch percept, and a neural substrate for pitch perception should show invariance not just to sound level but also across a range of spectrally different stimuli. Showing that neurons modulate their firing rates in response to changes in periodicity and that these changes correlate with psychophysical performance is not, by itself, sufficient to demonstrate pitch selectivity. Nevertheless, our data suggest that sensitivity to one of the key determinates of pitch is distributed throughout auditory cortex.

Simultaneous recordings of neural responses and behavioral measurements would show more directly that the neural coding strategies explored here are actually used by the brain as it solves perceptual tasks. Such studies have been performed successfully in other sensory modalities. For example, Luna et al. (2005) found that vibrotactile stimuli could be discriminated on the basis of five different neural codes but showed, using trial-to-trial correlations between neuronal responses and behavior, that only a spike count code could account for the animals' discrimination ability. Although there is evidence that behavioral context does not alter the tuning of auditory cortical neurons, suggesting that responses observed in passive listening conditions provide a valid measure of the representation of sound properties (Scott et al., 2007), other studies have shown that engaging in a task can suppress auditory responses (Otazu et al., 2009) and alter receptive field structure (Fritz et al., 2003). Examining neural coding and behavioral discrimination simultaneously in the same animal is

therefore likely to be essential to our ultimate goal of identifying the neural basis of perceptual judgments.

## References

- Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436:1161–1165.
- Bizley JK, Nodal FR, Nelken I, King AJ (2005) Functional organization of ferret auditory cortex. *Cereb Cortex* 15:1637–1653.
- Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW (2009) Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J Neurosci* 29:2064–2075.
- Brugge JF, Reale RA, Jenison RL, Schnupp J (2001) Auditory cortical spatial receptive fields. *Audiol Neurootol* 6:173–177.
- Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76:1698–1716.
- Cedolin L, Delgutte B (2005) Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *J Neurophysiol* 94:347–362.
- Chase SM, Young ED (2007) First-spike latency information in single neurons increases when referenced to population onset. *Proc Natl Acad Sci U S A* 104:5175–5180.
- Dahmen JC, Hartley DE, King AJ (2008) Stimulus-timing-dependent plasticity of cortical frequency representation. *J Neurosci* 28:13629–13639.
- De Cheveigne A (2005) Pitch perception monkeys. In: *Pitch: neural encoding and perception* (Plack CJ, Oxenham AJ, Fay RR, Popper AN, eds). New York: Springer.
- deCharms RC, Merzenich MM (1996) Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381:610–613.
- Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP (2008) Cortical activity patterns predict speech discrimination ability. *Nat Neurosci* 11:603–608.
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223.
- Furukawa S, Xu L, Middlebrooks JC (2000) Coding of sound-source location by ensembles of cortical neurons. *J Neurosci* 20:1216–1228.
- Gautrais J, Thorpe S (1998) Rate coding versus temporal order coding: a theoretical approach. *Biosystems* 48:57–65.
- Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. *Science* 319:1108–1111.
- Heil P (1997a) Auditory cortical onset responses revisited. I. First-spike timing. *J Neurophysiol* 77:2616–2641.
- Heil P (1997b) Auditory cortical onset responses revisited. II. Response strength. *J Neurophysiol* 77:2642–2660.
- Javel E (1980) Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones. *J Acoust Soc Am* 68:133–146.
- Jenison RL (2000) Correlated cortical populations can enhance sound localization performance. *J Acoust Soc Am* 107:414–421.
- Johansson RS, Birznieks I (2004) First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nat Neurosci* 7:170–177.
- Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61:597–608.
- Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60:1799–1822.
- Luna R, Hernández A, Brody CD, Romo R (2005) Neural codes for perceptual discrimination in primary somatosensory cortex. *Nat Neurosci* 8:1210–1219.
- Masquelier T, Guyonneau R, Thorpe SJ (2008) Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE* 3:e1377.
- McDermott JH, Oxenham AJ (2008) Music perception, pitch, and the auditory system. *Curr Opin Neurobiol* 18:452–463.
- Moore BCJ (2004) *An introduction to the psychology of hearing*, Ed 5. San Diego: Academic.
- Nelken I, Chechik G, Msrac-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.
- Nishikawa J, Okada M, Okanoya K (2008) Population coding of song element sequence in the Bengalese finch HVC. *Eur J Neurosci* 27:3273–3283.
- Otazu GH, Tai LH, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* 12:646–654.
- Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21:227–277.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* 24:6810–6815.
- Rees A, Palmer AR (1989) Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise. *J Acoust Soc Am* 85:1978–1994.
- Sayles M, Winter IM (2008) Ambiguous pitch and the temporal representation of inharmonic iterated rippled noise in the ventral cochlear nucleus. *J Neurosci* 28:11925–11938.
- Schnupp JW, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26:4785–4795.
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60:1823–1840.
- Scott BH, Malone BJ, Semple MN (2007) Effect of behavioral context on representation of a spatial cue in core auditory cortex of awake macaques. *J Neurosci* 27:6489–6499.
- Thorpe S, Delorme A, Van Rullen R (2001) Spike-based strategies for rapid processing. *Neural Netw* 14:715–725.
- Van Rullen R, Thorpe SJ (2001) Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput* 13:1255–1283.
- Van Rullen R, Thorpe SJ (2002) Surfing a spike wave down the ventral stream. *Vision Res* 42:2593–2615.
- Walker KM, Ahmed B, Schnupp JW (2008) Linking cortical spike pattern codes to auditory perception. *J Cogn Neurosci* 20:135–152.
- Walker KM, Schnupp JW, Hart-Schnupp SM, King AJ, Bizley JK (2009) Discrimination by ferrets of the direction of pitch changes in simple and complex sounds. *J Acoust Soc Am* 126:1321–1335.
- Wang L, Narayan R, Graña G, Shamir M, Sen K (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27:582–589.
- Wang X, Lu T, Bendor D, Bartlett E (2008) Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience* 19:484–494.
- Whitfield IC (1980) Auditory cortex and the pitch of complex tones. *J Acoust Soc Am* 67:644–647.
- Winter IM, Palmer AR, Meddis R (1993) The response of guinea pig auditory-nerve fibers with high spontaneous discharge rates to increments in intensity. *Brain Res* 618:167–170.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.