Behavioral/Systems/Cognitive

Temporal Codes for Amplitude Contrast in Auditory Cortex

Brian J. Malone, Brian H. Scott, and Malcolm N. Semple
Center for Neural Science at New York University, New York, New York 10003

The encoding of sound level is fundamental to auditory signal processing, and the temporal information present in amplitude modulation is crucial to the complex signals used for communication sounds, including human speech. The modulation transfer function, which measures the minimum detectable modulation depth across modulation frequency, has been shown to predict speech intelligibility performance in a range of adverse listening conditions and hearing impairments, and even for users of cochlear implants. We presented sinusoidal amplitude modulation (SAM) tones of varying modulation depths to awake macaque monkeys while measuring the responses of neurons in the auditory core. Using spike train classification methods, we found that thresholds for modulation depth detection and discrimination in the most sensitive units are comparable to psychophysical thresholds when precise temporal discharge patterns rather than average firing rates are considered. Moreover, spike timing information was also superior to average rate information when discriminating static pure tones varying in level but with similar envelopes. The limited utility of average firing rate information in many units also limited the utility of standard measures of sound level tuning, such as the rate level function (RLF), in predicting cortical responses to dynamic signals like SAM. Response modulation typically exceeded that predicted by the slope of the RLF by large factors. The decoupling of the cortical encoding of SAM and static tones indicates that enhancing the representation of acoustic contrast is a cardinal feature of the ascending auditory pathway.

Introduction

The information present in acoustic signals exists at multiple timescales, which has encouraged the distinction between the fine structure, the rapid changes of the pressure waveform that determine the spectral content of the sound, and the envelope, the contour describing changes in the overall amplitude of the pressure waveform (Rosen, 1992; Smith et al., 2002; Joris et al., 2004). Low-frequency envelope information (<20 Hz) is crucial for understanding human speech (Houtgast and Steeneken, 1973, 1985; Drullman et al., 1994b; Drullman, 1995; Fu and Shannon, 2000; Elliott and Theunissen, 2009) and even rhesus macaque vocalizations (Cohen et al., 2007). Sinusoidal amplitude modulation (SAM) is a useful and convenient tool for studying the neural representation of sound envelopes because SAM envelopes consist of a single component, which makes them elementary in the modulation domain, just as pure tones are elementary in the frequency domain.

Nevertheless, the modulation frequency which determines the periodicity of SAM is but one of four parameters required to specify a SAM signal, and the others—carrier frequency, carrier level, and modulation depth—are also robustly encoded in primary auditory cortex (AI) by precise temporal spiking patterns

(Malone et al., 2007). Of these, modulation depth is of particular interest because it determines the "contrast" of SAM signals. The temporal modulation transfer function (MTF), which measures the lowest detectable depth across modulation frequency, is a standard psychophysical measure in normal hearing (Kay, 1982), and more recently, in hearing mediated by cochlear implants (Busby et al., 1993; Cazals et al., 1994; Galvin and Fu, 2005). Importantly, modulation thresholds for users of cochlear implants are strongly correlated with phoneme recognition (Fu, 2002).

Despite the fundamental role of temporal modulations in both acoustic and electric hearing, very little is known about cortical sensitivity to shallow modulation depths (Middlebrooks, 2008). Typically, neurophysiological studies of SAM have used depths of 100%, whereas modulation detection thresholds for normal listeners have been reported as low as 2% (Zwicker, 1952, as shown in Kohlrausch et al., 2000) (Fig. 1). Prior studies that have measured responses to multiple modulation depths in awake primate cortex have focused on whether the cross-correlation between the peristimulus time histogram (PSTH) and the SAM envelope increased monotonically with modulation depth (Bieser and Müller-Preuss, 1996), or whether the best modulation frequencies (BMFs) for the MTF were invariant to changes in modulation depth (Liang et al., 2002), rather than on the limits of modulation sensitivity. Neither study used depths of <20%. Malone et al. (2007) demonstrated that the spike timing patterns produced by many cortical neurons could be used to discriminate modulation depths but did not report modulation detection or discrimination thresholds.

In the current study, we describe how single units recorded in the auditory cortex of awake macaques encode SAM stimuli of varying depths and pure tone signals of varying sound level. We
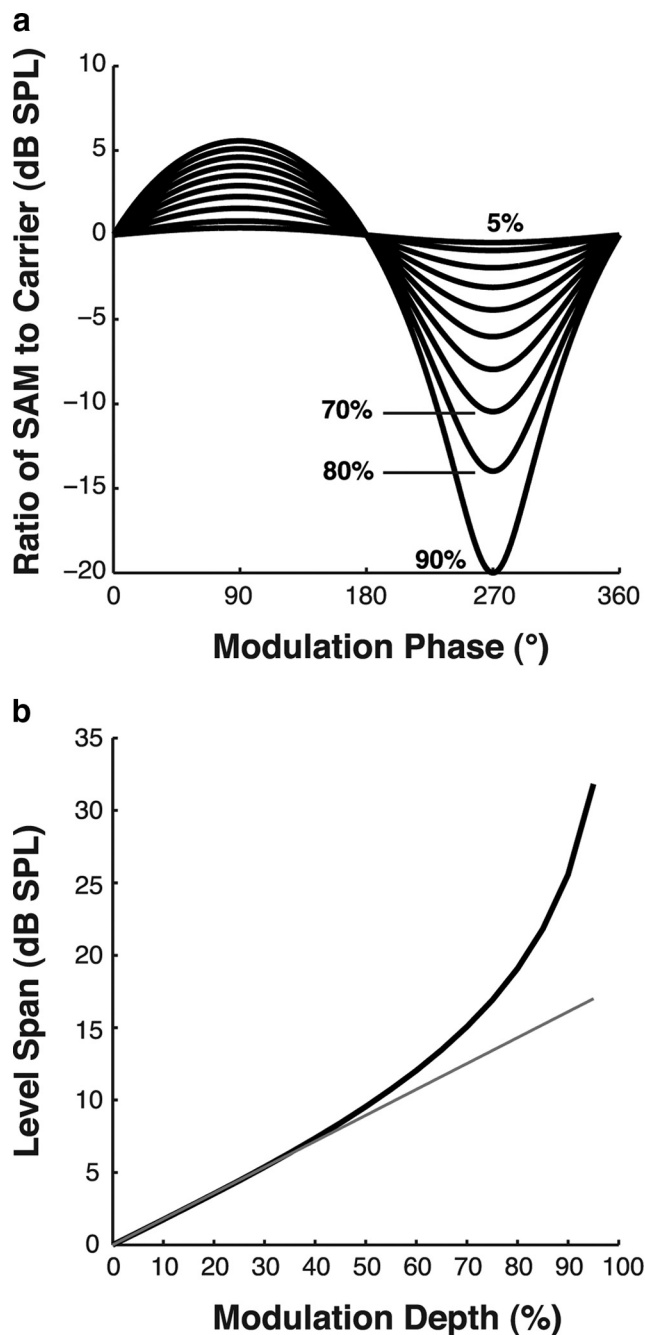
Figure 1. On a decibel axis, larger modulation depths produce larger decreases than increases in SPL and span much greater ranges. ***a***, Changes in the instantaneous SPL, in decibels, of the SAM stimulus relative to an unmodulated stimulus at the same carrier level. The curves represent modulation depths ranging from 10 to 90% in 10% steps, as well as the lowest depth tested, 5%. The instantaneous SPL achieves its minimum at 270° because all modulations were presented in sine phase. ***b***, Span, in dB SPL, of the modulation as a function of modulation depth. The level span is defined as the difference between the maximum and minimum instantaneous SPL (in decibels) within each modulation cycle. The gray line indicates the result if the level span increased linearly with modulation depth, which it nearly does up to values of 40%.

found that despite the dramatic reduction in cortical sensitivity to high-frequency SAM, the representation of low-frequency SAM is essentially undiminished relative to the auditory periphery and comparable to psychophysical performance in the most sensitive cells. Although the representation of static and dynamic changes in level is at least partially decoupled in the cortex, spike pattern discrimination of both SAM depth and tone level is robust.

Materials and Methods

Subjects, surgical preparation, and physiological recording

Two adult male monkeys (*Macaca mulatta*, designated X and Z) participated in these experiments. The general methods of animal training, stimulus delivery, and physiological recording have been described previously (Malone et al., 2002, 2007; Scott et al., 2007, 2009). All procedures were in accordance with the Society for Neuroscience guiding principles on the care and use of animals and approved by the Institutional Animal Care and Use Committee of New York University. All the data described in this report were obtained while the animals were sitting passively (Scott et al., 2007) with their heads fixed in a custom chair (Crist) within a double-walled anechoic chamber (Industrial Acoustics) while being monitored by video.

Physiological recordings were performed in both the left and right hemispheres of each animal by using a stainless steel chamber implanted with standard sterile surgical techniques. Localization of the neurons comprising the data for this report were confirmed for animal X through histology and postmortem magnetic resonance imaging (MRI). Physiological criteria, referenced to the anatomical data available for animal X, were used to define the locations of neurons encountered in animal Z (Scott et al., 2009).

Single-unit extracellular recordings were obtained by advancing tungsten microelectrodes (10–12 MΩ; FHC) with a stepping microdrive, such that electrode tracks were roughly parallel to the stereotaxic vertical plane (Pfingst and O'Connor, 1980). Electrical signals were amplified (variable gain), filtered (typically from 0.25 to 10 kHz), and passed to oscilloscopes, an audio speaker, and an event timer (MALab, Kaiser Instruments). Search stimuli including tones, bandpass noise, and sinusoidally amplitude- and frequency-modulated tones were used to identify single units, which were discriminated with multiple adjustable voltage/time windows. Action potentials and stimulus synchronization events were logged with a resolution of 1 μs by custom hardware (MALab, Kaiser Instruments) and stored by the host computer (Macintosh) for analysis and display.

Stimulus generation and analysis

Stimuli were generated digitally (MALab software and hardware) and presented in the closed field via electrostatic speakers (STAX Lambda) coupled to ear inserts (Custom Sound Systems) positioned within the ear canal. Phase and level at each ear were calibrated across frequency at the start of each session using a ½-inch probe microphone (Brüel and Kjær model 4133). All stimuli described in this report were gated on and off by a cosine-squared ramp (10 ms). Frequency tuning functions and rate level functions (RLFs) were obtained by blocked, repeated ($n = 10$) presentations of tone pips of 100 ms duration presented once a second. The RLFs were calculated using a 150 ms window beginning at tone onset. The minimum latency was calculated by taking the average of the first spike latency in each trial. Only values of >5 ms and <150 ms were included, and a valid latency estimate was required for a minimum of five trials to compute the average and SD.

Some aspects of the data sample described in this report have been published by Malone et al., 2007 (their Figs. 9–11), as have many of the details of SAM stimuli. Briefly, SAM stimuli consisted of a sinusoidal carrier tone (f_c) modulated sinusoidally by a second tone (f_m) such that $s(t) = A[1 + m\sin(2f_mt)]\sin(2f_ct)$. If the carrier frequency exceeds the modulating frequency by a large margin (i.e., $f_c \gg f_m$), the bracketed term in the equation above determines the time-varying amplitude of the stimulus at temporal resolutions well below the carrier frequency. As is evident from the equation, four parameters define a SAM stimulus: the carrier frequency (f_c), modulation frequency (f_m), carrier level (A), and modulation depth (m). By varying m while holding the other parameters constant, it is possible to characterize the neural modulation depth function (MDF).

The data in this report represent all instances in which an MDF was obtained during a detailed physiological survey of auditory cortex. SAM stimuli were presented in two consecutive trials of 10 s duration, separated by an interstimulus interval of 2 s. The long duration of the stimuli and interstimulus intervals ensured that onset and offset responses did not materially alter our results (Malone et al., 2007). In most cases, (121/
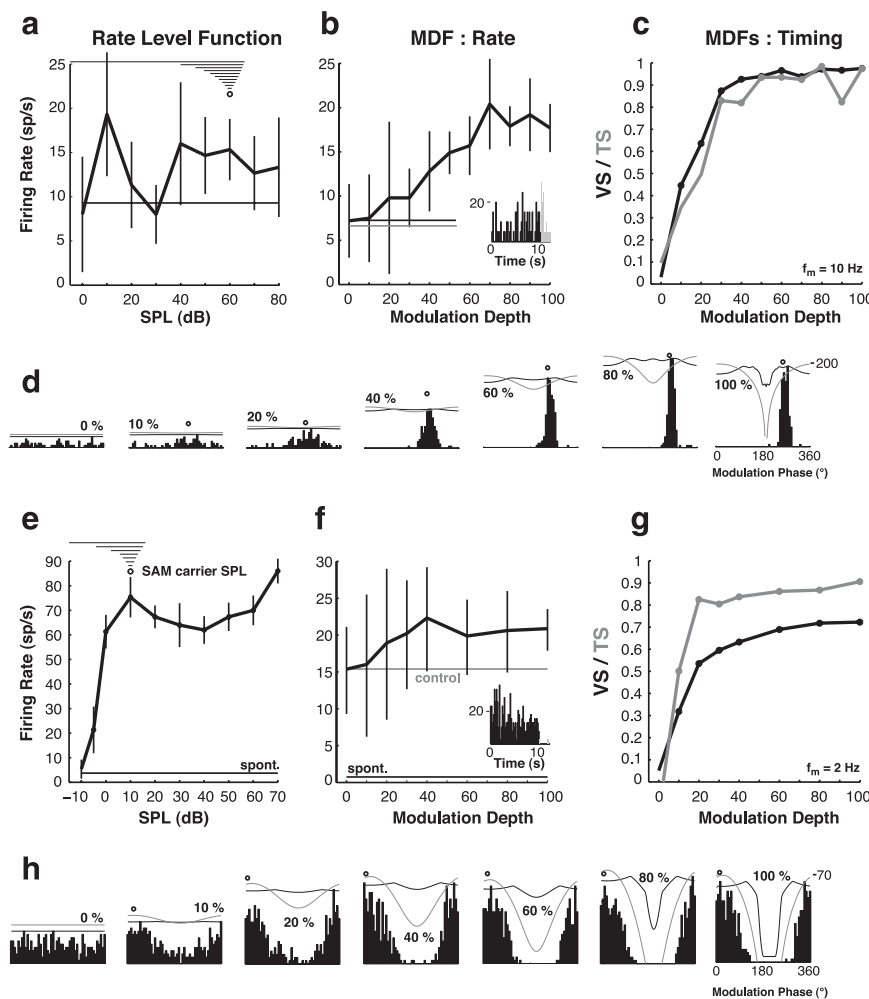
Figure 2. Cortical encoding of sound level varies substantially from neuron to neuron. ***a***–***h***, Shown here are the responses from two cells in auditory cortex (***a***–***d*** and ***e***–***h***). ***a***, ***d***, Rate level functions obtained for each of these cells. On all panels, vertical lines indicate ±2 SEM of measurement. Circles on a point of the RLF indicate that the firing rate was elevated relative to the spontaneous rate (Wilcoxon rank-sum; $p < 0.001$). The small circular icon indicates the carrier level used to obtain the rate modulation depth functions depicted in ***b*** and ***f***. The series of small horizontal lines above the carrier level icons (***a***, ***e***) illustrates the level span of the SAM stimuli comprising the MDF. The larger horizontal line indicates the spontaneous firing rate averaged over the interstimulus intervals. This convention is maintained in ***b*** and ***f***. The additional horizontal gray line indicates the response to the unmodulated control stimulus, which is also shown in PSTH form as an inset (gray bars in the insets represent spikes during the interstimulus interval). ***c***, ***g***, Vector strength (black) and trial similarity (gray) obtained across modulation depth. Filled circles on the VS and TS curves indicate significance at the $p < 0.001$ level (VS: Rayleigh test >13.816; TS: bootstrap; see Materials and Methods). The inset text indicates the modulation frequency. ***d***, ***h***, MPHs elicited by the modulation depths indicated on each panel. All MPHs have been rotated (90°) so that the instantaneous minimum SPL of the SAM stimulus is centered in the panel. The small circles indicate the mean phase of the response. The gray curves above each panel indicate the change in instantaneous SPL (Fig. 1a). The black curves indicate the modulation in firing rate predicted by the instantaneous SPL and the RLF. These curves are intended to indicate relative modulation, rather than absolute spike rates, and are vertically placed so their maximum aligns with the maximum spike counts in a single bin of the MPH. These curves have been rotated by the average minimum latency for each neuron.

137), the MDF or MTF contained a set of trials at a modulation depth of 0% (i.e., $m = 0$) to serve as an unmodulated control. Modulation depth was typically varied from 0 to 100% in 10 or 20% steps. In a few neurons where initial testing indicated particular sensitivity to low modulation depths ($n = 19$), a modulation depth of 5% was used. Responses to modulation depths of 25 or 75% were rounded up to 30 and 80% for population summaries. All SAM stimuli were presented in sine phase, as shown in Figure 1a. For graphical convenience, however, all modulation period histograms (MPHs) have been rotated by 90°, such that the amplitude minimum occurs in the center of the graph (Fig. 2; also Fig. 5 below).

The "instantaneous" sound pressure level (SPL) of the SAM stimulus was calculated by taking of the logarithm of the envelope, in decibels:

$20 \cdot \log[1 + m\sin(2f_\mathrm{m}t)]$. The results of this calculation for a range of modulation depths are shown in Fig. 1a. In the case of the largest depth indicated, 90%, the amplitude of the modulated stimulus varies from $+5.58$ to -20 dB relative to the carrier level. In contrast, a modulation depth of 5% represents a change from $+0.42$ to -0.45 dB. To help contextualize the magnitude of these changes in stimulus amplitude with respect to the RLF in dB SPL (re: 20 μPa), Figure 1b illustrates the "level span" of SAM stimuli as a function of modulation depth. We define the decibel span as the difference between the instantaneous maximum and minimum of the SAM stimulus relative to the carrier within each modulation cycle. For modulation depths of <40%, the decibel span increases quasilinearly, but at higher depths, the decreases in SPL each cycle are magnified. At 100% modulation, the dB increase is 6, while the decrease is briefly $-\infty$. As a result, the rapid fall and rise of the envelope occurring during a narrow portion of the modulation cycle can dominate the responses of some cortical neurons (Malone et al., 2007). Here, however, we are chiefly concerned with the neural representation of the small envelope changes introduced by low modulation depths, which sample a relatively small and symmetric range of amplitudes. All statistical tests were two tailed unless noted otherwise.

Modulation analysis
All data analysis was performed using MATLab (MathWorks). A detailed explanation of the analysis of responses to SAM signals has been provided previously (Malone et al., 2007). Average firing rate was calculated by binning responses to the 10 s duration stimuli into 1 s epochs and averaging over trials (two trials). Spontaneous firing rates were estimated by calculating the spike rate during each 1 s epoch of the 2 s interstimulus interval separating each trial. Significant differences in firing rate were determined by comparing the distributions of spike rates using a Wilcoxon rank-sum test ($p < 0.01$). Response class was determined by comparing the response to an unmodulated control tone (10 s) to the spontaneous firing rate. Neurons were classified as transient if the difference in firing rates was not significant, driven if there was a significant increase relative to the spontaneous rate, and suppressed if there was a significant decrease. The term "transient" was chosen because the responses in that class typically included a response at tone onset that did not endure throughout the duration of the control tone.

The MPH was formed by folding the response to SAM on the modulation period, resulting in a single-cycle histogram that depicts the change in the response as a function of phase. The MPHs contained 52 bins and were converted into spike rates by dividing the spike count in each bin by the duration of the bin in seconds and the number of modulation cycles used to obtain the MPH.

Two timing indices were used to quantify the relationship between the SAM stimulus envelope and the neural response. Vector strength (VS; Goldberg and Brown, 1969) was used to measure the degree to which the neural response was concentrated at a particular phase of the modulation cycle, such that $VS = (1/n) \cdot \sum_{i=1}^{n} ((\cos(2\pi \cdot f_\mathrm{m} \cdot t_i)^2 + \sin(2\pi \cdot f_\mathrm{m} \cdot t_i)^2)^{0.5}$,

where t_i is the time of occurrence of the ith spike, n is the total number of spikes, and f_m is the modulation frequency. A neuron was considered to be synchronized to the modulation envelope if the Rayleigh statistic, $2 \cdot VS^2 \cdot n$, exceeded 13.816 ($p < 0.001$) (Mardia and Jupp, 2000). If all spikes occur at the same modulation phase, then $VS = 1$. If all spikes are evenly distributed in the modulation cycle, $VS = 0$. To complement the VS metric, we also calculated an index of trial similarity (TS; Malone et al., 2007), which measures the reproducibility of the shape of the MPH rather than its dispersion. TS was calculated by taking the correlation between the MPHs obtained from separate, consecutively presented trials (10 s). In cases where only a single trial was available, the data were split, and MPHs based on the halves of the trial were correlated. Significance was assessed by computing a distribution of correlation coefficients for pairs of random spike trains over a range of bin widths and total spike counts. Generally, for a 52-bin MPH, TS values of 0.4 and 0.6 correspond conservatively to p values of 0.001 and 0.0001, respectively. When necessary, significance was determined by comparing the TS result to the distribution obtained at an equivalent firing rate. Negative values of TS were set to zero because we never obtained a significant negative result by the criterion above.

Modulation gain was calculated as $20 \cdot \log_{10} (2 \cdot VS/m)$, where m indicates the modulation depth, which varied from 0 to 1. We will describe modulation depth in terms of percentage (0–100%) in what follows. As an alternative to modulation gain, which is based on spike timing, we also computed the rate contrast ratio, which is based on changes in firing rate within the modulation cycle. This metric represents the ratio between the modulation present in the MPH and the modulation predicted by the neuron's static tuning for sound amplitude measured by the RLF. To minimize the effects of bursts, each MPH was converted into a bin by bin average firing rate, then smoothed by taking the three-bin average. The numerator of the rate contrast ratio was the difference between the maximum and minimum firing rates of the smoothed MPH. The denominator is based on the firing rate difference predicted by using the RLF as a lookup table for the distribution of "instantaneous" SPLs. To do so, the SPL is determined for each phase of the modulation cycle by adding the carrier level to the relative level of the SAM stimulus (Fig. 1a) to determine the instantaneous SPL and then finding the firing rate associated with that SPL from a linearly interpolated version of the RLF. For plotting purposes (Fig. 2), the resulting curve was smoothed with a Gaussian window (MATLAB; gausswin; $\alpha = 1/4$). The maximum firing rate difference within the actual MPH was divided by maximum firing rate difference from the RLF-derived, predicted MPH to obtain the rate contrast ratio.

We predicted the shape of the MPHs using a procedure similar to that described above. We calculated the group delay for each unit, defined as the slope of the best linear fit to the (phase-wrapped) mean phase of each MPH from 1 to 20 Hz, using the data from the MTF. We then rotated the MPH to compensate for the group delay. We compared the MPH to the stimulus envelope in decibels (Fig. 1) and to the predictions of the RLF-lookup model described above by computing the cross-correlations. The symmetry of the envelope of SAM signals implies that a neuron whose firing rate output bears a fixed relationship to the instantaneous SPL should produce a symmetrical MPH (a SAM signal presented in sine phase is symmetrical about 90° and 270°, as shown in Fig. 1a). After compensating for the group delay, each MPH from the MDF was rotated by 90° to center the stimulus minimum and halved. The first half of the MPH ($h1$) was then compared with the phase-reversed version of the second half ($h2$). A symmetry index (SI) was calculated by treating the spike count in each bin as a vector, and computing $1 - \| h_1 - h_2 \|/(\| h_1 \| + \| h_2 \|)$, where $\| \times \|$ represents the vector norm. The SI varies from 0 to 1, where 1 indicates perfect symmetry. We complimented this analysis by computing the cross-correlation between h_1 and h_2, the symmetry correlation, which compares the MPH profiles but is insensitive to differences in magnitude.

Spike train classification

To quantify the sensitivity of modulation depth coding, we used PSTH-based pattern classifiers to determine when the response to a SAM stimulus differed from the response to an unmodulated control tone

(modulation detection) or when the responses to two SAM stimuli could be distinguished (modulation depth discrimination). These methods have been described in detail previously (Foffani and Moxon, 2004) and have been applied to this dataset in prior publications (Malone et al., 2007). Briefly, responses to each stimulus in a set (e.g., an MDF) were divided into 1 s analysis epochs, averaged to form a "template" response for each stimulus, and binned to form a bin-dimensional vector representing the response. Each 1 s epoch of the data was binned similarly, and compared with the templates by computing the Euclidean distance between the test and the template vectors, and the match that minimized that distance was estimated to be the stimulus that produced the response. When the test and template were drawn from the same stimulus, the test was excluded from the average that produced the template (complete cross-validation). The binning was varied (2, 4, 8, 10, 20, 40, and 1000 ms) to encompass a range of temporal resolutions. When reporting classifier performance, we chose the bin width for each cell that resulted in the best classifier performance. When only a single bin (1000 ms) is used, the classification is based entirely on the spike rate averaged over the duration of the test epoch. Conversely, it is possible to remove average firing rate information and retain the relative distribution of spikes within the tests and templates by normalizing them by their respective vector norms. We refer to this as the "phase-only" classifier.

Classifier performance was evaluated by computing the percentage correct by summing along the diagonal of a confusion matrix whose columns indicate the actual stimulus, and whose rows indicate the estimate of stimulus identity produced by the classifier. Correct estimates fall along the diagonal, and the percentage correct can be computed by dividing the sum of the diagonal entries by the total number of estimates. Because modulation depth varies monotonically, rather than categorically, a classifier estimate of 30% for SAM at a depth of 40% is superior to an estimate of 20%, or 60%. With this in mind, each estimate was assigned an error cost proportional to the ordinal difference between the actual stimulus and the estimate. For example, if the depth was 40%, the error cost of an estimate of 40% would be 0, 30% would be 1, and 20% would be 2. Significance was assessed by simulating confusion matrices with random estimate assignments and generating distributions of both percentage correct and error cost data. Actual classifier performance was then compared with the relevant distribution obtained from a confusion matrix of equivalent size. When comparing results that involved confusion matrices of differing sizes, classifier performance was standardized as z-scores relative to the distributions obtained by bootstrapping. We also varied the duration of the analysis epoch to determine the accuracy of the classification process as a function of time. In doing so, however, we observed that the classifier could sometimes match a particular stimulus with many of the responses, producing a relatively low error cost when that stimulus fell in a row toward the middle of the confusion matrix. If the total number of estimates in a row of the confusion matrix exceeded 2.5 times of the expected value for a row (e.g., 10% of the total for a 10-by-10 matrix), discrimination performance was not considered significantly different from chance. When estimating the minimum time for successful classification, we identified the first significant ($p < 0.001$) interval where the average of all subsequent intervals was also significant by the same criterion.

We derived estimates of modulation detection and discrimination by assessing classifier performance for sets of pairwise comparisons of modulation depths by using 2-by-2 confusion matrices. For modulation detection, PSTH classification was applied to all pairs of modulation depths that included the unmodulated control. For modulation discrimination, all pairs of modulation depths were compared. Having noted that classification performance could vary with small changes in the number of bins because a prominent feature of the response was split across them, we took the average of performance obtained with five different bin counts centered on the bin width that provided the desired temporal resolution (e.g., for 20 ms resolution, we used 48, 49, 50, 51, and 52 bins). We verified that this procedure produced a more conservative estimate of performance by evaluating the distributions obtained when classifying random spike trains using both methods. When classifying responses in pairs, we also compensated for the bias introduced by subtraction of the test from one of the two templates, which introduces a slight difference in

the amount of data averaging (i.e., dividing by 19 instead of 20), by subtracting a second "test" epoch from the other template. Classification of the tone pip stimuli used to construct the RLFs was performed similarly, but five-bin averaging was unnecessary because their level dynamics were limited to onset and offset.

Analysis epoch was also varied to evaluate how classifier performance depended on the amount of data available to the classifiers. For the MDF data, performance was assessed by varying the analysis epoch from 50 to 1000 ms in 50 ms steps, using a temporal resolution of 10 ms. Cells tested at modulation frequencies of >50 Hz were excluded from this analysis. Analogously, we varied the analysis epoch used to classify the responses to 100 ms tone pips in 50 ms steps from 50 to 300 ms, and 400 ms. Because the optimal temporal resolution for this analysis was not known from prior work (Malone et al., 2007), we performed the classification at a range of temporal resolutions (1, 2.5, 5, 10, 25, and 50 ms), in addition to the full duration of the analysis epoch, which was used to compute the performance of the rate-only classifier.

Results

Summary of the data sample

The data in this report are derived from 137 single neurons recorded in the auditory cortex in two animals (X:102; Z:35), and four hemispheres (X, left: 34; X right, 68; Z, left: 6; Z, right: 29). Data from these cells has also been described in a prior paper focused on amplitude modulation coding (Malone et al., 2007). This dataset includes all MDFs obtained during a physiological survey of auditory cortex. It is likely that this sample is biased toward robust responses to modulated stimuli, since cells that responded poorly to the fully modulated ($m = 100\%$) SAM stimuli used to derive the MTF were unlikely to be further probed at low modulation depths. The majority of neurons were recorded from AI ($n = 79$, 58%; see Materials and Methods), though a sizeable fraction were recorded in the rostral field (R) ($n = 47$, 34%). The remainder were recorded in the caudomedial ($n = 5$), lateral ($n = 4$), and medial ($n = 2$) belt.

We attempted to obtain the MDF at the modulation frequency, carrier frequency, and carrier level that produced the most robust response modulation. When the frequency tuning function was available, the carrier frequency was chosen as the maximum. The choice of carrier level was more complicated because the maximum of the RLF obtained with tone pips (100 ms duration) did not necessarily produce the most robust modulation of the response to SAM tones. The carrier level was determined audiovisually by attempting to maximize the rate-synchrony product (effectively, the Rayleigh statistic) for fully modulated stimuli. As a result, the chosen carrier levels cannot necessarily be identified with a particular feature of the tonal RLF, such as the peak. Nevertheless, in 107 of 132 neurons where the RLF was available, the carrier level was within ±20 dB SPL of the best level of the RLF, and identical with it in 47 cases. The largest difference between the best level and carrier level was 60 dB. Such large differences are explained by the fact that some cells were most effectively modulated when the carrier level exceeded their best level, so that they fired robustly during the brief reduction in amplitude within each modulation cycle, or during the subsequent rapid increases from very low to moderate levels. Carrier level varied from 0 to 80 dB, with a mode of 60 dB (68/137). The other most commonly used SPLs were 40 ($n = 14$), 30 (12), and 50 (8) dB. SAM signals were nearly always presented binaurally (131/137), but monaural stimulation to the preferred ear was used when necessary to obtain a response.

The choice of modulation frequency proceeded similarly. The MDF was generally recorded immediately after the MTF, which facilitated selection of the modulation frequency. For a majority

of cells where the MTF was available, the modulation frequency chosen for the MDF matched the maximum of the Rayleigh statistic (73/121) or VS (71/121). In 84% of cases (102/121) the modulation frequency chosen for the MDF was equal or adjacent to the VS-derived temporal BMF (tBMF) and Rayleigh maximum in the list of tested modulation frequencies. Note that MDF runs were never taken below 1 Hz, even if 0.7 Hz was the tBMF or Rayleigh maximum in the MTF so that the analysis epochs of 1 s used for spike train classification always included an integer number of modulation cycles. The distribution of modulation frequencies used for the MDFs was concentrated below 20 Hz. The distribution in this range was as follows: 1 Hz: 19; 2 Hz: 28; 5 Hz: 34; 7 Hz: 1; 10 Hz: 31; 20 Hz: 13. Only 11 MDFs were obtained above 20 Hz: 30 Hz: 1; 50 Hz: 4; 100 Hz: 3; 120 Hz: 1; 200 Hz: 1; 300 Hz: 1. As a result, our description of modulation depth coding applies principally to low modulation frequencies, which dominate modulation encoding in auditory cortex of this species (Malone et al., 2007), and speech intelligibility in humans (Drullman et al., 1994a,b; Shannon et al., 1995; Elliott and Theunissen, 2009). Neurons recorded from AI were tested with significantly higher modulation frequencies than those in field R (Wilcoxon rank-sum, $p < 10^{-5}$), reflecting differences in their BMFs, resulting in median values of 5 and 2 Hz, respectively. All MDFs obtained with a modulation frequency exceeding 20 Hz were obtained in AI.

Dynamic level changes: modulation depth coding

Figure 1a indicates the effect of changing modulation depth on the sound level of a SAM stimulus in decibels of SPL (dB SPL) relative to the carrier level of an unmodulated tone. The curves indicate how the instantaneous SPL of the SAM stimulus varies throughout the modulation cycle. At the largest depth shown, 90%, the SPL reaches a maximum of approximately +5.6 and a minimum of −20 dB relative to a pure tone carrier of the same level. In contrast, a modulation depth of 5% produces a relative maximum and minimum of +0.42 and −0.45 dB, respectively, and a 10% modulation spans only +0.83 to −0.92 dB. Thus, modulations at the lower end of the depicted range involve amplitude excursions that are smaller than the resolution of the RLF defined by responses to static tone pips (e.g., 5 or 10 dB steps). Figure 1b depicts the changes in the decibel span (i.e., maximum − minimum) of SAM stimuli as modulation depth is varied from 0 to 95%. For modulations below 40%, the increase in the decibel span is very nearly linear, but further increases in modulation depth reflect the growing asymmetry in the absolute magnitude of the relative minimum and maximum at large depths, which equal $-\infty$ and +6 dB for a fully (100%) modulated signal.

It is easier to appreciate the relevance of the decibel span of SAM signals in the context of actual RLFs, such as those depicted for two neurons in Figure 2, a and e. As indicated by the ±2 SEM lines on the RLF, the neuron whose responses are depicted in Figure 2a was poorly tuned to the SPL of short-duration (100 ms) tone pips. The series of stacked horizontal lines at 60 dB indicate the span of the SAM stimuli comprising the MDFs in Figure 2, b and c. Despite the fact that differences of 20 dB in the tone level did not significantly impact the average firing rate, the response of this neuron was robustly modulated by dynamic changes in stimulus level of less than ±1 dB (Fig. 2c). Both VS and TS (see Materials and Methods) rose steeply with increasing modulation depths below 30%, and continued to increase more slowly at larger depths. The small PSTH inset in Figure 2b illustrates the response to a 10 s pure tone at the carrier level used for SAM (60
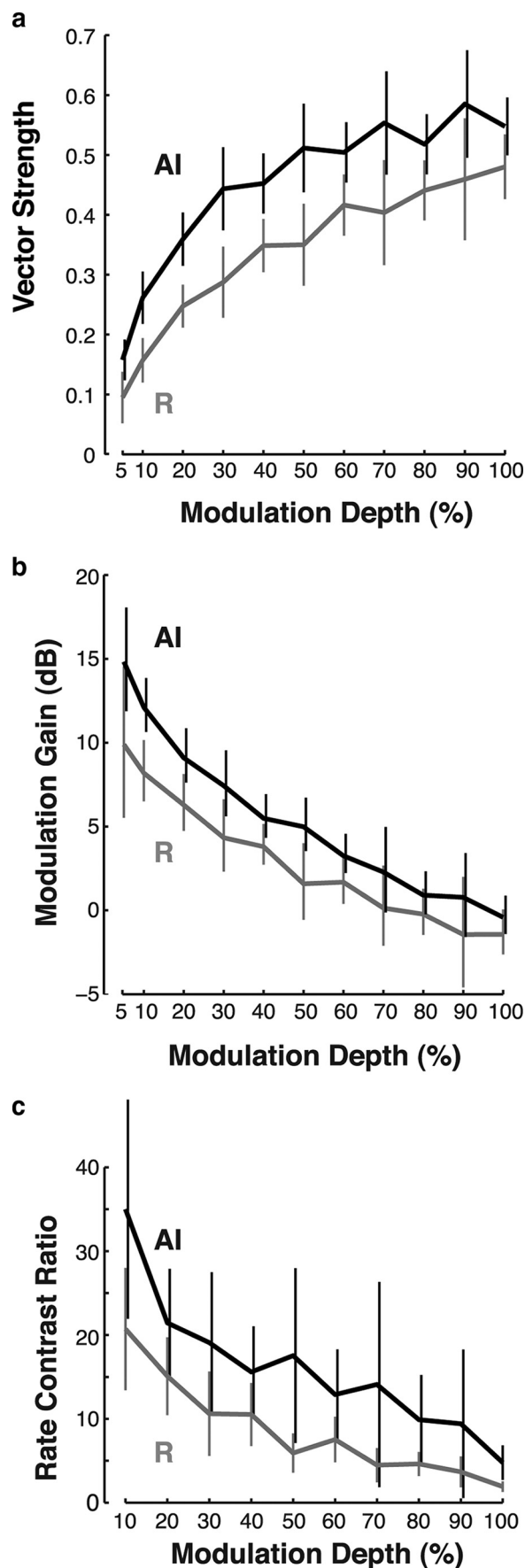
Figure 3. Cortical responses provide high-contrast representations of shallow modulation depths. *a*, Plot of the population average of vector strength against the modulation depth.

dB). For this cell, the unmodulated tone did not produce more spikes than the spontaneous rate, as indicated by the parity of the gray (control) and black (spont.) horizontal lines in Figure 2*b*. As the rate MDF indicates, however, increasing modulation depth produced an essentially monotonic increase in average firing rate.

Figure 2*d* contains a series of MPHs representing the responses to SAM folded on the modulation period (100 ms, for a 10 Hz modulation). The gray curves shown over each panel are analogous to the set of curves in Figure 1*a* indicating the instantaneous SPL of the SAM signal. The corresponding black curves represent the predicted shape of the MPH if the instantaneous firing rate could be obtained by comparing the instantaneous SPL of SAM to the RLF (see Materials and Methods). The vertical level of both sets of curves has been referenced to the height of the largest bin of the MPH in each panel for graphical convenience since the curves are primarily intended to convey the expected degree of modulation. It is clear that the MPHs obtained from this cell are substantially more peaked than this simple RLF-derived model would predict.

In contrast to the responses exhibited by the neuron in Figure 2*b* (inset), the neuron whose responses are depicted in Figure 2, *e* to *h*, exhibited a robust, sustained firing pattern for the 10 s control tone (Fig. 2*f*, inset). This is also reflected in the shapes of the MPHs, which achieve a monotonic increase in VS (Fig. 2*g*) through a progressive deepening of the trough corresponding to the cyclic amplitude minimum (Fig. 2*h*), rather than an increase in the peak height (Fig. 2*d*). As a result, the VS was lower overall than in the neuron discussed above, though comparable values were obtained with TS, which measures the reproducibility rather than the peakedness of the MPH shape. The sustained nature of the neuron's responses also resulted in a much closer match between the shapes of the MPHs, the amplitude envelope, and the RLF-derived prediction.

Modulation gain

As is evident from the MPHs, both of these neurons were quite sensitive to low modulation depths. Historically, modulation sensitivity has been assessed by computing the modulation gain (see Materials and Methods), which involves taking the ratio of twice the VS to the modulation depth (m from 0 to 1), and converting into decibels. The factor of two compensates for the fact that the VS of a half-rectified, fully modulated ($m = 1$) AM waveform is 0.5 (Rees and Palmer, 1989). Thus, a modulation gain of 0 dB indicates that the modulation of the response equals the modulation of the stimulus, under the assumption that the modulation of the response is a half-rectified sinusoid. Because VS is bounded from 0 to 1, the theoretical maximum modulation gain from 5% to 100% depth decreases from ~32 to 6 dB. Figure 3*a* is a box plot of VS at the modulation depths most commonly tested in the data sample. The increase in the median VS with increasing modulation depth is smaller than the increase in modulation depth itself, so modulation gain decreases with increasing modulation depth (Fig. 3*b*). This modulation gain function reflects a change in the population average for VS from roughly 0.2 at 10% modulation to just under 0.5 at 100% modulation (see Malone et al., 2007) (see Fig. 11*d* below).

←

b, Plot of the population average of modulation gain in decibels (see Materials and Methods) against the modulation depth. *c*, Plot of the rate contrast ratio (see Materials and Methods) against the modulation depth. On all curves, vertical lines indicate ±2 SEM, black curves represent data from AI, and gray curves represent data from R.
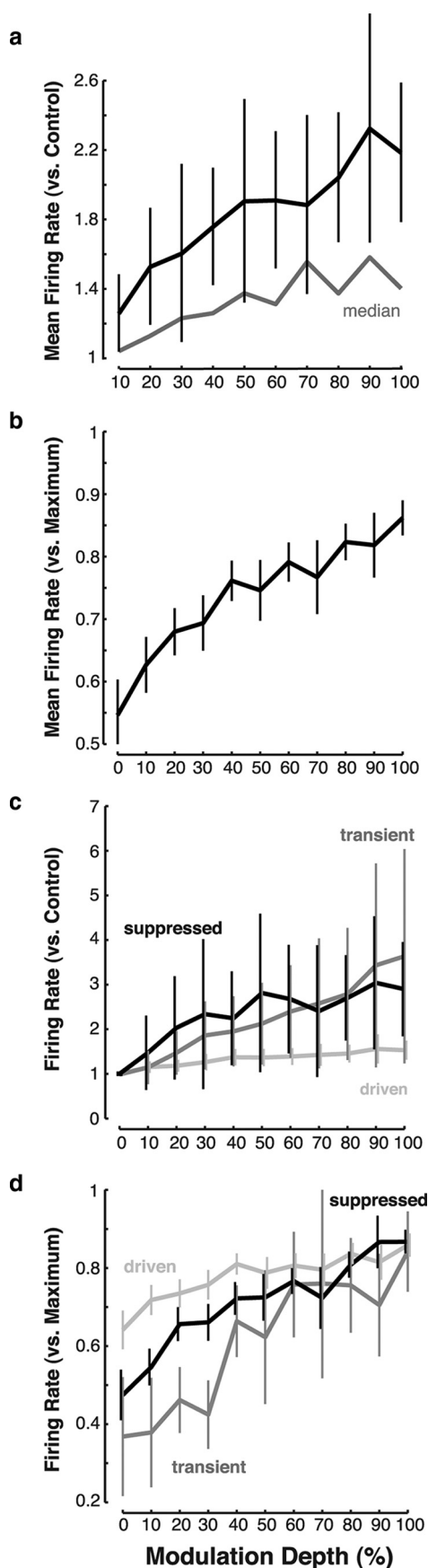
Figure 4. Firing rates averaged across the population increase with modulation depth to a degree predicted by the response to the unmodulated control tone. *a*, How the population average (black curve) firing rate varies across modulation depth when normalized by the firing

We computed the distribution of VS and TS values across all modulation depths for neurons located in AI ($n = 79$) and R ($n = 47$). Mean VS values were significantly (Wilcoxon rank-sum, $p < 10^{-10}$) higher in AI (0.45) than in R (0.35). A similar difference was obtained for TS (0.50 vs 0.42; $p < 10^{-4}$). Higher VS values necessarily imply higher modulation gains. At the lowest tested depths, modulation gains in AI and R were 15 and 10 dB (5%), and 12.3 and 8.3 (10%), respectively. In fact, modulations gain was higher in AI than in R over the entire range of modulation depths ($p < 0.05$), although the differences tended to larger at the lowest modulation depths (where the modulation gains themselves were larger).

Figure 3c plots the change in the rate contrast ratio (see Materials and Methods) as a function of modulation depth. The rate contrast ratio was calculated by taking the maximum firing rate difference present in the three-bin smoothed MPH, and dividing by the maximum firing rate difference present in the RLF-derived prediction (Fig. 2d,h, black curves). At low modulation depths, the cycle by cycle variation in firing rate was between 20 and 30 times larger than would be predicted by a simple look-up model based on the RLF. Thus, when the modulated responses of cortical neurons are referenced to the stimulus modulation, or to their own static tuning functions for amplitude, they appear robustly modulation sensitive.

Firing rate

Figure 4 depicts the relationship between firing rate and modulation depth. As is evident from Figure 4, *a* and *b*, the population average of firing rate increased monotonically with increasing modulation depth. The large ± 2 SEM bars on Figure 2a, however, indicate substantial variability in the relationship between the firing rate for modulated stimuli and for the unmodulated control tone. This variability was considerably reduced when the responses were normalized with respect to the firing rate maximum for each cell (Fig. 2b). We subdivided the data where control responses were available ($n = 129$) into 3 response classes defined by the strength of the response to the unmodulated control tone relative to the spontaneous rate (see Materials and Methods). If the response to the control tone was not significantly different (Wilcoxon rank-sum; $p < 0.01$) than the spontaneous rate, the response was considered to be transient ($n = 56$). Significant and sustained increases or decreases in the firing rate, relative to the spontaneous rate, were classified as "driven" ($n = 59$) and "suppressed" ($n = 14$), respectively. It is clear from Figure 4c that neurons with driven responses at the chosen frequency and carrier level exhibited relatively little variation in firing rate over the full range of modulation depths, but neurons with transient and suppressed responses to the controls roughly tripled their response rates as modulation depth increased.

The responses shown in Figure 2, *b* and *f*, are representative of this difference.

Figure 4d depicts the responses referenced to the MDF maximum like Figure 4b, and indicates the relative compression of the

←

rate elicited by an unmodulated control tone at the same carrier level and frequency. The gray curve shows the changes in the median firing rate. In all panels, vertical lines indicate ± 2 SEs of measurement. *b*, Same data normalized by the maximum firing rate achieved within the MDF. *c*, *d*, Similar to *a* and *b*, respectively, but the data have been subdivided by the MTF-derived response class, so the light gray curve represents the driven response class, the dark gray curve the transient class, and the back curve the suppressed class (see Materials and Methods). Vertical lines indicate ± 2 SEM.

dynamic range of MDFs for the driven response class. When the population means for modulation gain were calculated similarly, however, the curves for the different response classes overlapped throughout the modulation depth range (data not shown), indicating that the driven response class achieved equivalent improvements in VS with increasing modulation depth despite the relatively smaller increases in firing rate. This suggests that for many neurons, response modulation reflects redistribution of spike phases within the modulation period, rather than the addition of synchronized spiking activity.

The monotonic relationship between modulation depth and firing rate, VS, and TS was prevalent in the cortical population. We calculated a monotonicity index (MI) for the average firing rate by taking the ratio of the response to 100% modulation to the maximum response across all modulation depths, such that MI = 1 when the maximum response occurred for the largest modulation depth. Across the population, the population mean/median MI was 0.86/.93. When the MI was calculated for VS and TS, the population mean/median values were 0.89/.97 and 0.86/.96, respectively. Of the 94 neurons that exhibited a significant (Wilcoxon rank-sum; $p < 0.01$) change in firing rate within the MDF, 15 were nonmonotonic (relative to 100% modulation) by the same statistical criterion.

To place the changes in firing rate across modulation depth in context, we compared the dynamic ranges (i.e., maximum − minimum) of the MDFs and MTFs for 122 neurons. Responses to the unmodulated controls were not included in this analysis. There was a strong correlation between the dynamic ranges of the MDFs and MTFs ($r = 0.72$; $p < 10^{-19}$). Changes in firing rate across modulation frequency were significantly larger (Wilcoxon rank-sum; $p < 10^{-4}$) than the changes in firing rate over modulation depth, such that the average ratio of the MDF dynamic range to that of the MTF was 0.75.

Firing rates were very similar across the fields of the auditory core. The mean firing rates, averaged over all neurons and modulation depths, did not differ between AI and R (20.8 vs 19.8 spikes per second; $p > 0.17$), nor did the spontaneous rates (11 vs 10.2; $p > 0.35$). Firing rates did not differ between cortical fields at any particular modulation depth ($p > 0.10$).

Modulation detection: average rate, synchrony, and similarity
The 12 panels in Figure 5a contain example MPHs obtained with a modulation depth of 10%. With the exception of the neuron at the top right, which was included to illustrate the superposition of modulation to both the modulation frequency (10 Hz) and the carrier frequency (100 Hz), these examples represent the highest modulation gains observed in the data sample for 10% modulation. The cumulative distributions in Figure 5b indicate the percentages of cells with significant values of VS and TS values for 10% SAM. VS and TS values are also included on the panels in 5a. The rightward shift of the TS distribution reflects the fact that the minimum values of TS necessary to achieve significance are higher than those for VS (~0.5 vs 0.1) because of the difference in the metrics (see Materials and Methods).

Figure 5 indicates that a sizeable fraction of cortical neurons reliably encode amplitude excursions of a few dB SPL. We estimated the detection limit for modulation depth by finding the lowest significant ($p < 0.001$) values of VS and TS for each cell in the population (see Materials and Methods). We performed an analogous analysis by identifying the lowest depth that elicited a significant change in firing rate relative to the unmodulated control tone (Wilcoxon rank-sum; $p < 0.001$). Histograms of the minimum detectable modulation depths are shown in Figure 6.

a
Responses to SAM: 10% Modulation Depth
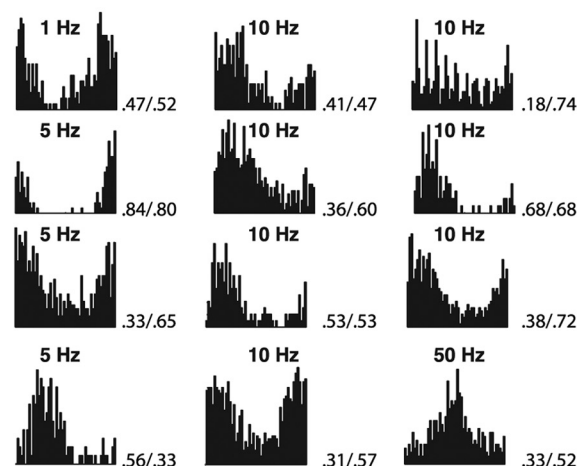
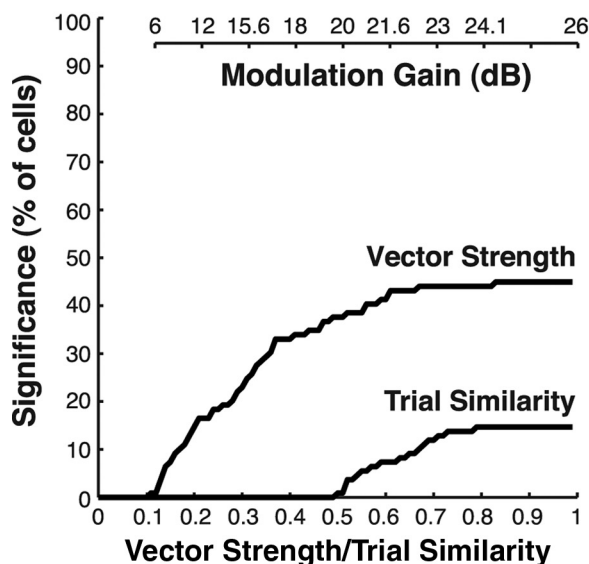## b
Responses to SAM: 10% Modulation Depth

Figure 5. Shallow (10%) modulations of amplitude produce robust cycle-by-cycle response modulation in the most sensitive cortical neurons. *a*, MPHs obtained for 10% SAM for 12 neurons. With the exception of the neuron in the top right panel, the responses of these cells produced the highest values of the Rayleigh statistic in the sample and thus represent the upper limit of modulation sensitivity we encountered. The modulation frequency used is indicated in bold, and the VS/TS values are shown to the right of each MPH. The MPH at the top right is modulated at both the modulation frequency (10 Hz) and the carrier (100 Hz). *b*, Population distribution ($n = 109$) of vector strength and trial similarity for 10% SAM as cumulative distribution curves. Note that the relative displacement of the two curves reflects different significance criteria (see Materials and Methods). For convenience, the modulation gain (decibel) equivalent to each of the VS labels is shown above.

Unfilled bars indicate that the detection threshold for the neuron was the lowest tested depth. In such cases, it is possible that the true detection threshold is lower than indicated. In most cases, however, the MDF spanned a sufficient range to capture the detection threshold by each of the metrics described above.

The VS-derived thresholds were the lowest, followed by TS, and finally, average firing rate. Discounting cases where the

a

b

c
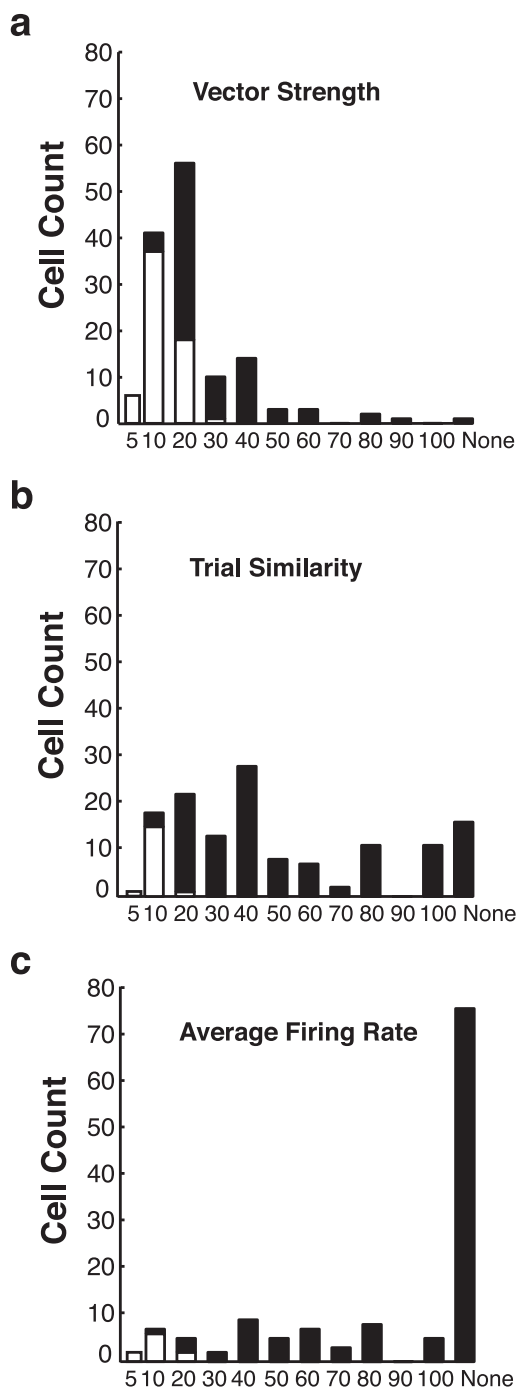
Minimum Detectable Modulation Depth (%)

Figure 6. Measures of temporal structure in the MPH predict lower modulation detection thresholds than measures of average firing rate. ***a*–*c***, Histograms indicate the lowest modulation depth that produced a significant VS (***a***), significant TS (***b***), or significant change in firing rate relative to the unmodulated control (***c***). White bars indicate that the result was obtained at the lowest modulation depth presented to the cell, allowing for the possibility that the actual threshold is lower. Black bars indicate that the threshold value obtained was not the lowest tested modulation depth. The designation "None" means that a threshold value was not obtained for any tested depth.

former—with one exception, every cortical neuron in this sample exhibited significant synchrony (VS) at some point in the MDF (136/137), and similar results were obtained for TS (121/137). When only average firing rate was considered, however, a minority of the neurons (53/129) exhibited significantly different firing rates from the control value at any depth. A depth of 50%, which corresponds to a total amplitude excursion of just under 10 dB SPL, would be detected by 95, 70, and 23% of cortical neurons using VS, TS, and rate, respectively. Changes in the statistical criterion for firing rate ($p < 0.05$) (Nelson and Carney, 2007) lowers some threshold estimates, but does not alter the number of nonsignificant cases (76/129). Given the consistently monotonic increases in firing rate with modulation depth (Fig. 4), it may seem surprising that this number is so high. We calculated the ratio of the variance in firing rate to the mean for both the control and SAM tones, and found that the median ratio was 33% larger for the control tones, which proved significant (Wilcoxon rank-sum; $p < 10^{-4}$). The larger second-to-second variability in firing rate may limit statistically significant differences with respect to the unmodulated control. There was a modest correlation between thresholds based on firing rate and TS ($n = 46$; $r = 0.35$; $p < 0.016$), and a strong relationship between those based on VS and TS ($n = 121$; $r = 0.69$; $p < 10^{-17}$), but the correlation between detection thresholds based on firing rate and those derived from VS was not significant ($n = 52$; $r = 0.07$; $p > 0.6$).

The cumulative distributions of modulation detection thresholds for each of the response classes (driven, transient, suppressed) were very similar (data not shown). Neurons with lower first-spike latencies (see Materials and Methods) tended to have lower firing rate-derived thresholds ($n = 51$; $r = 0.44$; $p < 0.002$). There was also a significant difference ($p < 0.012$) in latency between neurons that did (mean = 24.8 ms) and did not (mean = 29.6 ms) have defined modulation detection thresholds based on firing rate changes. The correlation between latency and VS- and TS-derived thresholds was not significant ($p > 0.05$) in either case.

Spike train classification

One concern when computing modulation detection thresholds in the manner described above is that it is not clear that the statistical criterion for significance is equivalent—the Rayleigh test for VS clearly differs from the bootstrap used for TS, for example. It is also not obvious how differences in VS values could be used to assess modulation depth discrimination (Nelson and Carney, 2007). Very different MPH shapes can produce identical VS values (Joris et al., 2004), indicating that VS cannot capture all of the changes in spike patterns produced by varying modulation depth. Accordingly, we reanalyzed the data present in the MDFs using PSTH-based classification techniques (see Materials and Methods). By generating sets of pairwise comparisons, it is possible to report the results in terms of modulation detection (where the unmodulated control is an element of the pair), and modulation depth discrimination more generally.

Figure 7*a* illustrates the results of this analysis for a neuron that exhibited particularly good modulation discrimination based on average firing rate information. The temporal resolution of the classifiers is determined by the binning applied to the PSTH—when only a single bin is used, the result is classification based on average firing rate alone (green circles). The inset shows the rate MDF for this neuron, one of very few nonmonotonic MDFs in the sample. Performance of the rate-only classifier is generally good, except in those cases where the firing rates in the
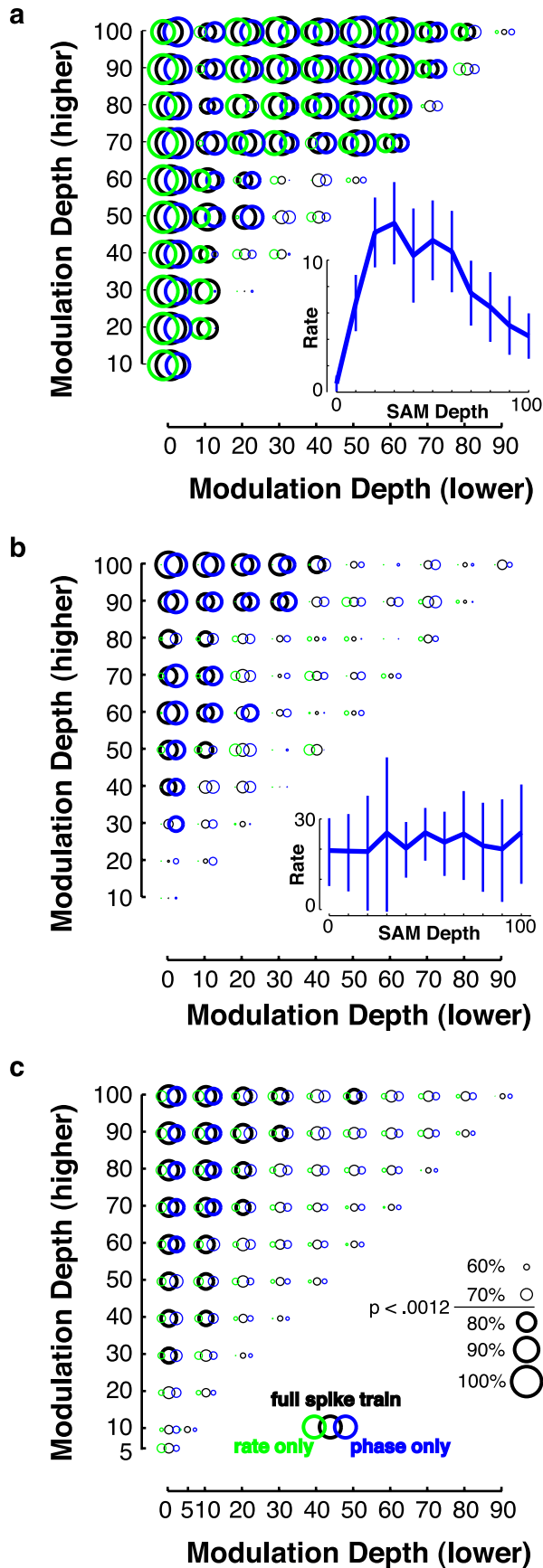
threshold could not be determined, the median thresholds were 20, 40, and 50%, respectively. However, the most striking difference between the timing and rate information present in the responses was the prevalence of defined thresholds for the

MDF are nearly equivalent (e.g., 10 vs 70–100% depths). In those cases, however, the SAM stimuli can typically be discriminated by the phase information present in the spike train (blue circles). When the rate MDF is flat, as it is in Figure 7b, spike timing information provides the sole basis for modulation depth discrimination. Performance of the phase-only classifier can even exceed that of the full spike train classifier due to the confounding influence of the firing rate (e.g., 0 vs 70% depth). We observed 24 cases where the phase-only classifier significantly exceeded chance performance ($p < 0.0012$ for 72.5%) for at least one paired comparison, but the rate-only classifier did not. The reverse was much less common, and occurred only twice. Typically, however, both firing rate and phase information contribute to the discrimination of SAM depth, though their relative contributions vary across stimulus pairs.

The population averages for classifier performance for all pairs drawn from the MDFs are shown in Figure 7c. It is clear from this plot that modulation discrimination is not strictly dependent on the difference in modulation depth. Performance also does not appear to depend on differences in the decibel span of the stimuli, which would predict better performance when discriminating modulation depths drawn from the higher end of the range (e.g., 80 vs 100%) (Fig. 1). Generally, mean performance tended to decrease along the diagonals which define equivalent differences in modulation depth, such that the population averaged performance for 0 versus 30% was significantly better than performance for 40 versus 70% (Wilcoxon rank-sum; $p < 0.0006$), or 70 versus 100% ($p < 0.008$). This trend in performance may reflect the fact that the population averages for firing rate (Fig. 4) and vector strength (Fig. 3) are more steeply sloped in the lower range of modulation depth, suggesting that changes in depth within those ranges produce proportionately greater changes in the spike patterns of cortical responses.

Figure 7c also indicates that the performance of the phase-only classifier was generally better than the performance of the rate-only classifier. Averaged over all cells and all comparisons, the phase-only classifier significantly outperformed the rate-only classifier (medians of 65 vs 60%; Wilcoxon rank-sum; $p < 10^{-48}$). As expected, the full spike train classifier (median = 70%) outperformed both the phase-only ($p < 10^{-46}$) and rate-only ($p < 10^{-163}$) classifiers.

Classifier performance was also dependent on the recording location. Neurons from AI outperformed neurons recorded from field R for the full spike train (Wilcoxon rank-sum; $p < 10^{-29}$), phase-only ($p < 10^{-15}$), and rate-only ($p < 10^{-30}$) classifiers. Examination of the median differences in performance for the different classifiers revealed that in AI the phase-only classifier outperformed the rate-only classifier by 3%. In field R this difference was significantly larger ($p < 10^{-6}$), at 5.5%. Analogously, the average performance difference between the full spike train and phase-only classifiers was significantly larger in AI than in R (3.5 vs 2.0%; $p < 10^{-8}$), while the average performance differ-

←

to the abscissa, and the depth for the second is mapped to the ordinate. Differences in modulation depth vary with the distance from the diagonal. The performance of the classifiers, in percentage correct, is indicated by the diameter of each circle, such that 50% = 0. The icons in the bottom right panel indicate how performance relates to circle size. Performance that is significantly above chance is indicated by heavier line weighting (72.5% corresponds to a p value of ~0.001). Black circles represent the performance of the full spike train classifier; blue circles: the phase-only classifier; green circles: the rate-only classifier. **a, b,** Results for individual neurons. The insets show the RLFs for each neuron (vertical lines indicate ±2 SEM). **c,** Classifier performance for each comparison averaged over the population.

Figure 7. Classifier performance for pairwise discrimination of SAM stimuli varying in modulation depth illustrates how both spike timing and spike rate information contribute to modulation depth discrimination. The modulation depth for the first stimulus in the pair is mapped

ence between the full spike train and rate-only classifiers was slightly smaller in AI than R (6 vs 8%; $p < 0.01$). These data suggest that average firing rate information features more prominently in AI than in R.

We also compared classifier performance for the different response classes. When the full spike train was used, the difference between the sustained (driven or suppressed) and transient classes was not significant (70 vs 69%; Wilcoxon rank-sum; $p > 0.08$). However, phase-only classifier performance for the sustained response class significantly ($p < 10^{-5}$) exceeded that for the transient response class (66 vs 63%). Conversely, rate-only classifier performance was significantly higher ($p < 10^{-4}$) for the transient response class, despite nominally equivalent medians (60%), because the transient distribution is skewed toward higher values above and below the median. This likely reflects the greater dynamic range in the MDFs of the transient response class (Fig. 4).

Modulation detection and discrimination: classification

The pairwise classification procedure illustrated in Figure 7, *a* and *b*, can also be used to create distributions of modulation detection thresholds analogous to those depicted in Figure 6. For detection, we identified the smallest modulation depth that could be successfully discriminated from the unmodulated control (Fig. 7*a,b*, lowest, thick-lined circle in the first column) (see Materials and Methods). The results of this analysis, separated by classifier type, are shown in Figure 8, *a*, *c*, and *e*. Detection thresholds for the full spike train and phase-only classifiers were very similar, reflecting a strong underlying correlation in the results on a cell-by-cell basis ($n = 87$; $r = 0.81$; $p < 10^{-20}$). The correlation between thresholds for the full spike train and rate-only classifier was similarly high among those cells where thresholds could be defined in both cases ($n = 54$; $r = 0.74$; $p < 10^{-9}$). These results are not surprising since the full spike train classifier is informed by both spike rate and spike timing information. The correlation between thresholds obtained with the rate-only and phase-only classifiers was weaker ($r = 0.46$) but still significant ($p < 0.0008$) despite the fact that the information available to them is mutually exclusive.

We further investigated the relationship between timing and rate information by computing the correlation between rate-only and phase-only classifier performance for every paired comparison in the sample ($n = 3329$) for all (111) neurons. There was a weak but highly significant relationship between performance for the classifier types ($r = 0.30$; $p > 10^{-69}$). The distribution of correlation coefficients obtained when the relationship between rate-only and phase-only classifier performance is calculated for individual cells evidences substantial heterogeneity, however. For example, this relationship was strong for the comparisons depicted in Figure 7*a* ($r = 0.59$; $p < 10^{-5}$), but completely absent for those in Figure 7*b* ($r = 0.02$; $p > 0.88$). The mean correlation coefficient for all cells was 0.22, within a range from −0.53 to 0.94, and significance was limited to 27 of 111 cells ($p < 0.01$). All significant correlations were positive, save one (−0.53).

The thresholds obtained with each of the 3 classifier types are compared with those obtained with average firing rate, VS, and TS in Table 1. Thresholds based on rate and the rate-only classifier should be highly correlated (Fig. 6*c* vs Fig. 8*c*) because both the classifier result and the statistical comparison depend on the dynamic range of the MDF and the second-by-second variation of the firing rate (see Materials and Methods). Firing rate information can also support depth discrimination by the full spike train classifier, and the correlation was substantial. Despite being
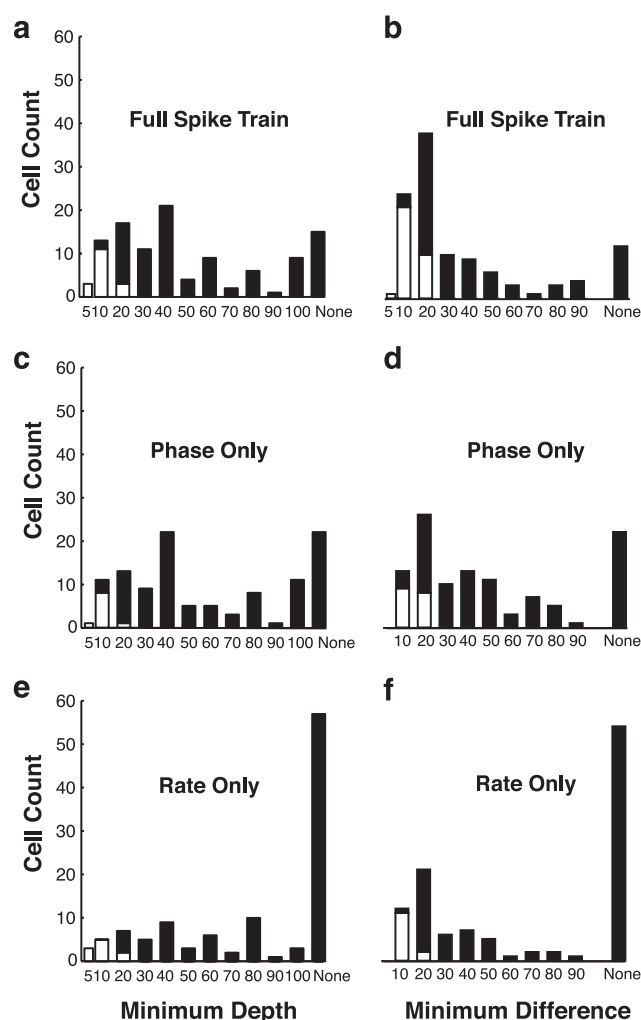

Figure 8. Distributions of minimum detectable modulation depths and differences obtained with spike train classification techniques indicate that spike timing information is crucial for many cells. *a*, *c*, *e*, Left histograms indicate the lowest modulation depth that could be reliably (72.5% corresponds to a *p* value of < 0.001) discriminated from the unmodulated control tone. *b*, *d*, *f*, Right histograms indicate the smallest difference in modulation depth (e.g., for 10%, 10 vs 20%, or 50 vs 60%) that could be reliably discriminated by the classifiers. The classifier type is indicated above each panel. White bars indicate that the result was obtained at the lowest modulation depth (*a*, *c*, *e*) or modulation depth difference (*b*, *d*, *f*) presented to the cell, allowing for the possibility that the actual threshold is lower. Black bars indicate that the threshold value obtained was not the lowest tested modulation depth.

Table 1. Correlations for modulation detection thresholds

Metric	Classifier type	*r* value	*p* value	*n*
Average firing rate	Spike train	0.77	$< 10^{-5}$	41
Average firing rate	Phase only	0.46	< 0.004	37
Average firing rate	Rate only	0.66	$< 10^{-8}$	40
Vector strength	Spike train	0.39	< 0.0002	95
Vector strength	Phase only	0.43	< 0.0043	88
Vector strength	Rate only	0.12	> 0.4	53
Trial similarity	Spike train	0.68	$< 10^{-14}$	89
Trial similarity	Phase only	0.74	$< 10^{-12}$	83
Trial similarity	Rate only	0.33	< 0.018	49

independent, in principle, however, the rate and phase-only classifier thresholds were also significantly correlated, though the relationship was weaker. This supports the observation above that the quality of spike rate and spike timing information covaries in some cortical neurons. For analogous reasons, TS-derived

thresholds (Fig. 8b) largely mirrored those obtained with the phase-only and full spike train classifiers since TS and the classifiers depend on the reproducibility of spike timing. The significant but unexpected correlation between TS and rate-only classifier thresholds compliments the similarly unexpected correlation between the phase-only classifier and average firing rate thresholds.

The most striking difference between Figures 6 and 8 is the difference between the thresholds derived from VS and those derived from the classifiers. It is possible that the Rayleigh criterion is relatively permissive, given that 3 of 137 neurons in our sample were "synchronized" to the unmodulated control tone even when the significance criterion was $p < 0.001$. Thresholds based on VS are far more concentrated at low depths, and only a single cell did not have a VS-defined threshold. The correlations between the thresholds for VS and the full spike train and phase-only classifier were significant, but there was little relationship with the rate-only classifier thresholds (Table 1).

Thresholds for modulation discrimination are represented by the histograms in Figure 8, b, d, and f. Whereas Figure 7c indicated the population average of classifier performance, Figure 8, b, d, and f, depicts the distribution of thresholds obtained for the same set of paired comparisons. The unmodulated controls were excluded from this analysis so all thresholds reflect modulation depth discrimination per se. Although mean thresholds for modulation discrimination were lower than for modulation detection, this difference may reflect the fact that modulation detection relies on a single comparison at each depth, while modulation discrimination for a given difference in modulation depths involves multiple comparisons for all depths of <90%.

Explaining the shape of the modulation period histogram
Having shown that cortical neurons are quite sensitive to amplitude modulation, we were interested in how the spike rate output relates to the instantaneous SPL of SAM stimuli. To quantify the extent of this relationship, we relied on the fact that during each modulation cycle of SAM, each SPL value occurs twice, as can be seen by drawing a horizontal line through the envelope profiles in Figure 1a. As a result, one can compare the firing rates elicited by each instance of a comparable SPL within the modulation cycle. Each MPH was rotated to center the bin corresponding to the SPL minimum of the SAM stimulus, then rotated again to compensate for the response latency estimated from the group delay. If a particular SPL elicits a particular firing rate, then the first and second halves of the MPH should be mirror images of one another. We quantified this in terms of a symmetry index (see Materials and Methods) that varied from 0 to 1, where 1 indicates perfect mirror symmetry. For example, the SI in Figure 2d for a depth of 80% is exactly 0, whereas the corresponding SI for 80% depth in Figure 2h is 0.79. Responses where the VS was not significant ($p < 0.001$) were excluded from this and following analyses because the phase distribution of unmodulated responses is not informative.

The mean SI for the population was 0.58, suggesting that cortical responses overall are consistently related to the ongoing SPL. To demonstrate this more forcefully, we iteratively computed distributions of SIs for all the MPHs in our sample, but instead of correcting for the response latency, we rotated each MPH by a phase value selected at random. This preserves the modulation present in the MPHs, but randomizes its relationship to the SPL-defined symmetry axis of the modulation period. On each iteration, we compared the actual and random phase SI distributions. In every instance ($n = 100$), the randomization significantly re-
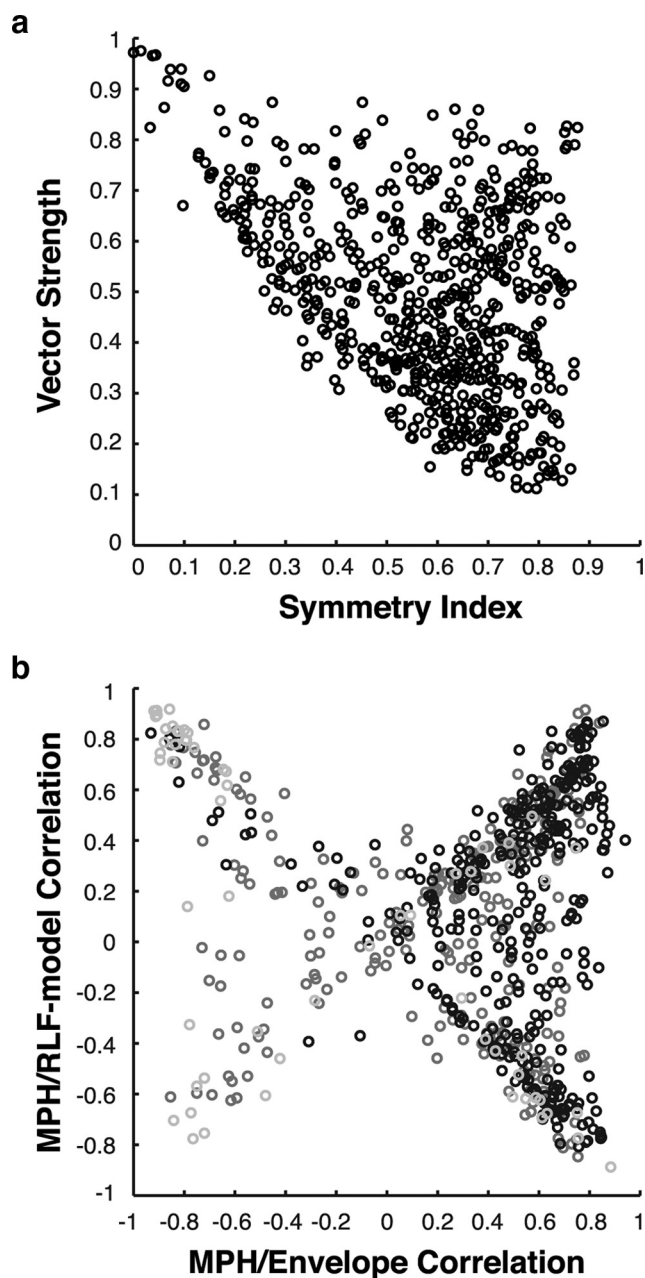
Figure 9. Analysis of MPH symmetry and the similarity of MPH shape to SAM envelopes indicates a dissimilarity in the encoding of static and dynamic level differences. ***a***, Distribution of SIs, which quantify the mirror symmetry of the response (the envelope of SAM is mirror symmetric) as a function of VS. ***b***, Scatterplot comparing the correlations between MPH shapes and the SAM envelope profiles (in decibels; abscissa) against the correlations between the MPH shapes and the MPHs predicted on the basis of the RLF (see Materials and Methods). Black circles indicate the responses of neurons with driven responses to the unmodulated control tone, dark gray circles indicate transient responses, and light gray circles indicate suppressed responses. In both panels, the data represent all tested neurons at all tested modulation depths, so a single neuron may contribute as many points to the graph as there were points in its MDF that elicited significant modulation.

duced the SI distribution ($p < 0.01$). The scatterplot in Figure 9a shows the relationship between VS and the SI for all the significantly modulated MPHs in the sample ($n = 692$). Although high SIs could be associated with both high and low VSs, low SIs were never associated with low VSs, as indicated by the triangular-shaped void on the lower left of the figure. This feature of the distribution, and the negative correlation ($r = -0.43; p < 10^{-31}$)

between VS and SI, is specific to the phase of cortical responses to SAM. When the phases of the MPHs are randomized, low VSs and SIs occur, and the correlation is substantially reduced (for 100 iterations, the mean r value was −0.09). On average, larger modulation depths produced less symmetrical responses (Kruskal–Wallis; $\chi^2 = 41.9$; $p < 0.0001$), as did faster modulation frequencies ($\chi^2 = 75.6$; $p < 10^{-14}$). As suggested by the examples in Figure 2, neurons with transient responses to the unmodulated control tones (SI mean: 0.53) had less symmetric MPHs than cells that were driven (SI mean: 0.61; Wilcoxon rank-sum; $p < 10^{-103}$) or suppressed (SI mean: 0.60; $p < 10^{-33}$).

Given the foregoing observations, we attempted to predict the shape of the MPHs by using the decibel profile of the stimulus envelope, as shown in Figure 1a. In this case, we simply computed the cross-correlation of the stimulus profile and the latency-corrected MPHs. Note that the previous analysis of symmetry is insensitive to the exact relationship between firing rate and SPL, and instead reflects whether increases and decreases of SPL are encoded similarly. Because previous work (Malone et al., 2007) indicated that in some cells, such as a nonmonotonically tuned neuron stimulated well above its best level, increases in firing rate could be associated with decreases in SPL, we attempted to use the shape of the RLF to predict the shape of the MPH response. In this case, the interpolated RLF was treated as a lookup table relating firing rate to instantaneous SPL within the modulation cycle (see Materials and Methods). Because most tested depths traverse a relatively narrow range of the RLF (Fig. 2a,e, icons), the RLF-lookup model effectively tests whether the local slope of the RLF predicts the nature of the relationship (i.e., direct or inverse) between SPL and firing rate changes.

Figure 9b shows the distribution of the correlations between the MPH shapes and the stimulus envelope on the abscissa, and between the MPH shapes and the RLF-lookup model predictions on the ordinate. In both cases, the marginal distributions are bimodal, but favor positive values, particularly for the correlations with the stimulus envelope. Overall, the correlations between the MPH shape and the stimulus envelope (mean/median: 0.32/0.50) were substantially higher (Wilcoxon rank-sum; $p < 10^{-19}$) than those based on the RLF-lookup model (mean/median: 0.11/0.21). The distinctive "X" shape of the scatterplot reflects a strong correlation between the absolute magnitudes of the predictions ($r = 0.72$; $p < 10^{-133}$). However, the relative densities of the upper left and lower right limbs of the "X" indicate that although the RLF model prediction can sometimes correct the direction of the firing rate and SPL relationship (upper left), more commonly, the relationship is direct despite a negative local slope in the RLF (lower right). When we iteratively recalculated these distributions after rotating each of the MPHs by a random phase value, the MPH shape predictions based on the envelope were dramatically reduced in every instance (Wilcoxon rank-sum; $p < 10^{-14}$ for all 100 iterations), resulting in a nearly uniform distribution centered on zero, and those based on the RLF were also significantly reduced in every case ($p < 0.0002$).

Although the RLF did not improve predictions of MPH shape overall, the response class designations based on the responses to the unmodulated control tones had a highly significant impact on the relationship between the MPH shape and the envelope (Kruskal–Wallis; $\chi^2 = 112.5$; $p \sim 0$). Across all tested depths, correlations from neurons driven by the tested carrier level ($n = 51$) were positive ~94% of the time (332/355; black circles), whereas for the transient and suppressed response classes, the proportions were 72% (199/276; gray circles) and 39% (24/61; light gray circles), respectively. The corresponding means were
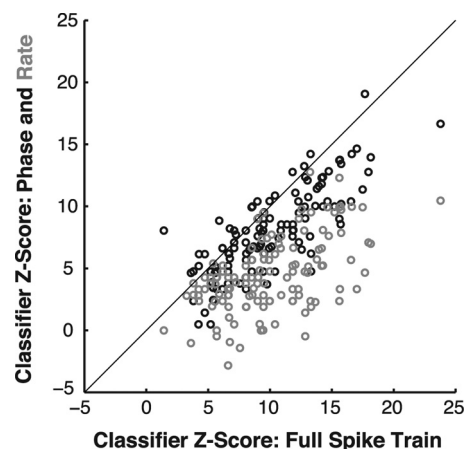

Figure 10. Temporal discharge patterns reliably discriminate tone pip stimuli varying in SPL. Results from the phase-only (black) and rate-only (gray) classifiers are plotted against classifier performance for the full spike train. The diagonal line indicates performance equal to that of the full spike train classifier.

0.50, 0.22, and −0.25. These results indicate that the steady-state response to the carrier level informs the shape of the MPH in cells with sustained responses.

Level difference coding

The preceding analyses demonstrate that the local slope of the RLF is not a reliable predictor of firing rate changes for dynamic stimuli like SAM tones. It is possible that the RLFs were not measured with sufficient tone repetitions to estimate the local slope accurately. However, the dynamic range of the RLF near the carrier level also dramatically underestimates the response modulation observed for SAM (Fig. 4c). It is also possible that averaging over a relatively long duration (150 ms) may have obscured temporal features of the PSTHs that encode static differences in level with better fidelity. To assess this possibility, we applied the same PSTH classification techniques to the PSTHs comprising the RLFs for each neuron. The RLFs were typically obtained at a resolution of 10 dB SPL from 0 to 80 dB. Additional points were added if the neuron was observed to be particularly sensitive during the initial run (e.g., −10 dB) (Fig. 2e). We corrected for differences in the number of stimuli presented to different neurons by standardizing classifier performance in terms of z-scores. The results of this analysis are shown in the scatterplot in Figure 10. Performance for the phase-only (black circles) and rate-only (gray circles) classifiers on the ordinate is compared against performance for the full spike train classifier along the abscissa. As is evident, classifier performance was substantially better than chance (0) for most neurons for all the classifiers. It is worth noting that the average z-scores for the level difference discrimination were substantially larger than those for modulation depth discrimination (full spike train: 9.9 vs 5.2; phase-only: 7.9 vs 4.0; rate-only: 4.5 vs 2.6). This may reflect the fact that the typical level differences comprising the RLFs, at ~10 dB, are larger with respect to psychophysical thresholds for intensity discrimination (Zwicker, 1952) than the 10% steps used for the MDFs are with respect to modulation depth discrimination thresholds.

Surprisingly, the phase-only classifier significantly outperforms the rate-only classifier (Wilcoxon rank-sum; $p < 10^{-14}$), even though the phase-only classifier effectively flattens the RLF itself. Although there was a significant performance advantage for the full spike train classifier relative to the phase-only classifier ($p < 10^{-24}$), their performances were strongly correlated ($r =$
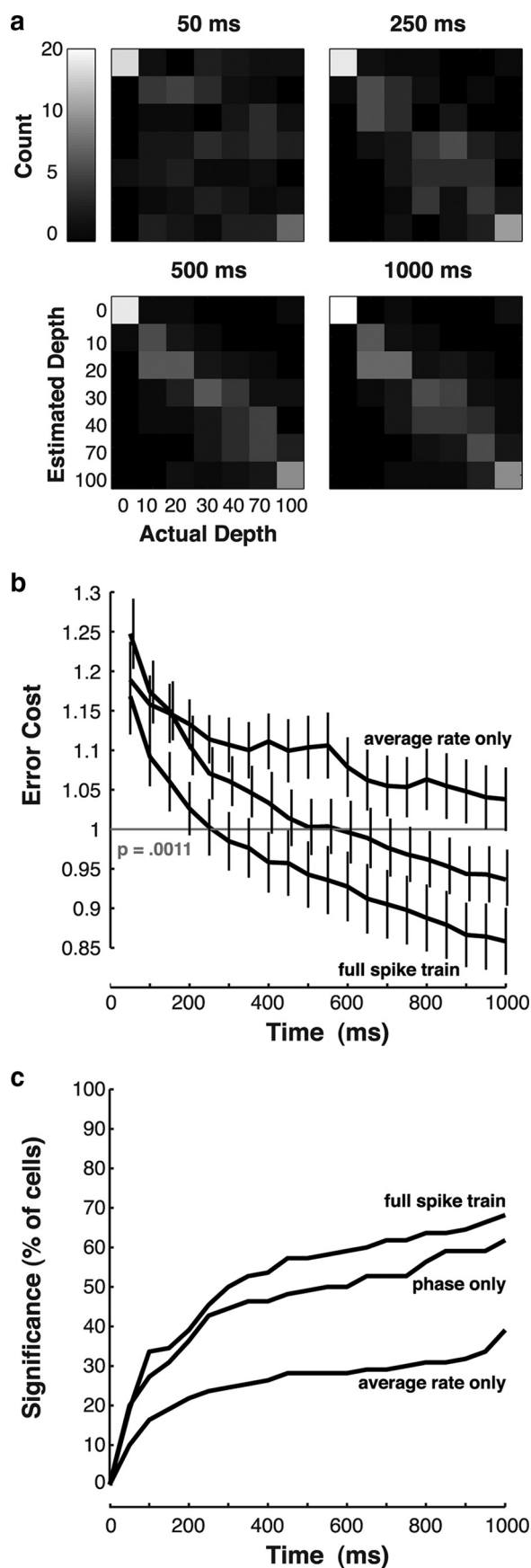
0.80; $p < 10^{-29}$). The correlations between the rate-only and full spike train classifier ($r = 0.54$; $p < 4.5 \cdot 10^{-11}$) and phase-only classifier ($r = 0.28$; $p < 0.002$) were substantially weaker. Together, these results suggest that the RLF measured by averaging spike rates over relatively long intervals (150 ms) fails to capture important features of level coding in the auditory cortex. Performance for modulation depth and static level discrimination was weakly but significantly correlated for both the full spike train ($r = 0.39$; $p < 10^{-4}$) and phase-only ($r = 0.23$; $p < 0.017$) classifiers, but not for the rate-only classifier ($r = 0.09$; $p > 0.37$). This finding suggests that encoding static level differences and dynamic changes in SPL with temporal spike patterns entails related processing demands.

Predicting classifier performance

Cell-by-cell averages of performance on the modulation depth and level discrimination tasks were not significantly ($p < 0.01$) correlated with SIs, or with the MPH and envelope profile correlation magnitudes. Analogous averages of VS were modestly but significantly correlated with modulation depth discrimination performance (full spike train: $r = 0.36$; $p < 0.0002$; phase-only: $r = 0.28$; $p < 0.0028$; rate-only: $r = 0.28$; $p < 0.0033$), but not level discrimination performance ($p > 0.1$ for all 3 classifiers). In contrast, average TS was significantly correlated with performance on both modulation depth discrimination (full spike train: $r = 0.61$; $p < 10^{-11}$; phase-only: $r = 0.54$; $p < 10^{-8}$; rate-only: $r = 0.36$; $p < 10^{-4}$) and level discrimination (full spike train: $r = 0.36$; $p < 10^{-4}$; phase-only: $r = 0.31$; $p < 0.0003$; rate-only: $r = 0.24$; $p < 0.0049$). The stronger correlations with TS suggest that the relationship between performance on these disparate tasks reflects cell-by-cell differences in the trial-by-trial variability of their responses.

In light of this observation, we tried to identify other response characteristics that predicted both static and dynamic level discrimination performance. We hypothesized that neurons that achieve higher maximum firing rates would perform better at MDF and RLF discrimination, since the dynamic range of the neuron, in terms of changes in firing rate, is analogous to the "contrast" present in the PSTH. The local maximum firing rate was determined by finding the largest single bin count, using a bin width of 10 ms, from all of the PSTHs obtained for tone pips. This simple metric was significantly correlated with both modulation depth (full spike train: $r = 0.41$; $p < 10^{-5}$; phase-only: $r = 0.36$; $p < 0.0002$; rate-only: $r = 0.31$; $p < 0.0012$) and static level discrimination (full spike train: $r = 0.65$; $p < 10^{-16}$; phase-only: $r = 0.56$; $p < 10^{-11}$; rate-only: $r = 0.37$; $p < 10^{-4}$). Cells with higher synchrony cutoffs, defined as the highest modulation frequency that produced a significant VS, were somewhat better at modulation depth discrimination when temporal information was available (full spike train: $r = 0.45$; $p < 10^{-5}$; phase-only: $r = 0.32$; $p < 0.0016$; rate-only: 0.22; $p < 0.0239$), but the relationship with level discrimination was marginal at best (full spike train: $r = 0.14$; $p > 0.1$; phase-only: $r = 0.21$; $p < 0.0210$; rate-

←

stimulus was subdivided into 20 epochs, so each cell of the matrix can vary from 0 to 20. The error costs for these four confusion matrices were 0.92, 0.47, 0.40, and 0.39 for 50, 250, 500, and 1000 ms, respectively. ***b***, How the population average of the error cost varies as a function of the epoch duration for each of the classifiers. Error cost for each neuron was normalized relative to the error cost that would be significant at the criterion level ($p < 0.0012$) for a confusion matrix of the equivalent size. ***c***, Set of cumulative distribution curves indicating the percentage of cells that achieved better-than-chance classifier performance in terms of error cost ($p < 0.0012$) as a function of the analysis epoch duration.

Figure 11. Classifier performance varies with the analysis epoch duration. ***a***, Four confusion matrices obtained at varying epoch durations (50, 250, 500, and 1000 ms). Correct identification of the SAM stimuli is indicated by the number of entries along the diagonal of the matrix. Each

only: $r = -0.03$; $p > 0.7$). Spike timing jitter, quantified as the SD of first spike latency for the tone at each neuron's best level, was inversely correlated with modulation depth discrimination performance (full spike train: $r = -0.32$; $p < 0.0012$; phase-only: $r = -0.28$; $p < 0.0048$), although the results for the rate-only classifier were marginal ($r = -0.23$; $p < 0.0165$). The modest correlation values reflect the fact that while neurons in the top quartile of classifier performance had uniformly low spike timing jitter, low spike timing jitter was not itself a guarantee of good discrimination performance. Analogous results were obtained for static level discrimination (full spike train: $r = -0.49$; $p < 10^{-8}$; phase-only: $p < 10^{-7}$; rate-only: $r = -0.17$; $p > 0.05$). There was a modest but significant negative correlation between average minimum response latency and modulation depth discrimination for the full spike train classifier ($r = -0.33$; $p < 0.0007$), but this relationship was marginal for the others (phase-only: $r = -0.25$; $p < 0.0115$; rate-only: $r = -0.22$; $p < 0.0277$). Maximum firing rate and spike timing jitter were significantly and inversely correlated ($r = -0.53$; $p < 10^{-9}$), but neither variable was significantly ($p > 0.05$) correlated with the synchrony cutoff.

Classification of both SAM and static tones occurs rapidly

PSTH-based classification techniques allow for changes in temporal resolution of the analysis, as well as the temporal interval to be analyzed. The optimal temporal resolution for the discrimination of modulation depth fell within a range of ~8–20 ms (Malone et al., 2007). Given this distribution, we fixed the temporal resolution of the full spike train and phase-only classifiers at 10 ms, and examined the effect of varying the analysis epoch. This allowed us to determine the minimum interval required to discriminate among the various modulation depths presented to each cell. The resolution of the rate-only classifier equaled the length of the analysis epoch. Figure 11a shows the confusion matrices for the full spike train classifier obtained for a single neuron for analysis epochs of 50, 250, 500, and 1000 ms. Discrimination of modulation depth in this cell was particularly rapid, and performance, assessed by the error cost (see Materials and Methods) was significantly above chance after only 50 ms. As is clear from the confusion matrix at the upper left of Figure 11a, the basis for this performance was the rapid identification of the unmodulated (0%) and fully modulated (100%) SAM stimuli in the set. As the analysis epoch increases, intermediate modulation depths are discriminated with greater accuracy, as indicated by the changes in the confusion matrices from 250 to 500 ms.

Figure 11b depicts the how classifier performance varies with increasing analysis epoch durations. Epoch duration varied from 50 to 1000 ms in 50 ms steps. Because chance performance varies with the number of modulation depths presented, we normalized for such differences in the data sample by referencing the error cost of each confusion matrix to the appropriate statistical criterion before averaging over the population. The performance of all the classifiers improves with time, but the improvements of the full spike train and phase-only classifier occur more rapidly, reaching the significance criterion indicated on the graph at ~250 and 500 ms, respectively. As the example in Figure 11a suggests, however, many neurons show significant modulation discrimination more rapidly. Conversely, some cells never achieve it, as indicated by the cumulative distributions in Figure 11c showing the percentage of cells with significant reductions in the error cost as a function of epoch duration. Although the majority (67/110) of neurons do not meet the statistical criterion when using the rate classifier, a substantial minority of those that do (18/43) do so within the first 100 ms. Similar proportions were

found for the full spike train (37/75) and phase-only (30/68) classifiers. Given the low modulation frequencies used to obtain the MDFs, better-than-chance performance was achieved within a single modulation cycle in roughly one third of cortical neurons (e.g., for the full spike train, 37/110 neurons). The median minimum epoch durations for the full spike train, phase-only, and rate-only classifiers were 1, 0.83, and 1.25 cycles, respectively. These low values reflect the use of very low (<20 Hz) modulation frequencies in most cells. The classifiers differed chiefly in the percentage of neurons that achieved the statistical criterion for performance, due to the comparatively poor performance of the rate-only classifier. Among those cells that did achieve it, however, the distributions of minimum epoch durations did not differ (Kolmogorov–Smirnov, $p > 0.3$ in all cases). Roughly half of the neurons in both AI (46/79, or 58%) and R (24/47, or 51%) exhibited significant performance, and among those that did, the distributions of minimum epoch durations did not differ significantly for any of the classifiers ($p > 0.1$ in all cases).

As Figure 11b suggests, classifier performance continued to improve with increasing analysis epoch duration. Examination of the error cost functions of individual cells confirmed that nearly all cells exhibited a monotonic improvement in classifier performance with increasing duration. Obviously, longer epoch durations increase the number of spikes available to the classifiers. For each epoch duration, we computed the total number of spikes comprising each confusion matrix, as well as the associated error cost, for all neurons in the sample. The correlation between the spike count and the error cost was significant for all classifier types, but the relationship was stronger for the full spike train ($r = -0.45$; $p < 10^{-110}$) and phase-only ($r = -0.46$; $p < 10^{-118}$) classifiers than for the rate-only classifier ($r = -0.22$; $p < 10^{-24}$). For the rate-only classifier, the additional spikes only improve the estimate of the average firing rate; for the other classifiers, however, the additional spikes may be necessary to define the contours of the PSTH templates.

We performed a similar analysis for the classification of the tone pips used to generate the RLF for each cell. Classifier performance was assessed at 50, 100, 150, 200, 250, 300, and 400 ms for all cells where the RLF was available ($n = 131$). Performance for the full spike train classifier increased from 50 to 100 ms (Wilcoxon rank-sum; $p < 10^{-4}$), but saturated thereafter, such that there was no effect of duration when the 50 ms analysis epoch was excluded (Kruskal–Wallis; $p > 0.7$). For the phase-only classifier, performance was statistically equivalent from 150 to 400 ms (Kruskal–Wallis; $p > 0.5$), with significant improvement from 50 to 100 ms ($p < 10^{-4}$), and marginal improvement from 100 to 150 ms ($p < 0.05$). Performance of the rate-only classifier varied nonmonotonically over duration (Kruskal–Wallis; $\chi^2 = 18.4$; $p < 0.006$), being best at 150 ms, and declining slightly at longer durations, most likely due to the effects of averaging spontaneous spikes occurring after the cessation of the 100 ms tone. As a result, performance was better at 150 ms than at 400 ms ($p < 0.0012$), for example. The performance difference between 50 and 100 ms was marginally significant ($p < 0.0385$). Cell by cell, level discrimination performance was significant ($p < 0.0012$) in 98, 92, and 60% of cases for the full spike train, phase-only, and rate-only classifiers, respectively. Average performance for the level discrimination task was best at a temporal resolution of 25 ms for the full spike train classifier, and 10 ms for the phase-only classifier, though in both cases performance was fairly similar at those resolutions.

Discussion

Perhaps the most prominent change in the representation of SAM in the ascending auditory system is the dramatic reduction in the upper limit of phase-locking to the modulation envelope (Langner, 1992; Joris et al., 2004). In contrast, however, the sensitivity of primate cortical neurons to small, low-frequency amplitude fluctuations is essentially undiminished, and in some cases enhanced, relative to the auditory nerve (Joris and Yin, 1992). This suggests that the ascending auditory pathway is designed to privilege these features of acoustic signals.

The information necessary to account for psychophysical discrimination thresholds in both humans (Lee and Bacon, 1997; Kohlrausch et al., 2000; Moore and Glasberg, 2001) and macaque monkeys (Moody, 1994) is present in the more sensitive cortical units in our sample, provided precise temporal discharge patterns rather than global averages of firing rates are considered. Spike discharge patterns also effectively discriminated static tone pips varying in level but otherwise identical in their envelopes. This result generalizes the notion that naturalistic stimuli are encoded by temporal discharge patterns, which has been demonstrated for a range of vocalization sounds in a variety of species, including conspecific calls in macaque monkeys (Russ et al., 2008; Recanzone, 2008), guinea pigs (Huetz et al., 2009), grasshoppers (Machens et al.2003), and zebra finches (Narayan et al., 2006; Billimoria et al., 2008), speech consonants presented to rats, (Engineer et al., 2008), speech phonemes presented to ferrets (Mesgarani et al., 2008), and even squirrel monkey twitter calls presented to ferrets (Schnupp et al., 2006; Walker et al., 2008). In contrast, few studies have shown that spike timing information contributes to the discrimination of static stimuli, such as tones that vary in location (Furukawa and Middlebrooks, 2002), or frequency (Moshitch et al., 2006).

Nevertheless, communication sounds possess complex dynamics, and encoding shallow modulation depths is important in processing them because modulation depth is particularly vulnerable to attenuation by the noise backgrounds and reverberant characteristics of ecologically relevant listening situations as distance from the sound source increases (Waser and Brown, 1986). Direct comparison to psychophysical thresholds is complicated by the fact that performance improves significantly with increasing carrier levels (Kohlrausch et al., 2000), while we collected data near the best level for each neuron, typically at low to moderate SPLs (30–60 dB). In rhesus subjects, Moody (1994) found average detection thresholds from roughly −12 to −20 dB (25 to 10% modulation) for 58 dB gated noise carriers modulated from 2.5 to 320 Hz. Human psychophysical thresholds for SAM tones like those used in this study were roughly −25 dB (5.6%) and −15 dB (17.8%) for 80 and 30 dB carriers, respectively (Moore and Glasberg, 2001). Because 10% was often the lowest modulation depth tested, the true thresholds for some neurons may be less than −20 dB, placing roughly a third of our data sample in the range of psychophysical detection thresholds. Moreover, typical cortical thresholds are apparently lower than those obtained in the central nucleus of the inferior colliculus in awake rabbits (Nelson and Carney, 2007, their Fig. 10), where modulation depths as low as 1.8% were tested. VS may overestimate modulation sensitivity, since VS produced lower thresholds than either TS or the spike train classifiers (compare Fig. 6a to 6b and 8a). It is unclear how VS could be estimated by the brain (Joris et al., 2004), and multiple methods can be used to infer psychometric thresholds from neural data (Scott et al., 2007; Walker et al., 2008). Nevertheless, VS-derived modulation gain in our data also

exceeds that reported for the auditory nerve (Joris and Yin, 1992), indicating that the ascending auditory pathway privileges sensitivity to low-frequency amplitude fluctuations.

In light of this sensitivity, it is interesting that Cazals et al. (1994) showed a strong correlation between speech intelligibility and the depth of the high-frequency (>50 Hz) rejection region in the MTFs of cochlear implant users. The neural representation of shallow modulation depths is also of interest because of the reduced dynamic range of electric hearing (Snyder et al., 2000; Geurts and Wouters, 2001; Middlebrooks, 2008), which may compress crucial features of speech into a narrow range of small depths (<10%). Middlebrooks (2008) observed a nonmonotonic dependence of VS on modulation depth in the auditory cortex of anesthetized guinea pigs when delivering SAM pulse trains via a cochlear implant. Our data strongly suggest that this finding is an artifact of electrical stimulation of the cochlea, and that modulation depth tuning is monotonic. Bieser and Müller-Preuss (1996) observed a monotonic increase in the PSTH and amplitude envelope correlation for increasing modulation depths. In contrast to their results, we also found a monotonic increase in firing rate with modulation depth, echoing Liang et al. (2002). The strength of this relationship depended on the character of the response to the unmodulated control at the chosen carrier level (Fig. 4d). Middlebrooks (2008) also observed roughly twofold changes in firing rate for modulation depths ranging from −25 (5.6%) to −15 dB (17.8%), which was similar to the dynamic range we observed for normal hearing over the full (0–100%) depth range.

Although the cortex represents low depth, low-frequency envelope information with similar sensitivity to the auditory nerve, the nature of the representation differs in important ways. In the auditory nerve, RLFs are relatively homogenous (Kim et al., 1990; Winter et al., 1990; Joris and Yin, 1992) and can be used as "input–output curves" for predicting the responses to SAM stimuli (Joris and Yin, 1992). Thus, the response modulation is proportional to the local slope of the RLF for a given modulation depth (Joris and Yin, 1992; Cooper et al., 1993). Cortical rate contrast ratios (Fig. 5c) are analogous to Cooper et al.'s (1993) modulated-rate predictions for auditory nerve fibers in the guinea pig, where predicted changes in firing rate over the modulation cycle were computed by taking one half (Yates, 1987) of the product of the decibel span (Fig. 1b) of the SAM stimulus and the slope of the RLF at the carrier level. In the auditory nerve, predictions based on the static RLF underestimate the response modulation (Yates, 1987), but far less so than in the cortex. Further, the RLF-derived predictions of cortical MPH profiles perform worse than predictions based simply on the envelope, or on the envelope prediction informed by the response class (e.g., driven, suppressed, transient) for equivalent duration tones at the carrier level (Fig. 10b).

The previous finding echoes earlier work which suggested that cortical representations of acoustic signals sacrifice the correspondence between firing rates and particular values of acoustic parameters (e.g., interaural phase differences) in favor of enhanced representations of stimulus changes (Malone et al., 2002). In the visual system, responses of retinal ganglion and more central cells are dominated by luminance contrast, rather than mean luminance. Our results suggest that a similar, albeit partial, transformation occurs in the ascending auditory pathway, where the relationship between instantaneous firing probability and absolute sound level becomes progressively less fixed, and the representation of changes in level are enhanced. Although tuning to mean level persists, as evidenced by cortical rate-level tuning,

responses to SAM suggest that adaptive mechanisms which modify and amplify the rate-level relationship predominate.

The decoupling of static and dynamic level coding is clearly related to the transition from sustained to transient responses to static stimuli like tone pips, and highlights the accumulation of response nonlinearities in central auditory structures. The engagement of these mechanisms may explain how gated tones at different levels produce PSTHs that can be distinguished even when average firing rate differences are removed (Fig. 10), and further evidences an enhanced cortical representation of envelope dynamics. Robust cortical sensitivity to small envelope fluctuations would seem counter to the argument that the increased cortical prevalence of nonmonotonic tuning to sound level and "closed" response areas contributes to a sound representation independent of level (Sadagopan and Wang, 2008). In fact, our data show that while the cortical representation of sound level dynamics is not independent of level (see also Malone et al., 2007), it is substantially less constrained by the RLF than in the periphery. Consequently, it is robust across a broad range of levels when spike timing information is considered. Analogously, Moshitch et al. (2006) found that global parameters defining the shape of tone-defined frequency response areas were not correlated with the information provided about tone identity by spike timing patterns. The fact that modulation detection thresholds are lower at higher carrier levels (Kohlrausch et al., 2000; Moore and Glasberg, 2001) seems paradoxical only when assuming that RLFs in the auditory periphery, which flatten at high levels, are the relevant input–output functions for the auditory system. Not only do RLFs change dramatically in the ascending auditory pathway (Semple and Kitzes, 1993; Phillips et al., 1994, 1995; Sadagopan and Wang, 2008), their predictive value for dynamic sounds is limited by adaptive mechanisms which ensure a high gain cortical representation of temporal amplitude contrast at all sound levels.

References

Bieser A, Müller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp Brain Res 108:273–284.

Billimoria CP, Kraus BJ, Narayan R, Maddox RK, Sen K (2008) Invariance and sensitivity to intensity in neural discrimination of natural sounds. J Neurosci 28:6304–6308.

Busby PA, Tong YC, Clark GM (1993) The perception of temporal modulations by cochlear implant patients. J Acoust Soc Am 94:124–131.

Cazals Y, Pelizzone M, Saudan O, Boex C (1994) Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants. J Acoust Soc Am 96:2048–2054.

Cohen YE, Theunissen F, Russ BE, Gill P (2007) Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. J Neurophysiol 97:1470–1484.

Cooper NP, Robertson D, Yates GK (1993) Cochlear nerve fiber responses to amplitude-modulated stimuli: variations with spontaneous rate and other response characteristics. J Neurophysiol 70:370–386.

Drullman R (1995) Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 97:585–592.

Drullman R, Festen JM, Plomp R (1994a) Effect of temporal envelope smearing of speech reception. J Acoust Soc Am 95:1053–1064.

Drullman R, Festen JM, Plomp R (1994b) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95:2670–2680.

Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. PLoS Comput Biol 5:e1000302.

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.

Foffani G, Moxon KA (2004) PSTH-based classification of sensory stimuli using ensembles of single neurons. J Neurosci Methods 135:107–120.

Fu QJ (2002) Temporal processing and speech recognition in cochlear implant users. Neuroreport 13:1635–1639.

Fu QJ, Shannon RV (2000) Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners. J Acoust Soc Am 107:589–597.

Furukawa S, Middlebrooks JC (2002) Cortical representation of auditory space: information-bearing features of spike patterns. J Neurophysiol 87:1749–1762.

Galvin JJ 3rd, Fu QJ (2005) Effects of stimulation rate, mode and level on modulation detection by cochlear implant users. J Assoc Res Otolaryngol 6:269–279.

Geurts L, Wouters J (2001) Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am 109:713–726.

Goldberg JM, Brown PB (1969) Response of binaural neurons of dog superior olivary complex to dichotic total stimuli: some physiological mechanisms of sound localization. J Neurophysiol 32:613–636.

Houtgast T, Steeneken HJM (1973) The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica 28:66–73.

Houtgast T, Steeneken HJM (1985) A review of the MTF-concept in room acoustics. J Acoust Soc Am 77:1069–1077.

Huetz C, Philibert B, Edeline JM (2009) A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. J Neurosci 29:334–350.

Joris PX, Yin TC (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. J Acoust Soc Am 91:215–232.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577.

Kay RH (1982) Hearing of modulation in sounds. Physiol Rev 62:894–975.

Kim DO, Sirianni JG, Chang SO (1990) Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power-spectrum. Hear Res 45:95–113.

Kohlrausch A, Fassel R, Dau T (2000) The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J Acoust Soc Am 108:723–734.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115–142.

Lee J, Bacon SP (1997) Amplitude modulation depth discrimination of a sinusoidal carrier: effect of stimulus duration. J Acoust Soc Am 101:3688–3693.

Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J Neurophysiol 87:2237–2261.

Machens CK, Schütze H, Franz A, Kolesnikova O, Stemmler MB, Ronacher B, Herz AV (2003) Single auditory neurons rapidly discriminate conspecific communication signals. Nat Neurosci 6:341–342.

Malone BJ, Scott BH, Semple MN (2002) Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques. J Neurosci 22:4625–4638.

Malone BJ, Scott BH, Semple MN (2007) Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. J Neurophysiol 98:1451–1474.

Mardia K, Jupp P (2000) Directional statistics. New York: Wiley.

Mesgarani N, David SV, Fritz JB, Shamma SA (2008) Phoneme representation and classification in primary auditory cortex. J Acoust Soc Am 123:899–909.

Middlebrooks JC (2008) Auditory cortex phase locking to amplitude-modulated cochlear implant pulse trains. J Neurophysiol 100:76–91.

Moody DB (1994) Detection and discrimination of amplitude-modulated signals by macaque monkeys. J Acoust Soc Am 95:3499–3510.

Moore BC, Glasberg BR (2001) Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. J Acoust Soc Am 110:1067–1073.

Moshitch D, Las L, Ulanovsky N, Bar-Yosef O, Nelken I (2006) Responses of neurons in primary auditory cortex (A1) to pure tones in the halothane-anesthetized cat. J Neurophysiol 95:3756–3769.

Narayan R, Graña G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. J Neurophysiol 96:252–258.

Nelson PC, Carney LH (2007) Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. J Neurophysiol 97:522–539.

Pfingst BE, O'Connor TA (1980) A vertical stereotaxic approach to auditory cortex in the unanesthetized monkey. J Neurosci Methods 2:33–45.

Phillips DP, Semple MN, Calford MB, Kitzes LM (1994) Level-dependent representation of stimulus frequency in cat primary auditory cortex. Exp Brain Res 102:210–226.

Phillips DP, Semple MN, Kitzes LM (1995) Factors shaping the tone level sensitivity of single neurons in posterior field of auditory cortex. J Neurophysiol 73:674–686.

Recanzone GH (2008) Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. J Neurosci 28:13184–13193.

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.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373.

Russ BE, Ackelson AL, Baker AE, Cohen YE (2008) Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99:87–95.

Sadagopan S, Wang X (2008) Level invariant representation of sounds by populations of neurons in primary auditory cortex. J Neurosci 28:3415–3426.

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.

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.

Scott BH, Malone BJ, Semple MN (2009) Representation of dynamic interaural phase difference in auditory cortex of awake rhesus macaques. J Neurophysiol 101:1781–1799.

Semple MN, Kitzes LM (1993) Focal selectivity for binaural sound pressure level in cat primary auditory cortex: two-way intensity network tuning. J Neurophysiol 69:462–473.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primary temporal cues. Science 270:303–304.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90.

Snyder RL, Vollmer M, Moore CM, Rebscher SJ, Leake PA, Beitel RE (2000) Responses of inferior colliculus neurons to amplitude-modulated intracochlear electrical pulses in deaf cats. J Neurophysiol 84:166–183.

Walker KM, Ahmed B, Schnupp JW (2008) Linking cortical spike pattern codes to auditory perception. J Cogn Neurosci 20:135–152.

Waser PM, Brown CH (1986) Habitat acoustics and primate communication. Am J Primatol 10:135–154.

Winter IM, Robertson D, Yates GK (1990) Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres. Hear Res 45:191–202.

Yates GK (1987) Dynamic effects in the input/output relationship of auditory nerve. Hear Res 27:221–230.

Zheng Y, Escabí MA (2008) Distinct roles for onset and sustained activity in the neuronal code for temporal periodicity and acoustic envelope shape. J Neurosci 28:14230–14244.

Zwicker E (1952) Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. Acustica Akust Beih 3:125–133.