

Brian J. Malone, Brian H. Scott and Malcolm N. Semple

J Neurophysiol 98:1451-1474, 2007. First published Jul 5, 2007; doi:10.1152/jn.01203.2006

You might find this additional information useful...

This article cites 75 articles, 31 of which you can access free at:

<http://jn.physiology.org/cgi/content/full/98/3/1451#BIBL>

Updated information and services including high-resolution figures, can be found at:

<http://jn.physiology.org/cgi/content/full/98/3/1451>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

This information is current as of March 6, 2008 .

Dynamic Amplitude Coding in the Auditory Cortex of Awake Rhesus Macaques

Brian J. Malone, Brian H. Scott, and Malcolm N. Semple

Center for Neural Science, New York University, New York, New York

Submitted 14 November 2006; accepted in final form 3 July 2007

Malone BJ, Scott BH, Semple MN. Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol* 98: 1451–1474, 2007. First published July 5, 2007; doi:10.1152/jn.01203.2006. In many animals, the information most important for processing communication sounds, including speech, consists of temporal envelope cues below ~20 Hz. Physiological studies, however, have typically emphasized the upper limits of modulation encoding. Responses to sinusoidal AM (SAM) are generally summarized by modulation transfer functions (MTFs), which emphasize tuning to modulation frequency rather than the representation of the instantaneous stimulus amplitude. Unfortunately, MTFs fail to capture important but nonlinear aspects of amplitude coding in the central auditory system. We focus on an alternative data representation, the modulation period histogram (MPH), which depicts the spike train folded on the modulation period of the SAM stimulus. At low modulation frequencies, the fluctuations of stimulus amplitude in decibels are robustly encoded by the cycle-by-cycle response dynamics evident in the MPH. We show that all of the parameters that define a SAM stimulus—carrier frequency, carrier level, modulation frequency, and modulation depth—are reflected in the shape of cortical MPHs. In many neurons that are nonmonotonically tuned for sound amplitude, the representation of modulation frequency is typically sacrificed to preserve the mapping between the instantaneous discharge rate and the instantaneous stimulus amplitude, resulting in two response modes per modulation cycle. This behavior, as well as the relatively poor tuning of cortical MTFs, suggests that auditory cortical neurons are not well suited for operating as a “modulation filterbank.” Instead, our results suggest that <20 Hz, the processing of modulated signals is better described as envelope shape discrimination rather than modulation frequency extraction.

INTRODUCTION

Our current understanding of the neural representation of communication sounds, including speech, would be enriched by examining the responses of cortical neurons to the relatively slow amplitude modulations (<20 Hz) that dominate speech signals and contribute most directly to their intelligibility (Drullman 1995; Fu and Shannon 2000; Houtgast and Steeneken 1973, 1985). Surprisingly little is known about the cortical representation of AM, in part because physiological studies using AM have typically concentrated on changes in modulation frequency rather than the changes in sound amplitude that define the modulation itself. Given this perspective, it is natural that many investigators have focused on the upper limits of modulation frequency coding, rather than how low frequency amplitude changes (e.g., <5 Hz) are encoded by cortical neurons. For example, it has long been known that cortical neurons are sensitive to the shape of modulation

waveforms (Swarbrick and Whitfield 1972), but the physiological basis of envelope shape discrimination has received relatively little attention. Although “envelope shape” has sometimes been used to distinguish between canonical modulation waveform shapes, such as square, sinusoidal, or triangular modulators, we intend the phrase to embrace all features of the time-varying envelope and thus differences in the depth of modulation or stimulus level.

The most commonly studied form of modulation, sinusoidal AM (SAM), has often been used to characterize the temporal aspects of the responses of central auditory neurons (Bieser and Muller-Preuss 1996; Creutzfeldt et al. 1980; Eggermont 1991, 1994; Frisina et al. 1990; Gaese and Ostwald 1995; Krishna and Semple 2000; Langner 1992; Langner and Schreiner 1988; Rees and Moller 1983, 1987; Schreiner and Urbas 1988). Four parameters are needed to specify a SAM stimulus: carrier frequency, carrier level, modulation frequency, and modulation depth. Historically, however, many studies of cortical responses to SAM have focused almost exclusively on modulation frequency because that is what cortical neurons were thought to encode. For example, it has been shown in awake macaque monkeys (Malone et al. 2000) and marmosets (Liang et al. 2002) that the modulation frequencies eliciting the strongest responses in primary auditory cortex (AI) are correlated when responses to SAM and sinusoidal FM (SFM) are compared. On this basis, Liang et al. (2002) concluded that “..it is the ‘temporal modulation’, and not the amplitude or FM *per se* that most auditory cortical neurons appear to extract from a complex acoustic environment (p. 2257).” While we agree that the correlation suggests that common temporal constraints influence how individual AI neurons respond to SAM and SFM, the broader conclusion that AI neurons respond to an abstracted “temporal modulation” is unwarranted.

Rees and Moller (1987) emphasized that neurons of the inferior colliculus “do not function as an array of stimulus invariant modulation frequency detectors” of the sort appropriate to a modulation filterbank, but rather “carry a selectively emphasized version of the input signal’s amplitude envelope which is modified by the prevailing stimulus conditions.” This notion gained support from the observation of Krishna and Semple (2000) that the best modulation frequency (BMF) of many neurons in the inferior colliculus (IC) changed substantially when the carrier level was varied over more than a 20-dB range. In contrast, Liang et al. (2002) reported that BMF was relatively invariant when carrier level and modulation depth were varied in AI. These competing views could be reconciled

Address for reprint requests and other correspondence: M. Semple, Center for Neural Science, New York University, New York, NY 10003 (E-mail: mal.semple@nyu.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

by evidence that successive levels in the auditory pathway are increasingly unaffected by changes in SAM parameters such as carrier level or modulation depth. Nevertheless, our results indicate that responses of central auditory neurons remain sensitive to SAM parameters other than modulation frequency, and are inadequately characterized by the modulation transfer function (MTF) describing how average firing rate and response synchrony vary with modulation frequency, and by derivative summary measures such as the BMF.

The most obvious difference between a neural code for modulation frequency and a neural code for changes in sound amplitude pertains to how cortical neurons represent changes in SAM stimuli that affect the stimulus amplitude but do not affect the modulation frequency, such as changes in carrier level or modulation depth. If cortical neurons code for stimulus amplitude, changes in such parameters should interact with an individual neuron's tuning for sound amplitude because that tuning is the basis for coding amplitude changes. For example, a unit's nonmonotonic tuning for sound pressure level (SPL) should be reflected in its response to SAM, even if such tuning introduces response components at modulation frequencies that are not present in the acoustic signal. Thus amplitude coding would compromise the modulation frequency code.

To reveal what AI neurons actually encode, we will concentrate on a data representation called the modulation period histogram (MPH), which shows the occurrence of action potentials relative to the period of the modulating stimulus waveform. It is common to collapse this distribution into summary measures such as vector strength and mean phase, which are plotted against modulation frequency in the modulation transfer function (MTF). By using detailed MPH examples from individual neurons, we will show that the responses of many cortical neurons cannot be adequately captured in this manner, particularly at low modulation frequencies. In this modulation range, it is reasonable to define the "instantaneous" SPL of the SAM stimulus, because the modulation period is long relative to the carrier period. Calculation of the "SPL profile" of the SAM stimulus allows for direct comparison of the amplitude envelope to the cycle-by-cycle response profile depicted by the MPH. Examination of the data in this format reveals that cortical responses to modulation frequencies in the range most important for communication sounds unequivocally encode changes in sound amplitude and are robustly sensitive to all parameters defining the SAM signal.

METHODS

Subjects, surgical preparation, and physiological recording

Two adult male monkeys (*Macaca mulatta*, designated X and Z) participated in these experiments. All procedures pertaining to animal use and welfare in this study were reviewed and approved by the New York University Institutional Animal Care and Use Committee. Before implant surgery, anesthesia was induced with ketamine and sodium thiopental, and a surgical plane was maintained with isoflurane. This first implant was a head-holder that mated to a specially designed primate chair (Crist Instruments, Hagerstown, MD). After behavioral training, a recording chamber (CalTech Engineering Services, Pasadena, CA) was implanted above the auditory cortex in the left hemisphere of each animal. The initial placement of the recording chamber on monkey Z was slightly rostral to allow recordings across the rostral (R) and rostrotemporal (RT) fields (Hackett et al. 1998). The back of the initial chamber and the front of the chamber in its

second placement straddled the low-frequency portion of primary AI. On completion of the mapping of the left hemisphere, the recording chamber was removed, and the skull was permitted to regrow under a protective layer of acrylic (Palacos). Meanwhile, a new recording chamber was implanted above the putative location of field R on the right hemisphere, which allowed for limited access to AI caudally. The initial implant for animal X was centered over AI in the left hemisphere, and allowed for a complete mapping of AI and portions of the surrounding auditory cortex. When this site was completed and covered, a new recording chamber was centered on the putative low-frequency border of AI/R in the right hemisphere.

All penetrations were made vertically with respect to the cylinder implants and thus roughly parallel to the stereotaxic vertical plane. Animal Z is still involved in experiments, so assignment of recording locations to cortical fields is based on physiological criteria, such as the tonotopic progression in AI and the distribution of response latencies (Scott et al. 2000). Subsequent histology and postmortem magnetic resonance imaging in animal X confirmed the recording locations to be within primary auditory cortex. We also assigned a relative cortical depth to the neurons in our sample by normalizing the recording depth with respect to the first and last points in each penetration where audible "hash" responses could be detected ($n = 270$). Expressed in quintiles from the shallowest to deepest depths, we obtained the following distribution: 19, 24, 27, 20, and 10%. Although we cannot unequivocally assign our recordings to particular laminae, it is likely that the neurons in our sample came predominantly from the middle and upper layers.

Both animals were extensively trained on binaural lateralization tasks. During recordings, blocks of psychophysical trials alternated with passive listening, when the SAM stimuli described in this report were presented. Behavioral and recording sessions were all conducted in a double-walled sound attenuated chamber (Industrial Acoustics) while the animals were continuously monitored using closed circuit television. Single-unit activity was recorded with tungsten microelectrodes (FHC, Bowdoin, MA) advanced into the brain through a stepping motor microdrive (CalTech Engineering Services, Pasadena, CA). Recording location was referenced to a stereotaxic positioning system that mounted directly on the implant. Depths of all recordings were referenced to entry into the brain. Entry into the superior temporal plane was typically marked by a sudden increase in activity after a long silent interval and the first appearance of auditory responsiveness.

Stimulus generation and data acquisition

Stimulus waveforms were generated by digital synthesizers and custom hardware (MALab, Kaiser Instruments). Stimulus characteristics were specified in software running on the host computer (Macintosh), which communicated with a dedicated microprocessor (MALab) using an IEEE-488 interface. After digital attenuation and D/A conversion, the signal was transduced by electrostatic earphones (STAX Lambda) in custom housings (Custom Sound Systems) fitted to ear inserts. Before each experiment, the SPL expressed in decibels (re: 20 μ Pa) at each ear was calibrated under computer control for level and phase from 40 Hz to 30 kHz, using a previously calibrated probe tube and condenser microphone (Brüel and Kjær 4134).

Electrical signals from the brain 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). The occurrence of discriminated action potentials and stimulus synchronization events were logged with a resolution of 1 μ s and stored by the host computer for analysis and display.

Stimulus protocols

All stimuli described in this report were gated on and off by a cosine-squared ramp (10 ms). Responsive AI neurons were initially

characterized with a battery of pure tone stimuli of relatively short duration (typically 100 ms, but occasionally 200 ms). These tonal stimuli were used to determine the frequency tuning function at the best sound level (dB SPL), and the rate-level function at the neuron's best frequency. SAM stimuli were typically presented at best frequency and level. In cases where there was no clear best level because of saturating responses in the range of moderate SPLs, 60 dB SPL was used, which provided a consistent carrier level for the generation of the composite modulation period histograms (Fig. 15). This was the most common carrier level used (particularly in animal X) and was used in 178 (49%) cells. SAM stimuli used in this study were typically presented in two consecutive trials of 10 s, separated by a 2-s interstimulus interval. Long stimulus durations were chosen to minimize the effects of onset responses while maximizing the number of modulation periods. This choice was crucial for the low modulation frequencies emphasized in this study. We verified that the effects of onset responses were negligible by recalculating spike counts and spike timing metrics for a subset of the data sample ($n = 124$). The common practice of excluding responses during the first 100 ms (1% of the stimulus duration) does not significantly impact the distributions of the spike timing metrics described below (Wilcoxon ranked sum, $P < 0.7$). In fact, this correction, which affected 7 of 333 spikes on average, cannot be resolved when the distributions of the spike counts themselves are compared ($P > 0.4$). Because stimulus runs generally included an unmodulated control tone of similar duration (10 s), we were also able to compare "sustained" responses elicited by modulated stimuli to the adaptation characteristics of each neuron. In general, cortical neurons continued to respond robustly throughout the duration of SAM stimuli, such that firing rates showed average decreases of roughly 10% from the first to second half of the stimuli (0–5 vs. 5–10 s) and average decreases of roughly 20% from the first to fourth quarter (0–2.5 vs. 7.5–10 s). The distributions of temporal measures such as vector strength and trial similarity were not significantly altered from the first to second halves of the stimuli (Wilcoxon ranked sum, $P > 0.2$), nor from the first to fourth quarters ($P > 0.1$).

The SAM stimuli consisted of a sinusoidal carrier tone (f_c) modulated sinusoidally by a second tone (f_m) such that $s(t) = A[1 + m \cdot \sin(2 \cdot \pi \cdot f_m \cdot t)] \sin(2 \cdot \pi \cdot f_c \cdot t)$. Because the carrier frequency was always much greater than the modulating frequency (i.e., $f_c \gg f_m$), the bracketed term defined the time-varying amplitude of the stimulus. The overall amplitude of the stimulus is set by A , and the depth of the modulation is determined by m , which varies from 0 to 1 (often described as 0–100% modulation). We shall refer to the sound level of a SAM stimulus (e.g., 60 dB SPL) in terms of its carrier SPL (A) for simplicity, although the actual sound level will vary over the course of the modulation period when $m > 0$, as explained below. Nearly all neurons in the sample were presented with fully (100%) modulated SAM signals at 0.7, 1, 2, 5, 10, 20, 50, 100, and 200 Hz. If the cell exhibited a synchronized discharge to 100 Hz, additional frequencies at modulation frequencies $\leq 1,000$ Hz were presented in steps of 100 Hz. In some cases, intermediate values (e.g., 3 Hz) were also chosen to get more precise estimates of the slopes of the MTF. In many cases, tuning to modulation depth was explored at a range of depths (typically, 0–100% in 10 or 20% steps). Because of the limited recording time available, variations in all four SAM parameters could not feasibly be presented in all neurons. Consequently, runs that varied carrier level and frequency were performed somewhat less frequently, generally in those cells where the isolation was particularly stable and the responses were particularly robust.

To reference the SAM stimulus to the decibel scale used to measure the rate level function for each neuron, we computed the "instantaneous" relative amplitude of SAM signals (in dB) by taking the logarithm of the envelope, in decibels: $20 \cdot \log\{[1 + m \cdot \sin(2 \cdot \pi \cdot f_m \cdot t)]\}$. A family of curves describing the instantaneous amplitude of the SAM signal relative to an unmodulated carrier signal at various modulation depths is shown in Fig. 2A. With increasing modulation depth, these curves become less sinusoidal, and the falling and rising

phases of the envelope become more prominent. There is also an asymmetry in the increases and decreases of SPL within each modulation cycle. When $m = 0.1$, the SAM signal increases 0.83 dB (0°) and decreases -0.92 dB (180°) relative the carrier. For $m = 0.9$, these values are 5.5 and -20 dB, respectively. Thus the changes in sound level (dB) for large modulation depths are dominated by the rapid fall and rise of the envelope within a relatively small portion of the modulation cycle centered on 180° . For the low modulation frequencies considered in this study, where the envelope is well defined, it is possible to generate estimates of the instantaneous SPL of SAM signals by adding the relative amplitude to the carrier SPL, as shown for 100% modulated signals at various carrier SPLs in Fig. 3B. For example, a 60-dB SPL SAM signal presented at a modulation depth of 90% ($m = 0.9$) varies from 65.5 dB SPL to 40 dB SPL during each modulation cycle.

Data analysis

The dominant representation of cortical responses to SAM in this paper is the MPH, which shows the distribution of spike counts for the different phases of the modulation cycle. The MPH is constructed by folding the peristimulus time histogram (PSTH; Fig. 1B) around the modulation period. The modulation period is inversely related to the modulation frequency (e.g., f_m of 2 Hz results in a modulation period of 500 ms). Both the carrier and modulation waveforms were presented in sine phase, resulting in the MPH shown in Fig. 1C. To facilitate interpretation of shapes of the MPHs, however, the responses were shifted by 90° (Fig. 1D), so that responses to the most dramatic changes in stimulus amplitude (instantaneous SPL) are centered in the MPH representation.

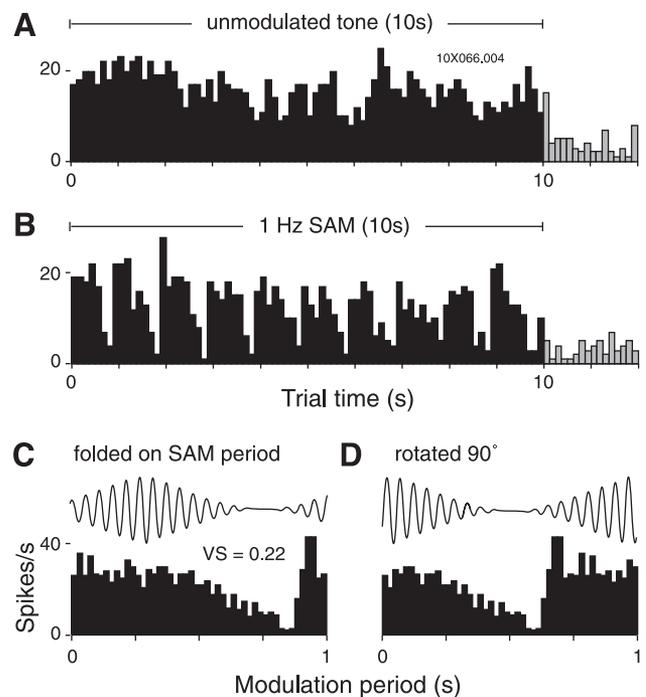


FIG. 1. Conventions for the display of sinusoidal AM (SAM) data. *A*: peristimulus time histogram (PSTH) of the responses to unmodulated control tone. *B*: PSTH of responses to a fully (100%) modulated SAM stimulus presented in sine phase at 1 Hz. *C*: modulation period histogram (MPH) representation of response shown in *B*, folded on duration of modulation period (1 s). *D*: MPH shown in *C* has been rotated by 90° , so that the center of MPH defines transition from decreases in sound amplitude to increases in sound amplitude. Thus instantaneous level minimum always occurs in the center of MPH, and maximum occurs at edges. This adjustment has been made for all subsequent MPHs depicted in later figures.

Calculations of spike rate were based on the entire stimulus duration. To evaluate the significance of differences in average firing rate, responses were averaged across repeated trials and binned in 1-s epochs. The average firing rate was calculated for each epoch. Calculation of spontaneous rates was based on firing rates for 1-s epochs drawn from all interstimulus intervals in a given stimulus run. Significance was assigned for all comparisons according to the outcome of a heteroschedastic *t*-test ($P < 0.01$).

To allow for comparison with other studies, response synchronization at the modulation frequency was quantified in terms of vector strength (VS) (Goldberg and Brown 1969). Each spike is treated as a unit vector whose angle corresponds to the phase at which it occurred in the modulation cycle. These unit vectors are summed to produce a resultant vector whose length corresponds to the magnitude of the Fourier component of the response at the modulation frequency. Normalizing the resultant vector by the total number of spikes (n) results in the VS, which is bounded from 0 (e.g., spike counts are equal at all phases) to 1 (all spikes occur at the same phase). The direction of the resultant vector indicates the mean phase of the MPH. Computationally, VS is calculated from the MPH as follows: $VS = [(r_i \cos_i)^2 + (r_i \sin_i)^2]^{0.5}$, where r is the spike count in the i th bin of the MPH. To assess the statistical significance of the VS, the Rayleigh statistic ($2VS^2n$) was computed, and values >13.8 (Mardia and Jupp 2000) were considered to be significant ($P < 0.001$). The synchrony cut-off was considered to be the highest tested modulation frequency that resulted in a Rayleigh statistic >13.8 . Additional details concerning the problems with applying the VS metric to cortical responses are presented in RESULTS.

As an alternative to VS, we introduce a spike timing index, trial similarity (TS), based on the correlation (i.e., the Pearson's correlation, or product-moment coefficient of correlation) between MPHs constructed independently from trial 1 and trial 2 of the SAM stimulus presentation. All spikes occurring within the stimulus duration were included in the analysis. Although useful alternatives to VS have been proposed (Joris et al. 2006; Kajikawa and Hackett 2005), we have chosen to analyze TS because of its simplicity in the context of data collected in two long duration trials. Nevertheless, one could in principle correlate MPHs generated from data distributed across any number of trials, provided that the total stimulus duration was comparable and the conditions for generating the MPHs are properly met. For example, one could generate the first MPH from odd numbered trials and the second from even numbered trials.

Unlike the VS metric, which measures how densely spikes are clustered around a single phase of the MPH, the TS index depends only on the reproducibility of the MPH shapes across trials. Whereas VS measures the synchrony of the neural response, TS measures its fidelity. To calculate the correlation between the MPHs obtained for each stimulus trial, it is first necessary to choose the numbers of bins that comprise the MPHs, and the value of TS will depend on the number of bins used for the correlation. Empirically, MPHs based on 52 bins adequately capture the temporal features of the neural responses, and MTFs based on TS show similar high-frequency cut-offs to those based on VS. The significance criterion for the TS metric is the likelihood that a given correlation coefficient could have been produced by chance (i.e., for 2 random spike trains). We created significance criteria by simulating thousands of pairs of random spike trains and calculating TS across a range of binwidths and spike counts. For the 52 bin MPHs used in this study, TS values of 0.4 and 0.6 correspond conservatively to P values of 0.001 and 0.0001, respectively. Note that it is possible for TS to be negative. Because we never observed a case where TS was significantly negative by the criteria above, however, negative values were simply set to zero.

Stimulus estimation and spike train classification

To determine how much information cortical spike trains provided about stimulus identity, we used a PSTH-based pattern classifier to

estimate the stimulus on the basis of 1 s of data (for complete details of the method, see Foffani and Moxon 2004). Note that only stimuli whose modulation frequency was an integral multiple of 1 Hz were analyzed in this way. For each stimulus in a given set (i.e., the stimuli comprising a modulation depth function, modulation transfer function, or carrier level function), a "template," representing the average response to that stimulus, was formed by folding the response at 1-s intervals and binning the responses into a bin-dimensional vector. The average spike count per bin was obtained by dividing by the number of seconds of data (i.e., 20), unless the template contained the data epoch to be matched—the "test." In such a case, the test was subtracted from the appropriate PSTH before binning and calculating the average spike count per bin (i.e., dividing by 19). This form of classification is referred to as "complete cross-validation" (Foffani and Moxon 2004). The test was binned similarly.

Each of the 20 tests per stimulus was matched to the template that minimizes the Euclidean distance between the test vector and template vectors. The results of this matching process are stored in a confusion matrix whose columns represent the stimulus that was actually present and whose rows represent the estimate of stimulus identity produced by the classifier. If every test epoch is correctly associated with the stimulus that elicited it, all values of the confusion matrix along the diagonal will be 20 and all off-diagonal entries will be 0. Percent correct for a given stimulus set is obtained by summing along the diagonal and dividing by the product of data epochs (20, in every case we analyzed) and the number of stimuli in the set (e.g., 8, for our typical MTF consisting of responses to 1, 2, 5, 10, 20, 50, 100, and 200 Hz SAM).

The method described above also allows us to parse the contributions of spike timing and spike rate information to the performance of the classifier. For example, the size of the binning applied to the tests and templates will impact the performance of the classifier because it determines the amount of temporal detail available to it. To capture this aspect of its performance, we generated complete confusion matrices for bins 1, 2, 4, 8, 10, 20, 40, and 1,000 ms wide for each stimulus set. Given the use of 1-s data epochs and PSTHs, the inclusion of a single, 1,000-ms-wide bin in the analysis allows us to determine how successfully the stimulus can be estimated based on the spike rate alone. Conversely, it is also possible to eliminate information pertaining to the distribution of firing rates across stimuli by normalizing both the tests and the templates by their respective vector norms. Geometrically, this corresponds to mapping all tests and templates to a hypersurface located at a unit distance from the origin. Subsequent to this normalization, the only information retained by the test and template vectors is the relative distribution of spike probability within a 1-s window. Because spike phase information is retained, we refer to this as the "phase only" classifier. Normalization by the total spike count, rather than the vector norm, produces essentially identical results: the correlation coefficients for classifier performance, in percent correct, across the different normalization schemes were 0.96, 0.99, and 0.97 for modulation depth, modulation frequency, and carrier level, respectively.

To assess whether the classifier performance was significantly better than would be expected by chance, we simulated confusion matrices based on random draws from a given stimulus set over many (10,000) iterations and generated a distribution of percentage correct based on the bootstrap results. If classifier performance exceeded all bootstrapped values, it was considered to be significant ($P < 0.0001$). Because the typical number of stimulus set elements was not constant across stimulus type (e.g., 8 for modulation frequency vs. 11 for modulation depth), it is not possible to compare classifier performance across different stimulus types directly because the baseline for chance performance varies inversely with stimulus set size. To circumvent this limitation, we standardized classifier performance as a z -score with respect to the appropriate bootstrap distribution by dividing the difference between the actual classifier result and the bootstrap mean by the bootstrap SD.

Although similar classifiers are often applied to stimulus sets that vary categorically (e.g., a set of vocalizations), SAM parameters such as modulation depth varied monotonically. Percentage correct is insensitive to the relative quality of the classifier estimate—for a SAM stimulus modulated at 40% depth, an estimate of 30% is the same as an estimate of 0 or 100%, i.e., a miss. We accounted for the relative quality of the classifier estimates by assigning a cost to each estimate, so that values along the diagonal of the confusion matrix equal zero, and off-diagonal values are multiplied by the distance to the diagonal in each column. Thus for a 40% depth modulation in an 11 by 11 confusion matrix spanning 0 to 100% modulation in 10% intervals, an estimate of 30% entails a cost of 1, whereas 0% has a cost of 4, and 100% has a cost of 6. Significance is assessed by the bootstrap method described above, except the distribution simulated was based on total cost, summed over the confusion matrix, rather than percentage correct. For population comparisons, the total cost for classifier performance on a given stimulus set was normalized by the theoretical maximum cost for a confusion matrix of equivalent size, producing a cost index from 0 (perfect) to 1. In practice, the additional sensitivity provided by the cost index, relative to percentage correct, proved unnecessary for modulation frequency and carrier level because classifier performance was particularly strong in such cases.

RESULTS

Summary of the data sample

We will describe data obtained from the responses of 361 neurons tested with SAM stimuli. These data represent a subset of an extensive physiological survey of auditory cortex focused on A1, but perhaps including a few neurons on the borders of adjacent fields. To allow for response class categorization, only recordings that included responses to the unmodulated control tone were included in the data sample. Because we could not detect any obvious differences in the responses from either animal (X or Z) or hemisphere (left or right), we combined the data from four hemispheres in two animal subjects (see METHODS).

Carrier frequencies varied from 0.1 to 32 kHz. At least 30 cells were characterized for each octave with respect to 0.5 kHz (i.e., <0.5, 0.5–1, 1–2 kHz, etc.). The SAM stimulus was presented binaurally in most cases because it elicited more robust responses than monaural stimulation (binaural summation was much more common than suppression). Modulation transfer functions were typically generated at the neuron's best frequency and level. If the rate-level function exhibited a plateau of similar responses that included 60 dB SPL, we used that value as the carrier level. Carrier levels ranged from –10 to 90 dB SPL, but roughly one half of the neurons were tested with a carrier level of 60 dB SPL.

SAM stimuli were very effective for AI neurons in awake rhesus macaques. A neuron was considered to be responsive to SAM if it exhibited either a significantly synchronized response to at least one modulation frequency or a significantly different firing rate from the response to the unmodulated control tone for at least one modulation frequency (see METHODS). By this criterion, only 6 of 361 (0.6%) neurons were considered to be unresponsive to 100% modulated SAM signals presented at the cell's best carrier frequency and level (or 60 dB SPL). We typically did not record SAM responses for cells that had been deemed generally unresponsive during initial testing with tonal stimuli, so this result likely overestimates the prevalence of SAM responsiveness in AI. Neverthe-

less, we attempted to obtain an MTF for all neurons exhibiting robust responses to pure tones, so the data sample can be considered representative of such neurons.

In addition to characterizing cortical responses to tones of short duration (typically 100 ms), we also measured responses to an unmodulated tone of the same duration as the modulated stimuli (10 s). The response to a pure tone of the same carrier frequency and level as the SAM stimuli served as a reference for the responses to modulated tones (Fig. 1A). A striking aspect of cortical responses in awake animals is the fact that roughly one third (113/361; 31%) had significantly elevated firing rates relative to the spontaneous rate when calculated over the duration (10 s) of the control tone. In an additional 12% (42/361) of neurons, the firing rate was significantly suppressed over the duration of the control tone. Thus the generally accepted notion, derived from studies of anesthetized animals, that cortical neurons do not give sustained responses to pure tone stimuli of long duration (see Middlebrooks 2005) does not hold for nearly one half (43%) of our data sample (Malone et al. 2002; Wang et al. 2005). It should be noted that if mechanisms of adaptation act on long time scales (Malone et al. 2002; Ulanovsky et al. 2004), the very long duration (10 s) of the control tones makes our estimate of the prevalence of sustained responses more conservative than a response classification based on shorter stimuli would likely be.

Construction of the MPH

The MPH represents the occurrence of action potentials relative to the phase of the modulation cycle. Figure 1 shows the construction of the MPH and the conventions for its display. The periodic modulation of spike rate for a 1-Hz SAM stimulus (modulation depth = 100%, or $m = 1$) is clearly evident in the PSTH representation shown in Fig. 1B. By folding the responses on the modulation period (1 s), the distribution of discharge rates within the modulation period is more easily seen in the MPH representation (Fig. 1C). The stimulus envelope was presented in sine phase, but in Fig. 1D (and in all subsequent MPHs), we shifted the responses by 90° (cosine phase) to facilitate interpretation of the shape of MPH with respect to the instantaneous stimulus amplitude. In this representation, the instantaneous amplitude minimum (270°) of the stimulus occurs in the middle of the MPH, and the maximum (90°) occurs at its lateral extremes, which occur at neighboring phases because the modulation phase axis is cyclic. Inspection of Fig. 1D revealed that, for this neuron, the MPH response profile differs from the sinusoidal amplitude envelope suggested by the cartoon of the carrier waveform shown above it. Within each cycle, the instantaneous probability of discharge declines rather slowly as the instantaneous level declines, but the response rises abruptly when the amplitude increases from its minimum. This behavior was typical of neurons that exhibited sustained responses to the unmodulated control tone (Fig. 1A).

The fact that the response shown in Fig. 1 is responsive throughout most of the modulation period results in a relatively low VS (0.22). Nevertheless, the shape of the MPH response profile was extremely robust. The shapes of the MPHs obtained for separate trials were strongly correlated (TS = 0.83). This suggests that, despite the low VS, the fidelity of the cortical response is quite good because the response profile is a con-

sistent, albeit transformed, representation of the modulated signal.

Changes in modulation depth are reflected in the MPH response profile

The nature of the cortical representation of slow amplitude changes will be addressed in a number of examples that compare the MPH response profile to the instantaneous amplitude of a SAM stimulus, calculated by taking the ratio of the SAM stimulus to the unmodulated control and converting this ratio into decibels (see METHODS). By adding this value to the carrier level, one can construct what we term the ‘‘SPL profile’’ of the SAM stimulus, which specifies the approximate sound pressure level (in dB re: 20 μ Pa), of the modulated waveform at each phase of the modulation period. This conversion facilitates comparisons between the SPL profile of the stimulus, the MPH response profile, and the neuron’s rate-level tuning function, which was also measured in terms of SPL.

Although the stimulus envelope of SAM is sinusoidal by definition, the log-transformed SPL profiles depicted in Fig. 2A are not. Deviations from a sinusoidal profile in decibels increase with increasing modulation depth, as shown. Not only are the increases and decreases in SPL asymmetric about the nominal carrier SPL, but the range of relative amplitudes

spanned by SAM also increases nonlinearly with increasing modulation depth. Although small modulation depths sample a relatively narrow range of actual stimulus levels, the response depicted in Fig. 2 is clearly modulated for $m = 0.2$ (+1.58/–1.93 dB), and increasing depths result in increasingly robust modulation. Note that in all cases this neuron responds to the decrease in amplitude (see Fig. 2B stimulus icons, in gray) with an increase in firing rate, as would be predicted from the fact that the carrier level (60 dB SPL) is higher than this strongly nonmonotonic unit’s best level (20 dB SPL; Fig. 2C, inset). For $m = 0.9$, a second response mode corresponding to the rising phase of the stimulus envelope begins to emerge, as the SPL profile encompasses progressively lower SPLs and the slope of its rising phase sharpens. The emergence of an ‘‘onset’’ peak at large modulation depths is consistent with the onset response this unit displayed for short-duration pure tones (Fig. 2C). This type of change in the response profile as modulation depth increased to the top of its range (0.8 to 1) was common in our sample. When the modulation is of sufficient depth, the SPL profile sweeps through the neuron’s preferred range of sound levels twice: once during the falling phase and again during the rising phase. Because these features of the SPL profile occur at neighboring phases of the modulation cycle, the resultant peaks are closely apposed, as shown in Fig. 2B. Although such changes in the shape of the MPH reduce the VS, the TS metric

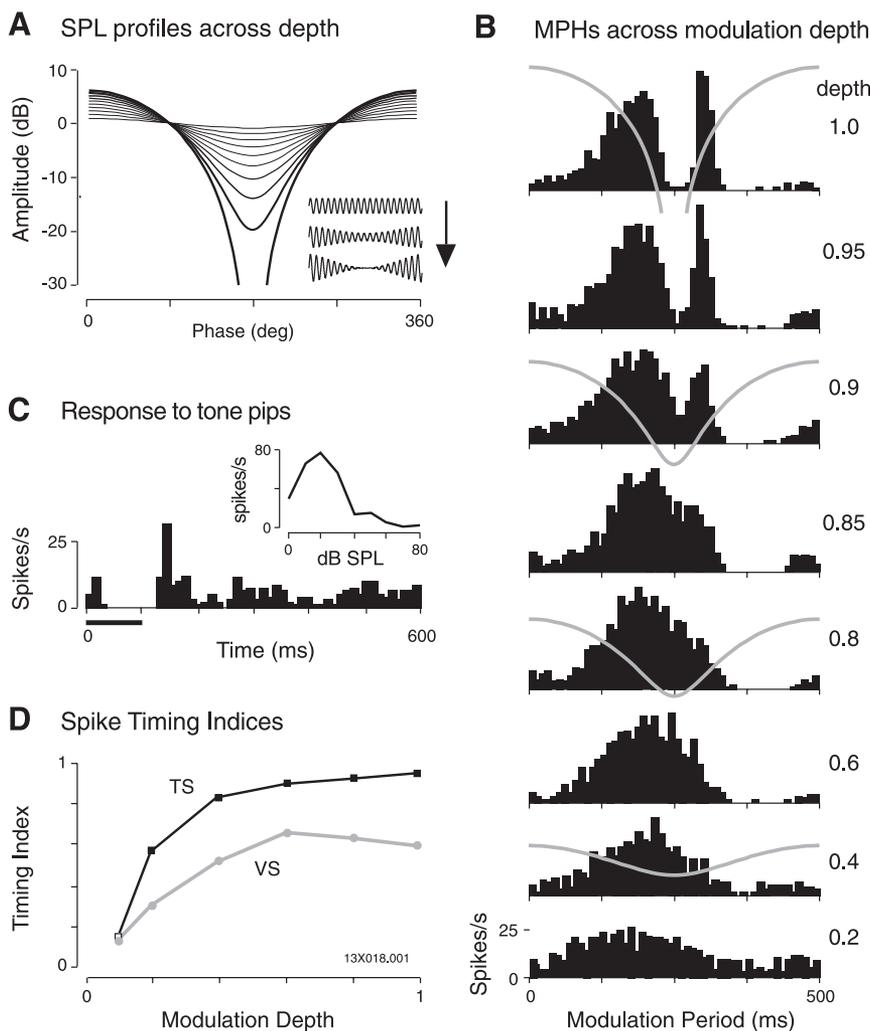


FIG. 2. Changes in modulation depth are reflected in changes in the MPH response profiles. *A*: family of curves depict the SPL profiles, which are the instantaneous levels of the SAM stimulus, in decibels with respect to carrier level, as a function of phase. Each curve represents a different modulation depth from 0.1 to 1 in 0.1 steps. Icons of SAM time waveforms indicate effect of increasing modulation depth. Rotated phase is indicated on abscissa on this and subsequent figures. *B*: series of MPHs depict changes in response profile for increasing modulation depths (0.2 to 1). Icons representing shape of corresponding SPL profiles are overlaid in gray. Stimulus was 2-Hz SAM presented at a carrier frequency of 600 Hz and a carrier level of 60 dB. *C*: PSTH of response to 100-ms tones at carrier frequency and level shows that this sound amplitude was predominantly suppressive and resulted in a sizable offset response. *Inset*: rate level function for this neuron, which peaked at 20 dB. *D*: curves indicate values of vector strength (gray line) and trial similarity (TS; black line) across modulation depth. In this and subsequent figures, significant ($P < 0.001$) values for vector strength (VS) and TS are indicated by filled symbols.

continues to increase, indicating that this feature of the neuron's response was highly reliable (Fig. 2D). The relationship between the response profile and the SPL profile for this cell suggests that the rate-level function (RLF) can sometimes serve as a useful heuristic for predicting the neuron's responses to changes in amplitude. Although the presence of two modes in the MPH accurately conveys information about changes in SPL, it necessarily confounds a simple periodicity code for modulation frequency.

Changes in carrier level profoundly affect cortical response profiles

If cortical neurons are primarily sensitive to the amplitude of SAM signals, we would expect that changes in carrier level would profoundly change the shapes of the MPH, particularly in neurons that are sharply tuned for SPL. For example, nonmonotonic tuning for SPL was common in AI in these animals, with 38% of neurons exhibiting decreases in firing rate >50% (relative to best level) for increasing SPLs, 15% exhibiting milder (<50%) decreases, 7% exhibiting a saturating plateau, and 31% responding strictly monotonically (Scott 2004). The range of SPLs spanned by a SAM stimulus is jointly determined by the carrier level and the modulation depth. In the case of 100% modulated ($m = 1$) signals, however, the carrier level effectively determines only the instantaneous SPL maximum, because all signals are briefly "off" for a portion of the modulation cycle (180°). Figure 3A depicts the way that changes in carrier level cause vertical displacements of the signal on the SPL axis.

Figure 3C indicates that the responses of an example neuron to brief duration tones were strongly nonmonotonic, with a best level of 20 dB SPL. As the carrier level for 1-Hz modulation was increased from 10 to 70 dB, the shapes of the MPHs changed profoundly. From 20 to 40 dB, the MPH trough corresponding to the lowest instantaneous SPLs progressively narrowed, as a greater proportion of the modulation cycle consisted of moderate SPLs. As the carrier level increased further, the cell responded to decreases in SPL with increases in firing rate, resulting in a response peak centered between 90 and 180°. These features of the MPHs are compatible with the fact that this neuron responded in a sustained fashion to the long duration control tone at 50 dB and showed offset responses for tone pips at higher SPLs.

The set of response profiles in Fig. 3B suggests that the cell fired robustly when the instantaneous SPL falls within the range of its preferred SPLs. However, the SAM-derived RLF (Fig. 3C) deviates substantially from the RLF defined with 100-ms tone pips. The SAM-derived RLF is shifted to higher SPLs and is substantially less nonmonotonic. Apparently, the two response peaks elicited by rapid transitions through the neuron's preferred SPL range offset the fact that high carrier level stimuli spend a smaller fraction of the modulation period in that range. Because the shape of the SPL profile is effectively constant across carrier level, response profile changes as carrier level is varied can only be explained by the interaction of the SPL profile and the neuron's tuning for SPL. The divergence of the SAM-derived RLF and tone pip RLF indicates that the neuron is also sensitive to the dynamics of the amplitude changes.

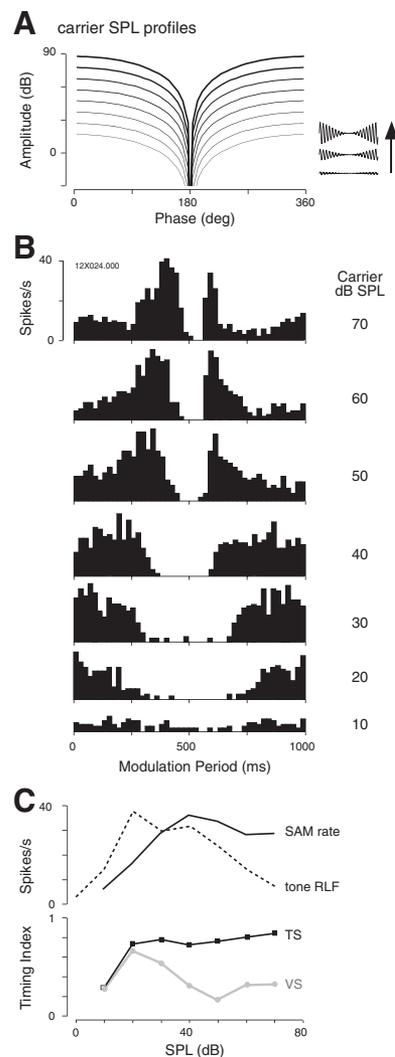


FIG. 3. Changes in carrier level are reflected in changes in MPH response profiles. *A*: family of curves depicts SPL profiles for 100% modulated SAM stimuli with carrier levels from 10 to 80 dB. Icons of SAM time waveforms indicate effect of increasing carrier level. *B*: series of MPHs depict changes in response profiles for increasing carrier levels (10–70 dB). *C*: curves in *top panel* compare firing rates obtained with SAM (solid line), and for 100-ms tone pips (dashed line) as carrier level was varied. Curves in *bottom panel* display changes in VS (gray line) and TS (black line) measured for responses to SAM across carrier level (filled symbols indicate significance for timing indices at $P < 0.001$).

Here again, the appearance of novel response features in the response profiles at high carrier levels shows that this neuron is capable of coding more than the modulation frequency of the SAM stimulus. These response features degrade synchrony to the modulation frequency (Fig. 3C). The progressive narrowing of the trough in the response profile from 20 to 50 dB causes a progressive reduction in VS. From the perspective of SPL coding, however, the narrowing of the trough in the response profile is readily explained in terms of the neuron's level tuning, and implies that the fidelity of the amplitude representation is maintained. Inspection of the TS curve, which increases with carrier level, confirms this impression, indicating that the neural representation of the SPL profile is more rather than less robustly coded at high carrier levels.

Changes in carrier frequency are captured by changes in cortical response profiles

Unlike the carrier level, modulation depth, and modulation frequency, the carrier frequency of the SAM stimulus is not directly tied to the SPL profile. For this reason, we presented SAM stimuli at each neuron's best frequency and rarely varied this parameter systematically. Nevertheless, the fact that a neuron's response area is a joint function of frequency and level implies that the choice of carrier frequency should impact the shape of the MPH for SAM. Anecdotally, we have observed that changes in the responses to tone pips at different frequencies (e.g., a change from predominantly onset to offset responses) are typically mirrored in analogous changes in the MPH response profiles, including predictable changes in the phase of the dominant response peak.

Figure 4 shows how changes in carrier frequency (6, 10, and 16 kHz) were reflected in the responses of an example neuron. At all carrier frequencies, the neuron fired most strongly at the SPL profile minima (Fig. 4A; note that as the modulation period shortens, a fixed neural latency will cause increasing apparent lag in the response peak relative to the SPL mini-

mum). Overall, the 6-kHz carrier elicited the highest average firing rates, and the firing rate varied relatively little with modulation frequency (Fig. 4B). Increasing the carrier frequency resulted in more variable rMTFs, including suppression relative to the control tone for the 16-kHz carrier at modulation rates >50 Hz. The impression conveyed by the tMTFs based on VS (Fig. 4B) is somewhat more complicated, because the VS tMTF appears band-pass at 10 kHz, but not at 6 or 16 kHz. Study of the MPHs revealed that responses to the 10-kHz carrier are weakly bimodal, which explains the reduction in VS below 20 Hz. Although the differences between the MPH shapes for the 10- and 16-kHz carriers are subtle, they are highly reproducible, as indicated by the uniformly high TS values from 2 to 20 Hz. Thus MPH response profiles contain information that can be used to distinguish the carrier frequencies of SAM stimuli.

Changes in modulation frequency are captured by changes in cortical response profiles

As modulation frequency is increased, the changes in amplitude indicated by the SPL profile occur more and more rapidly. The resultant changes in the response profile reveal temporal constraints on the coding of those changes operant in the recorded cell and its input pathway. Figure 5B depicts response profiles for 100% modulated stimuli ($m = 1$) spanning a range from 1 to 50 Hz. These data were obtained from the same cell featured in Figs. 2 and 3. Comparison of the response profiles obtained with 2-Hz modulation at 100% depth in this and previous figures indicates that the features of the SAM stimulus were robustly encoded in the moment-by-moment discharge rate of this neuron.

Summary measures derived from the MTF fail to convey important aspects of the way this neuron encodes the SPL profile of the SAM stimulus. Figure 5B shows the typical data representation for SAM responses: MTFs for average firing rate (rMTF; black line) and synchrony, measured as VS (tMTF; dashed line). Based on these functions, the best modulation frequencies for rate (rBMF) and temporal synchrony (tBMF) are 10 and 20 Hz, respectively. The substantial increase in response synchrony at 10 and 20 Hz occurs because the neuron is no longer capable of producing separate response peaks associated with the rising and falling phases of the SPL profile, as it does for modulation frequencies <5 Hz. The fact that the tMTF is band-pass is merely an artifact of the VS calculation, because the separate peaks in the response profiles for slow modulations distribute the spike times more evenly in the modulation period. It would be inaccurate to infer from the tMTF that the temporal fidelity of this neuron's representation of modulated stimuli is best at 20 Hz or that the neuron is tuned for that modulation frequency. At low modulation frequencies, the shape of the response profile captures the direct rate code for amplitude used by AI neurons. Accordingly, the TS curve is effectively flat over this range. In contrast, the shape of the tMTF captures only what the VS metric encodes—the precision of phase-locking for a presumed unimodal distribution of spike phases. This is reflected in the sharp increase in the VS from 5 to 10 Hz, where the MPH changes from a bimodal to a unimodal shape (Fig. 5).

Inspection of the response profiles at the lowest modulation frequencies also indicates that the response profiles have higher

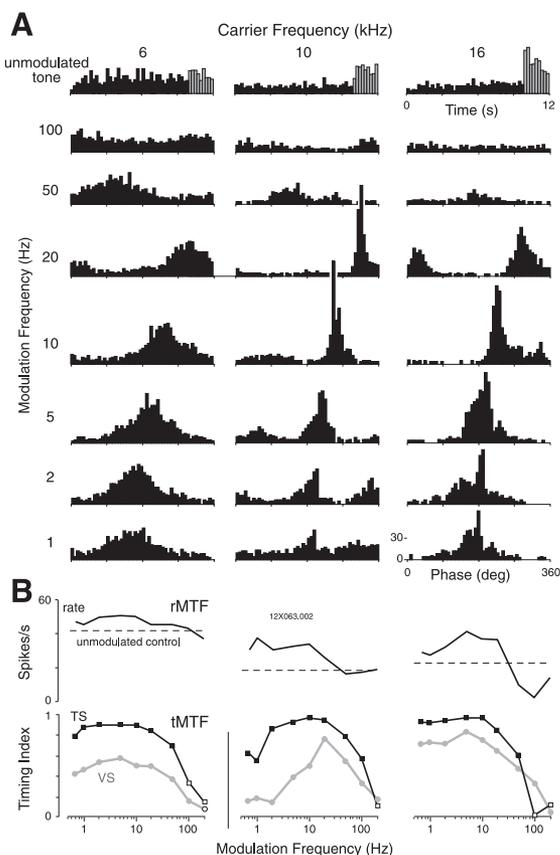


FIG. 4. Changes in carrier frequency are reflected in changes in MPH response profiles. *A*: matrix of MPHs depicts changes in response profiles across modulation frequency (*rows*) for 3 carrier frequencies (*columns*). Gray histograms in PSTHs for unmodulated control tone in the 1st row indicate firing rate during interstimulus interval. All SAM responses are from the same neuron tested with 100% modulation at 60 dB. *B*: curves in *top panels* depict corresponding modulation transfer functions (MTFs). Average discharge rate elicited by unmodulated control at each carrier frequency is indicated by dashed line. Curves in *bottom panels* depict corresponding VS (gray lines) and TS values (black lines; filled symbols indicate significance of timing indices at $P < 0.001$).

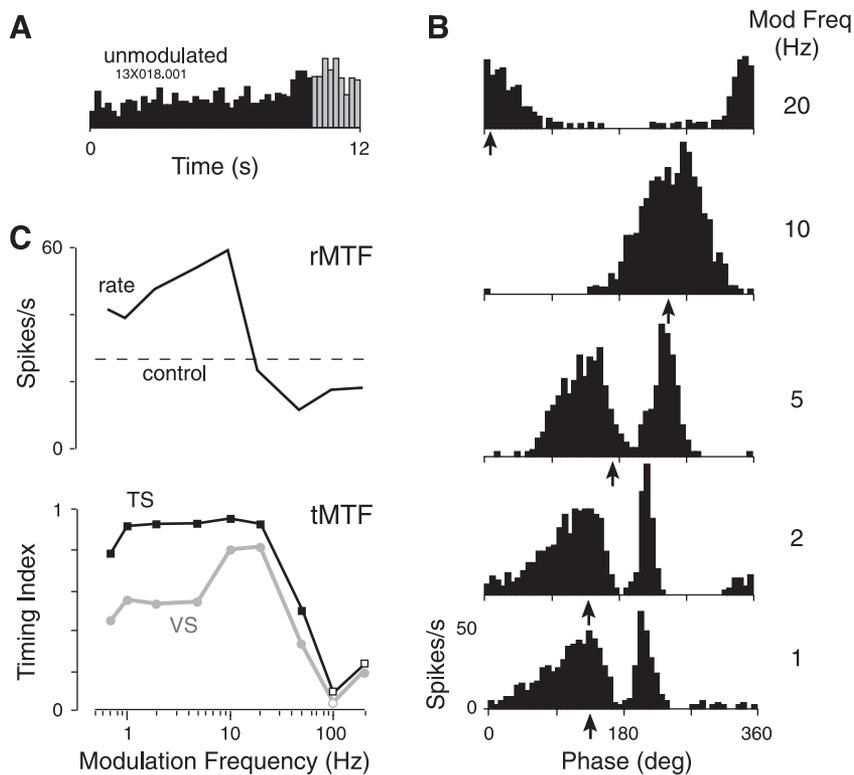


FIG. 5. Changes in modulation frequency are reflected in changes in MPH response profiles. *A*: response to unmodulated control tone indicates that response of this cell was suppressed by a 10-s tone at carrier frequency (600 Hz) and level (60 dB). *B*: series of MPHs shows responses to modulation frequencies indicated to *right* of each histogram. Rightward shift of mean response phase at increasing modulation frequencies reflects greater impact of response delay for shorter modulation periods (arrows). Modulation depth was 100%. *C*: curve in *top panel* depicts rMTF, the changes in firing rate across modulation frequency. Response to unmodulated control tone is indicated by dashed line. Curves in *bottom panel* depict tMTFs, the changes in VS (gray line) and TS (black line) across modulation frequency (filled symbols indicate significance of timing indices at $P < 0.001$).

frequency components than are present in the sinusoidal envelope of the stimulus. The logarithmic transform that generates the SPL profile from the amplitude envelope introduces increasingly higher frequency components at larger modulation depths, caused by the steep falling and rising phases of the SPL profiles. Nevertheless, the response profiles in Fig. 5 are clearly not faithful replicas of either the amplitude envelope or the SPL profile. The nature of this transformation appears to be related, at least in part, to the neuron's strongly nonmonotonic rate level function. At higher modulation frequencies, however, this relationship becomes obscured because the way amplitude changes can be encoded is limited by the neurons' maximal discharge rates.

Number of spikes per modulation cycle determines the resolution of a discharge rate code for amplitude

The distinction between a synchrony code for modulation frequency and a discharge rate code for amplitude can only be made when there are sufficient spikes within each modulation cycle to distinguish them. The average spike count per modulation cycle impacts a synchrony code and a rate code very differently. Firing rates affect the synchrony estimate because VS is maximal (i.e., 1) when all spikes fall in the same phase of the modulation period. Thus increasing the firing rate above one spike per modulation cycle will tend to increase the dispersion of spike times within the modulation period, lowering the VS. If cortical neurons serve to extract the modulation frequency, a synchrony code based on perfect phase-locking to a single and arbitrary point on the stimulus envelope would suffice. This code is achievable with one or fewer spikes per modulation cycle. In contrast, the resolution of an amplitude code based on instantaneous discharge rate is crucially

dependent on the number of discharges that a neuron can fire within each modulation period. Whereas a synchrony code for modulation frequency marks a point in the MPH, subject to intrinsic variability in response phase, a discharge rate code for amplitude changes must describe a function through the period, subject to intrinsic variability in response rate. In effect, the average number of spikes per modulation cycle limits the resolution of a rate code for sound amplitude.

Figure 6 shows the distributions of average spikes per cycle across all tested modulation frequencies. On a plot with logarithmic axes, the average numbers of spikes per modulation cycle is well described by a power function, intersecting an average of one spike per cycle between 10 and 20 Hz. Averaged across all neurons, mean spikes per modulation period falls from 18.8 at 1 Hz to 1.05 at 20 Hz. Our data showed that the multi-peaked response profiles evident at very low (<5 Hz) modulation frequencies typically become unimodal at or above 10 Hz (e.g., Fig. 5). Multi-peaked response profiles (e.g., Figs. 2, 3, and 5) were typically associated with large depth, low frequency modulations presented well above the best level of a strongly nonmonotonic neuron exhibiting sustained responses to the unmodulated control tone. Because not all cells were tested at the requisite high carrier levels, we cannot assess the absolute prevalence of the phenomenon. However, in a sub-population of neurons tested across a wide range of carrier levels with 100% depth, fully modulated SAM signals <20 Hz, 13 of 25 neurons exhibited multi-peaked discharges, indicating that they are not uncommon if the appropriate stimulus conditions exist. Nevertheless, the modulation frequencies that resulted in a clearly multimodal MPH never exceeded 20 Hz in our sample. In effect, cortical neurons default to a synchrony code for SAM above this critical range for direct amplitude coding. We will consider evidence that cortical neurons use a

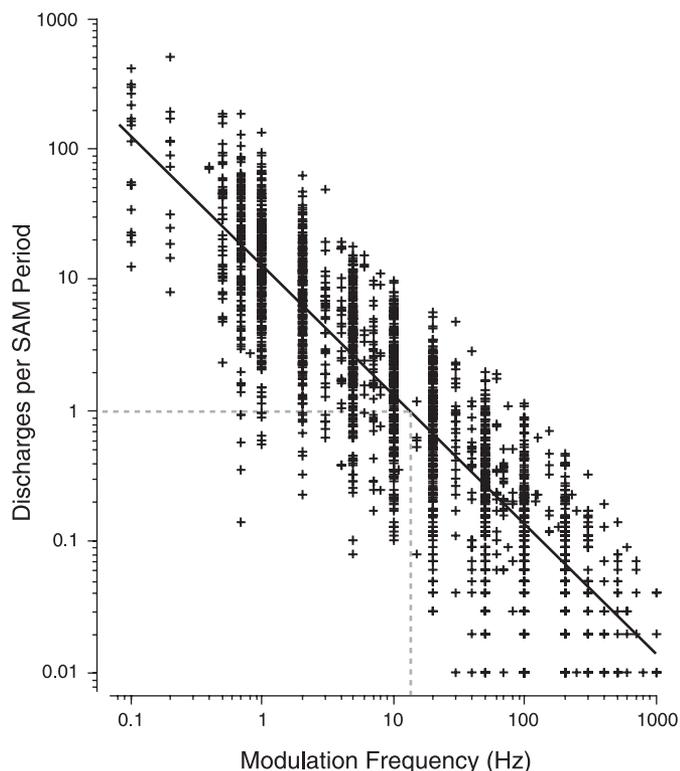


FIG. 6. Number of discharges per modulation period decreases with increasing modulation frequency ($n = 2,379$). Each neuron contributed as many points to the graph as there were tested modulation frequencies in its rMTF. Best linear fit to data ($r^2 = 0.77$) is shown by solid line. Dashed gray line extending from 1 discharge per period indicates that, on average, cortical neurons no longer fire more than a single action potential per period at ~ 11 Hz.

nonsynchronized rate code for modulation frequency in a subsequent section.

Response profiles of cortical neurons uniquely encode different SAM stimuli

The cortical response profiles appear to “multiplex” the parameters of SAM stimuli because those parameters jointly define the instantaneous amplitude of the signal. The foregoing examples showed that changes in modulation depth, carrier level, and modulation frequency are each captured by changes in the shapes of MPH response profiles at low modulation frequencies. Figure 7 shows the results of four slices through the space of possible SAM stimuli. For a given carrier frequency, the stimulus space of a SAM stimulus has three dimensions, as indicated by the axes of carrier SPL (x), modulation depth (y), and modulation frequency (z). It is clear from the SPL profiles that graded changes in a given stimulus parameters are represented by graded changes in the associated response profiles. As a result, the shapes of the MPHs suffice to identify the details of the stimulus envelope, which is jointly defined by these three parameters. In contrast, the summary measures based on rate, synchronization, and trial correlation appearing in the insets do little to enlighten precisely how the changes in amplitude are encoded by this neuron.

Increases in modulation frequency for a 20-dB SPL carrier (aligned diagonally in black) produce a progressive rounding of the peak associated with the SPL maximum (changes in the mean phase of the response reflect the group delay). Increases

in carrier level for a 2-Hz modulation (aligned horizontally in gray) also produce striking changes in the phase of the response, but these changes are attributable solely to the neuron’s tuning for SPL, because the modulation frequency is constant (latency differences related to changes in SPL are unlikely to have a significant impact when the modulation period is very long, as it is here). At 40 dB SPL and above, the cell responds to decreases in SPL near the falling phase of the stimulus envelope, consistent with its nonmonotonic pure tone RLF and preferred SPL of 25 dB SPL (data not shown; note, however, that the SAM-derived RLF in *b* is flat). Comparison of the response profiles at 20 and 70 dB SPL shows that the response phases are effectively inverted. Analogously, this neuron responded to short (100 ms) 20-dB SPL tones throughout the tones’ duration, whereas 70-dB SPL tones elicited only a phasic discharge at sound offset (data not shown).

This example further shows that the RLF can be very useful for predicting qualitative features of cortical responses to SAM at low modulation frequencies (particularly in cells with sustained responses to long duration tones). It is noteworthy that the reversal in response phase at 20 and 60 dB SPL is apparent even at the lowest modulation depths, where the change in actual stimulus levels is quite small (less than ± 2 dB for $m = 0.2$). For this reason, it is more accurate to say that the neuron generally responded when the stimulus approached its preferred SPL, because it clearly did not actually achieve that value for moderate depths. Thus it would be overly simplistic to suggest that a static measure of SPL tuning such as the RLF could simply be used as a lookup table for predicting cortical responses to those same SPLs in a dynamic context. A more detailed model relating to the shape of the RLF to shape of the MPH, taking the additional temporal factors that shape cortical responses (e.g., spike frequency adaptation, or synaptic depression) into account, is beyond the scope of this paper. Nevertheless, the heuristic value of the RLF in predicting SAM responses in many neurons is evidence for rate-based coding of sound level at low modulation frequencies.

Like Fig. 7, Fig. 8 includes a matrix of response profiles for SAM signals varying in depth, carrier level, and modulation frequency. In this neuron, however, complete modulation transfer functions (0.7 to 200 Hz) were obtained at 100% depth for carrier levels spanning a 40-dB SPL range. This neuron responded nonmonotonically to tone pips (Fig. 8A). Here we show results with SAM whose carrier level was near the peak (20 dB SPL), on the slope (40 dB SPL), and near its nadir (60 dB SPL). Similar reversals of response phase at very low modulation depths are evident in the responses to 20- and 60-dB SPL carriers (gray histograms on the diagonal), as are graded changes in the response profiles with increasing modulation depth, including the appearance of a narrow peak near the rising phase of the envelope for 100% modulation at 60 dB SPL. Again, stimulation at a low carrier SPL elicited responses coincident with the highest SPLs within the modulation period, with significant synchrony evident even at 200 Hz (data not shown). Responses to 40 dB SPL SAM were similar to those at 20 dB SPL, although the average firing rates are higher at modulation frequencies > 10 Hz, as shown in the rMTF (Fig. 8B). At 60 dB SPL, however, responses to both the falling and rising phases of the envelope are evident ≤ 20 Hz, although the earlier peak becomes substantially attenuated when the latter begins to dominate at 5 Hz and above. The rMTF at 60 dB SPL

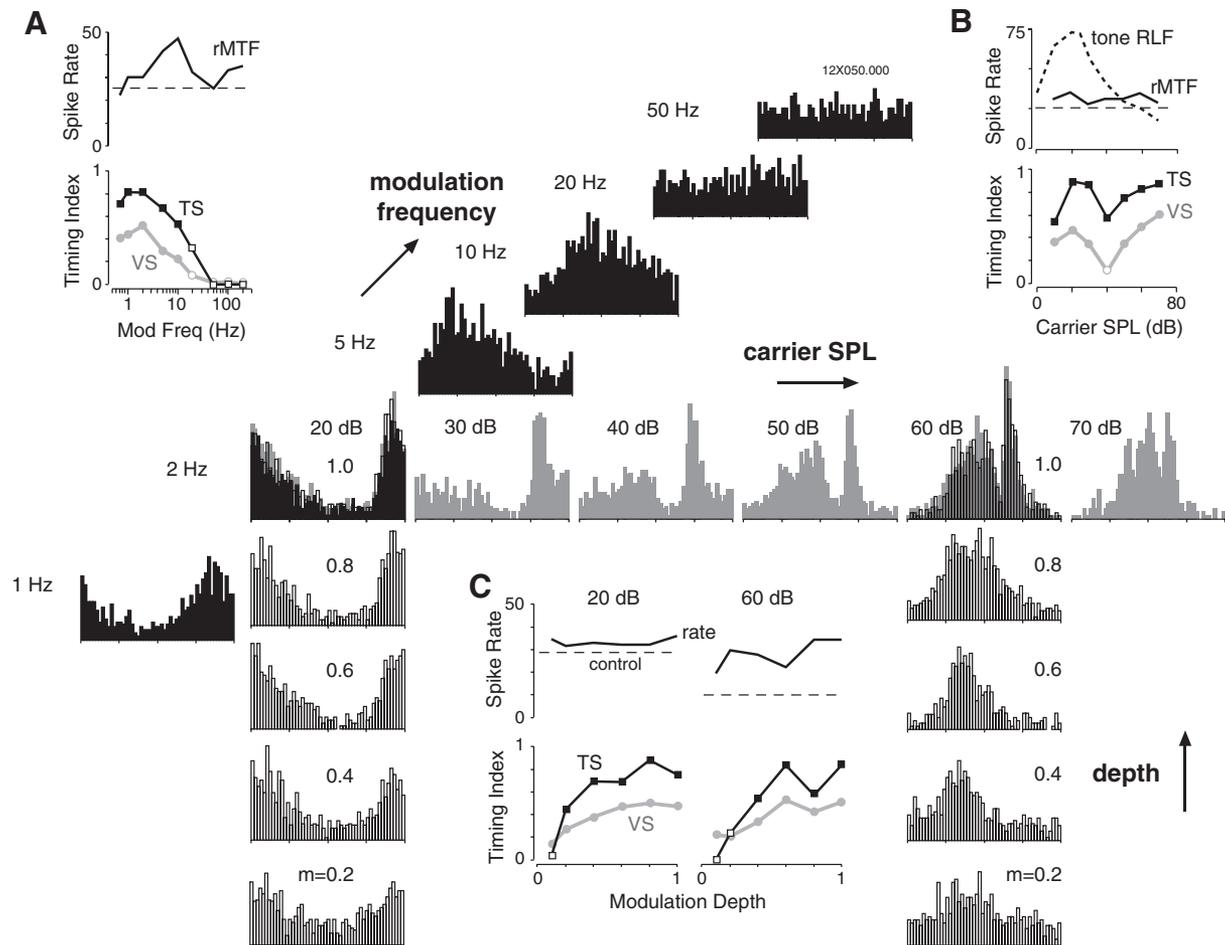


FIG. 7. Matrix of response profiles shows how changes in modulation frequency, modulation depth, and carrier level produce different but highly reproducible effects on shape of MPH. Response profiles corresponding to changes in modulation frequency for a 20-dB carrier level are aligned along the diagonal axis in black. Associated rMTFs and tMTFs appear in *top* and *bottom* panels of *inset a* (filled symbols indicate significance of timing indices at $P < 0.001$). Response profiles corresponding to changes in carrier level for a 2-Hz modulation are aligned along horizontal axis in gray. Associated curves depicting changes in firing rate and timing indices for different carrier levels appear in *top* and *bottom* panels of *inset b*. Response profiles corresponding to changes in modulation depth are shown in vertical columns on the *left* (2 Hz at 20 dB) and *right* (2 Hz at 60 dB). Associated curves depicting changes in firing rate and time indices appear in *inset c*. Because data were obtained from separate runs, the 100% modulated, 2-Hz stimulus at 60 dB was repeated twice, and the analogous 20-dB stimulus was repeated 3 times (carrier frequency was constant at 700 Hz). Resulting MPHs are overlaid to show robustness of MPH shapes for repeated trials.

drops rapidly with increasing modulation frequency because of the diminished response to the falling phase of the envelope, which comprises the majority of the spikes at 1 and 2 Hz. Thus the rBMF for a 60-dB SPL carrier is 1 Hz compared with 5 and 10 Hz for the 20- and 40-dB SPL carrier levels, respectively.

The relative ordering of the tMTFs in Fig. 8B indicates a progressive loss of synchrony as the carrier level is increased. At low modulation frequencies, the cortical reductions in VS at high carrier levels typically reflect increases in firing rate during the rapid falling and rising phases of the SPL profile, which distributes spikes more widely throughout the modulation cycle. For example, the sharp peak that occurs during the rising phase of the SPL profile is not captured by the tMTF at 60 dB SPL. The effect on the VS is diluted by spikes occurring at distant phases, despite the obvious temporal precision of the neuron's responses. Given the shape of the RLF, this neuron might be expected to respond poorly to 60 dB SPL SAM because the stimulus spends relatively little time in the range of the neuron's preferred SPLs. The rMTFs and tMTFs seem to confirm this expectation. The TS values did not vary significantly across carrier level (Wilcoxon ranked sum; $P > 0.1$). On

the other hand, the changes in TS across modulation frequency were highly correlated for the different carrier levels ($r^2 > 0.7$; Fig. 8B), which suggests the existence of shared temporal limits on the neuron's ability to encode SPL changes across the tested range.

The foregoing examples suggest that the shapes of the MPHs can, in many cases, be uniquely associated with the SAM stimuli that elicited them. We explicitly tested this notion by applying a PSTH-based response classifier to spike trains elicited by SAM stimuli that varied along a single parameter axis: carrier level, modulation depth, or modulation frequency (see METHODS). Figure 9 shows the confusion matrices (A–C) obtained when applying the classifier to the MPHs depicted in Figs. 2, 3, and 5, respectively. In all three cases, classifier performance was substantially better than chance, as indicated by the distribution of correct estimates of stimulus identity along the diagonal. Although the elimination of firing rate information by response normalization (see METHODS) did reduce performance in these examples, the “phase only” confusion matrices in the central column evidence a fairly modest reduction in classifier performance. In contrast, the “rate only”

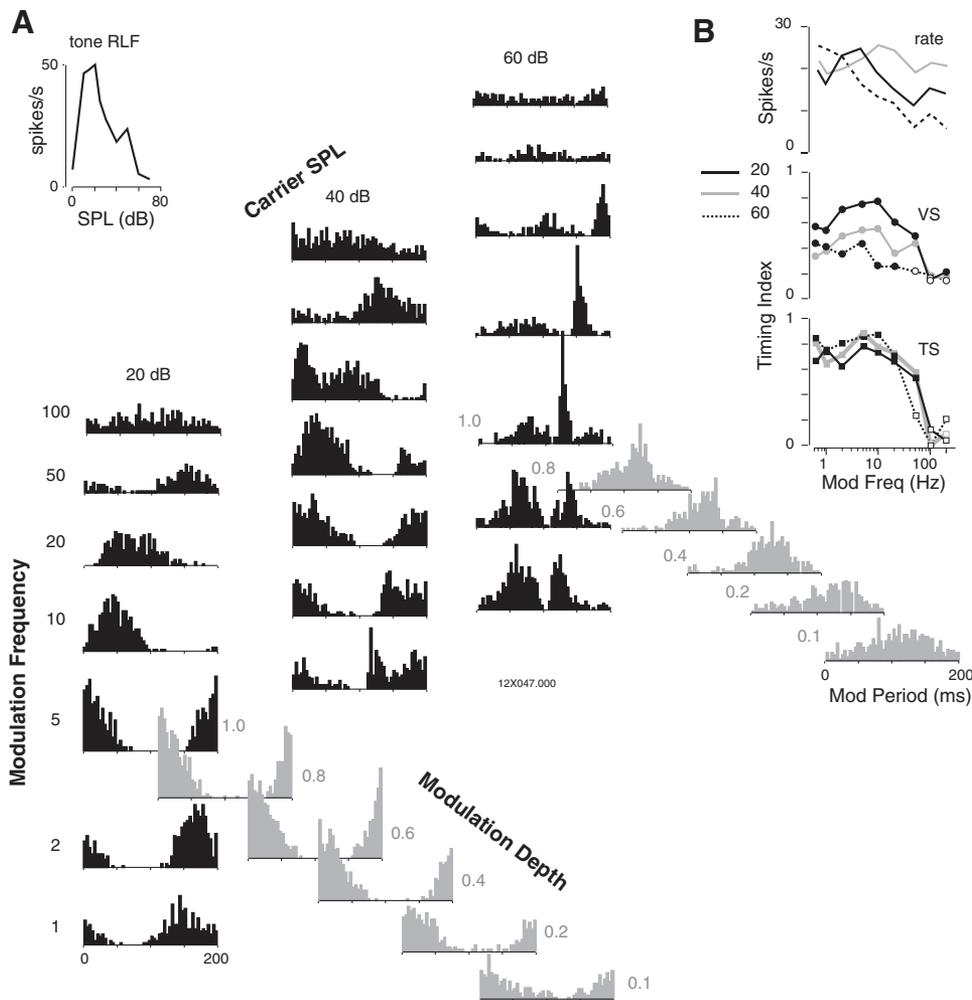


FIG. 8. Matrix of response profiles shows how changes in carrier level, modulation depth, and modulation frequency produce distinct MPH shapes in a cortical neuron. Rate level function obtained with tone pips (100 ms) is shown in *inset a*. Responses to 3 different carrier levels (20, 40, and 60 dB) across modulation frequency are aligned in vertical columns from *bottom left to top right* of main figure in black. Effects of varying modulation depth for 5-Hz SAM are shown along diagonal for the 20- and 60-dB carriers in gray (carrier frequency was constant at 4,000 Hz). *Inset b*: rMTFs (*top*) and MTFs based on VS (*middle*), and the correlation metric, TS (*bottom*; filled symbols indicate significance of timing indices at $P < 0.001$).

confusion matrices to the right indicate poorer performance when the basis for the estimate is limited to average spike rate.

To explore these issues at the population level, we identified a subset of neurons for which we had obtained MPHs for a set of SAM parameter values. We were able to identify 25 neurons that had been tested across a large range (mean = 52 dB) of carrier levels at low (<20 Hz) modulation frequencies. We also identified a subset of neurons ($n = 145$) where we had obtained both MTFs and modulation depth functions (MDFs) at the best modulation frequency. After discarding data based on nonstandard trial lengths, we applied the classifier to the remaining 124 MTFs and 111 MDFs. These data represent the neuronal population characterized below.

Because our PSTH-based classifier operates on vectors of binned spike counts, the choice of binwidth determines the temporal resolution at which the test spike train is compared with the different response templates. The curves in Fig. 9, *D–F* show how classifier performance varied as a function of binwidth for the examples shown (the spike train and phase-only confusion matrices are depicted at the optimal binwidth, i.e., the peaks of the black and gray curves, respectively). To assess this relationship at the population level, we computed percent correct as a function of bin size (1, 2, 4, 8, 10, 20, 40, and 1,000 ms) for each neuron, normalized each curve by its peak, and generated population composite curves by averaging the results across cells for each SAM parameter. Although

optimal binsize did vary from neuron to neuron, the curves for most individual neurons were relatively flat from 8 to 40 ms. The composite curves for all three SAM parameters exhibited significant (Wilcoxon ranked sum, $P < 0.001$) increases when the binwidth increased from 1 to 2, 2 to 4, and 4 to 8 ms, but were essentially flat ($P > 0.01$) from 8 to 40 ms, indicating that binning over roughly 10 to tens of milliseconds produced the best classifier performance in nearly all neurons. All three composite curves dropped significantly ($P < 0.0001$) at the 1,000-ms bin, corresponding to the “rate only” estimate. The shapes of the composite curves for the phase-only classifiers were essentially identical to those for the spike train classifiers, although the slopes tended to be shallower for small bins for the MTF and MDF discriminations.

Scatterplots comparing the performance of the spike train, phase-only, and rate-only classifiers for each SAM parameter are shown in Fig. 10. Performance of the spike train classifier is mapped to the abscissa, whereas performance of the phase-only (black) and rate-only (gray) classifiers are mapped to the ordinate. Discrimination of depth was relatively poor, as indicated by the clustering of points near the origin in Fig. 10*A*. Nevertheless, classifier performance significantly ($P < 0.0001$; see METHODS for a description of the bootstrapping procedure) exceeded chance in 59 of 111 neurons (53%) when the original spike train was used, in 47 neurons (42%) when only phase information was used, and in 21 (19%) neurons when only

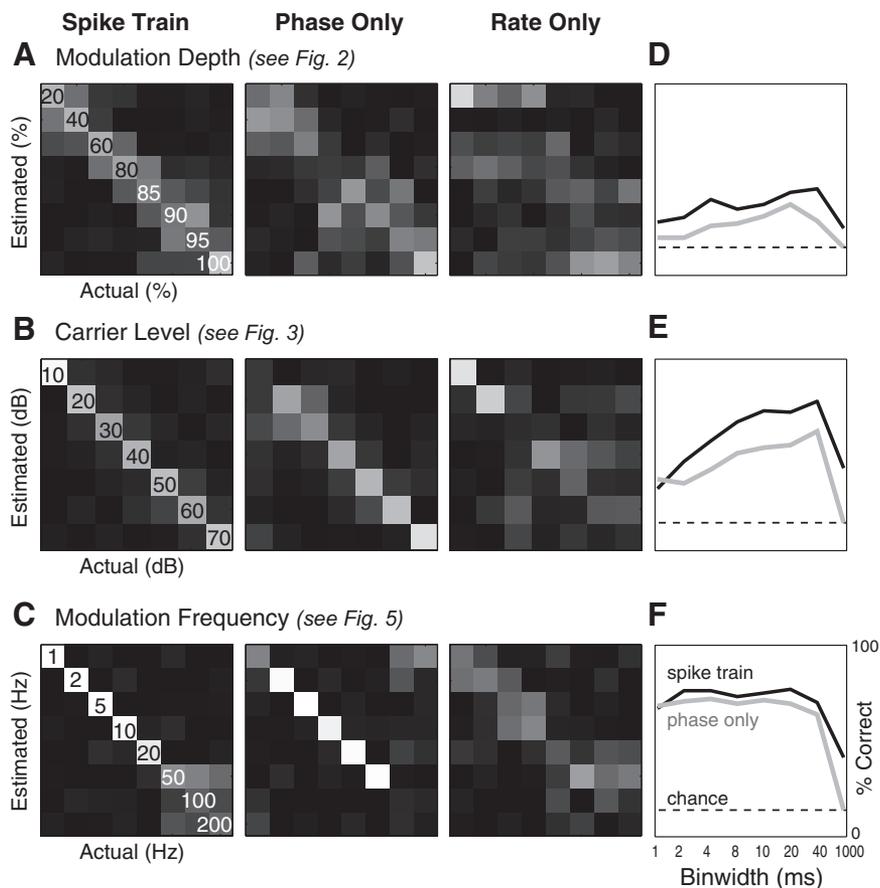


FIG. 9. Set of 9 confusion matrices depicted here show performance of spike train, phase-only, and rate-only classifiers for MPHs that had been previously shown in Figs. 2 (A: modulation depth), 3 (B: carrier level), and 5 (C: modulation frequency). Note that Figs. 2 and 5 were based on data from the same cell. Numbers along diagonal for confusion matrices in *leftmost column* indicate stimulus values to be discriminated by classifier. Grayscale values for each square indicate how often classifier chose stimulus value in that row when actual stimulus had the value of the column. Perfect classification (20/20) is indicated by white squares along diagonal. Curves depicted in D–F indicate percent correct as a function of binwidth chosen for classifier. Performance of spike train classifier is shown in black and that of phase-only classifier in gray. Performance of rate-only classifier (binwidth = 1,000) corresponds to rightmost point on each black curve. Chance performance is indicated by dashed line on each panel.

average rate information was used. When the relative quality of the estimates was accounted for by assigning a cost index (see METHODS), significant performance rose to 77 (86), 69 (77), and 29% (32), respectively. When the performances of the spike train and phase-only classifiers were compared, there was a marginal but significant benefit to using the full spike train with respect to both percent correct (Wilcoxon ranked sum, $P < 0.0075$; spike train mean: 0.28; phase-only mean: 0.25) and cost ($P < 0.0156$; spike train mean: 0.31; phase-only mean: 0.34). Nevertheless, the phase-only classifier significantly out-

performed the rate-only classifier both in terms of percent correct ($P < 0.0002$; mean: 0.21) and cost ($P < 5.6 \times 10^{-11}$; mean: 0.41).

As is evident in Fig. 10B, discrimination of carrier level was substantially better than that for modulation depth, although it is worth noting that we typically tested fewer carrier levels (typically 6 or more, with a minimum of 4 and maximum of 9), than modulation depths (typically 6, or 11, i.e., 0 to 100% in 20 or 10% steps) so the baseline for chance performance is also somewhat higher for carrier level. We corrected for the dis-

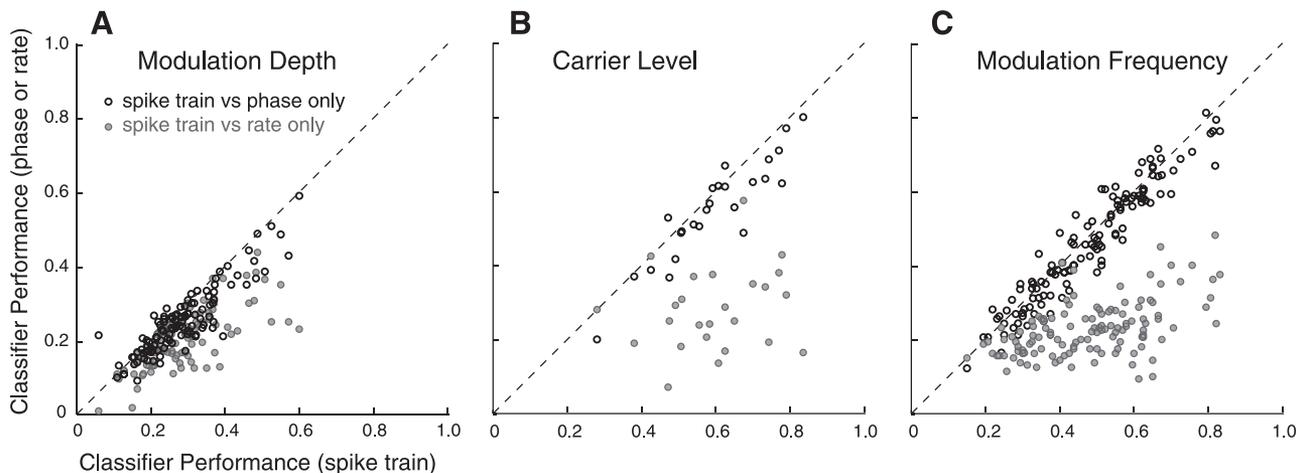


FIG. 10. Scatterplots in each panel compare performance (in percent correct) of spike train classifier, which is plotted on abscissa, to performances of phase-only (black circles) and rate-only (gray circles) classifiers, which are plotted along ordinate, for stimulus sets that varied in modulation depth (A; $n = 111$), carrier level (B; $n = 25$), and modulation frequency (C; $n = 124$).

crepancy in stimulus set size by converting percent correct into z-scores with respect to the bootstrapped distribution (see METHODS). When this is done, classifier performance remains substantially better for carrier level than for modulation depth for the spike train classifier (Wilcoxon ranked sum, $P < 8.1 \times 10^{-11}$; means of 13 vs. 5.2) and phase-only classifier ($P < 4.5 \times 10^{-11}$; means of 11.8 vs. 4), but not the rate-only classifier ($P = 0.1960$; means of 2.9 vs. 4). As the alignment of the black circles on the unity line in Fig. 10B suggests, the spike train and phase-only classifiers performed similarly (Wilcoxon ranked sum, $P = 0.3037$; means of 0.6 vs. 0.55). Given the foregoing observation, it is not surprising that the rate-only classifier performed substantially worse than the phase-only classifier, ($P < 2.1 \times 10^{-7}$; means of 0.28 vs. 0.55). Performance was significantly better than chance in 96 (24 of 25), 92 (23), and 44% (11) of the neurons in our sample for the spike train, phase-only, and rate-only classifiers, respectively.

Discrimination of modulation frequency, as expected, was also quite good in many neurons (Fig. 10C). Surprisingly, however, discrimination of modulation frequency was not significantly better than discrimination of carrier level for the spike train (Wilcoxon ranked sum, $P = 0.7728$; z-scored means of 13.7 vs. 13), phase-only ($P = 0.5239$; 13 vs. 11.8), or rate-only ($P = 0.1844$; 4 vs. 2.9) classifiers. Thus the notion that modulation frequency is preferentially represented by the responses of cortical neurons was not supported, because carrier level is represented with equivalent fidelity. Figure 10C also shows that average firing rate information is, at best, ancillary to modulation frequency discrimination, because the spike train and phase-only classifiers exhibit statistically equivalent performance (Wilcoxon ranked sum, $P = 0.3954$; means of 0.48 vs. 0.46), and the phase-only classifier dramatically outperforms the rate-only classifier ($P < 2.3 \times 10^{-29}$; means of 0.46 vs. 0.20). Performance was significantly better than chance in 95 (118 of 124), 94 (116), and 36% (45) of the

neurons in our sample for the spike train, phase-only, and rate-only classifiers, respectively.

To assess how well particular stimuli could be discriminated on the basis of cortical responses, we reanalyzed the set of confusion matrices for the population. Instead of summing over the diagonal to generate the global percent correct, we summed over columns to generate the percent correct for each stimulus in the set individually. We averaged percent correct for every instance of a given stimulus in the population, for each classifier, to produce the curves shown in Fig. 11, A–C. Performance for modulation depth was best at extreme values (0 or 100%), which is compatible with the observation that unmodulated control tones produced weak, flat PSTHs, whereas fully modulated stimuli often produced robust PSTHs with unique features, such as those related to the emergence of double peaking in the MPH. Conversely, SAM stimuli in the middle range of modulation depth produce excursions in instantaneous SPL around the carrier level within a fairly limited range of SPLs (see Fig. 2A), so the resulting MPHs tended to be relatively similar in shape, which apparently resulted in poorer discrimination. The discrimination curves for carrier level (Fig. 11B) indicate that the spike train and phase-only classifiers outperform the rate-only classifier at all carrier levels >0 dB, where low average firing rates support reliable discrimination based on rate alone, and phase information is limited. Finally, the discrimination curves for modulation frequency indicate that average rate information is generally poor, with some slight improvement at the highest modulation frequencies. It is not immediately obvious why the spike train and the phase classifier differ for the 1-, and to a lesser extent, 2-Hz stimuli, but it is possible that PSTH shapes corresponding to either double-peaked or “notched” (e.g., Fig. 3, 40 dB) MPHs, where spikes are widely distributed throughout the modulation period, are confused with responses above the synchrony cut-off when firing rate information cannot be used to distin-

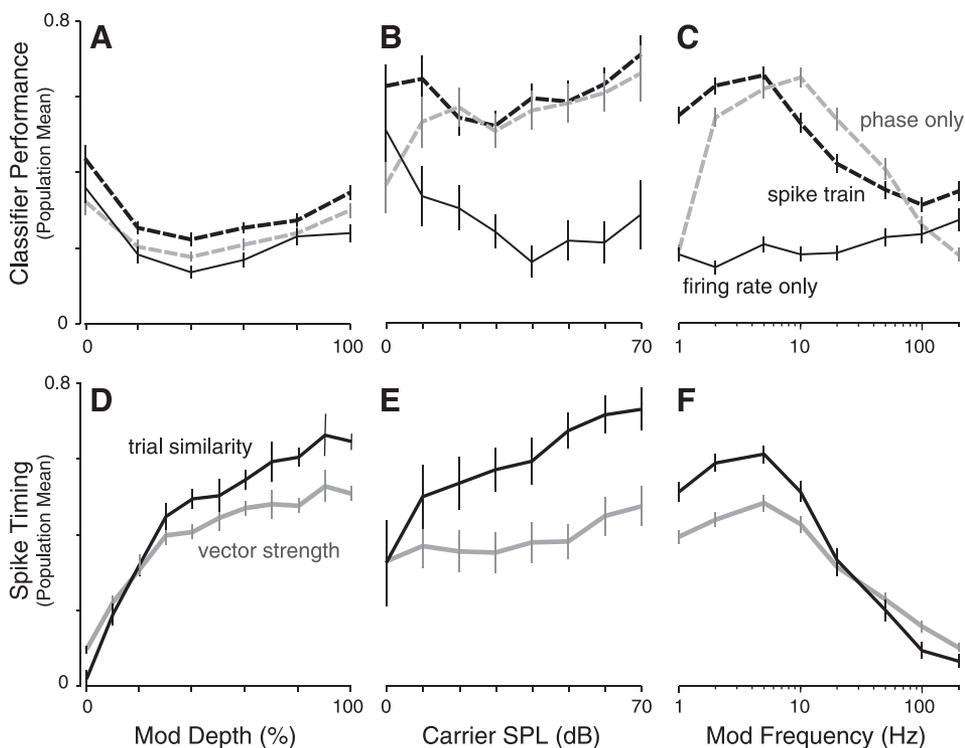


FIG. 11. Sets of curves in the *top row* (A–C) indicate population average of performance of different classifiers on a stimulus by stimulus basis. Composite curves for spike train, phase-only, and rate-only classifiers are shown in dashed black, dashed gray, and black, respectively. Vertical bars indicate ± 2 SE of measurement. Horizontal axes are shared with panels in the *bottom row*, which show how population averages of vector strength and trial similarity vary across modulation depth (D), carrier level (E), and modulation frequency (F). Vertical bars indicate ± 2 SE of measurement.

guish them. In neurons with low firing rates, it is likely that the PSTH shapes are inconsistent when so few modulation cycles (i.e., 1 or 2) are available to define them. In light of this, the superior performance of the phase-only classifier at 10 and 20 Hz suggests that, when more modulation cycles are available, second-to-second fluctuations in firing rate can worsen performance by promoting spurious, rate-based stimulus estimates.

In Figs. 2–5, 7, and 8, we showed examples where the values for TS exceeded those for VS, indicating that the latter metric failed to capture the fact that cortical responses could reliably produce MPH shapes that the VS metric evaluates as poorly synchronized. Figure 11, *D–F*, shows the population averages for VS and TS across modulation depth, carrier level, and modulation frequency, respectively. Higher modulation depths and carrier levels tend to increase the discrepancy between these metrics in favor of TS. Conversely, lower modulation frequencies (<10 Hz) favor TS with respect to VS, in support of our claim that cortical neurons essentially default to a synchrony code at 20 Hz or higher, where the two curves merge. Overall, however, TS and VS values were highly correlated when the values averaged in Fig. 11*F* ($n = 1245$) are compared ($r = 0.80$; $P < 8.7 \times 10^{-282}$). As a result, the shapes of tMTFs derived from TS and VS within individual cells are also highly correlated, such that the population mean of such correlations was 0.84.

If TS more effectively captures the ability of cortical neurons to represent a particular stimulus, we would expect that it should be more highly correlated with classifier performance on a cell by cell basis. To test this hypothesis, we computed the mean VS and TS for each stimulus set and correlated these values with the performance of the spike train and phase-only classifiers. For modulation depth, the (Pearson product moment) correlation between percent correct for the spike train classifier and mean VS was weak ($r = 0.22$; $P = 0.0205$). A much stronger correlation was obtained with the mean TS ($r = 0.44$; $P < 8.7 \times 10^{-7}$). The corresponding correlations for the phase-only classifier were 0.17 ($P = 0.0807$) and 0.36 ($P = 0.0001$). When we repeated the analysis for carrier level, the correlation between spike train classifier performance and mean VS was 0.25 ($P = 0.2275$), compared with 0.70 ($P = 0.0001$) for mean TS; the corresponding values for the phase-only classifier were 0.20 ($P = 0.3291$) and 0.80 ($P < 2.2 \times 10^{-6}$). As expected given the nature of the discrimination, mean VS was most highly correlated with spike train classifier performance for modulation frequency ($r = 0.44$; $P < 3.2 \times 10^{-6}$). The correlation with mean TS was substantially higher ($r = 0.75$; $P < 1.7 \times 10^{-23}$), however. We obtained analogous results for the phase-only classifier, with corresponding values of 0.43 ($P < 6.8 \times 10^{-7}$) and 0.80 ($P < 3.1 \times 10^{-29}$). Thus mean TS provides a demonstrably better metric for assessing the cortical representation of modulated signals than does mean VS.

Qualitative features of response profiles to SAM stimuli can be predicted by responses to an unmodulated carrier tone

The foregoing single-unit examples indicate that, at low modulation frequencies, cortical responses can sometimes be related to features of the SPL profiles of SAM stimuli by considering each neuron's tuning for sound amplitude. Because changes in the parameters defining SAM signals produce

changes in the shapes of neuronal response profiles consistent with their tuning for sound level, heterogeneity in the shapes of cortical RLFs predict a corresponding heterogeneity in the shapes of cortical response profiles for a given stimulus. Given the myriad temporal factors that shape cortical responses, however, it is likely that a RLF measured with 100-ms tone pips will not always be relevant to the tuning for sound amplitude expressed at longer stimulus durations. In the absence of complete RLFs obtained with long stimuli, we simply classified responses to unmodulated control tones presented at the same carrier level and duration as either sustained and nonsustained.

Sustained responses differed significantly (see METHODS) from the spontaneous rates estimated during the interstimulus intervals: driven responses exceeded the spontaneous rate ($n = 113$); "suppressed" responses were significantly below it ($n = 42$). Nonsustained responses to the unmodulated control tone ($n = 201$) were not significantly different from the spontaneous rate. It is important to note that these are response categories for a particular stimulus, not cell categories. For example, changing in the SPL of the unmodulated control tone presented to a neuron that is nonmonotonically tuned to SPL could change its designation from "driven" to "suppressed." Nevertheless, sustained responders generally continued to give sustained responses as stimulus parameters varied and encoded parameters of the SAM stimuli robustly. As a result, previous figures have emphasized neurons with sustained responses because they appear to track changes in stimulus amplitude more effectively.

Response class designation was not significantly associated with either carrier frequency or level (ANOVA, $P > 0.10$). Response classes were not distinguishable on the basis of summary measures derived from the MTF, such as the distributions of rBMF and tBMF (ANOVA, $P > 0.10$), although there was a trend for VS at the tBMF to be slightly higher for the suppressed response class (ANOVA, $P < 0.0314$). As one would expect from how the classes were defined, the average firing rates elicited by the pure tone control varied across response types (ANOVA, $P < 0.0001$), such that the driven class was higher (mean = 22.9 spikes/s) than the nonsustained (13.2 spikes/s) and suppressed classes (9.9 spikes/s). Conversely, the distributions of spontaneous rates varied among groups (ANOVA, $P < 0.0001$), being highest for the suppressed class (mean = 23.3 spikes/s), intermediate for nonsustained responses (12.01 spikes/s), and lowest for the driven response class (6.6 spikes/s).

Figure 12 depicts the responses of four cells that exhibited sustained changes in firing rate during the unmodulated control stimulus. Driven responses are shown in the first two columns. Characteristically, there was a trough coincident with the periodic amplitude minimum at the lowest modulation frequencies. In contrast, suppressed responses were characterized by a response peak coincident with the periodic amplitude minimum. In one case (third column), the response to the decrease in amplitude was accompanied by a second response peak at the rising phase of the envelope.

The MTFs for both rate and synchrony appear in the bottom two rows of Fig. 12. For columns 1–4 respectively, synchrony cut-offs were 40, 60, 50, and 200 Hz, and tBMFs were 10, 6, 20, and 20 Hz. None of these responses was particularly well tuned in terms of average firing rate, although all possessed a

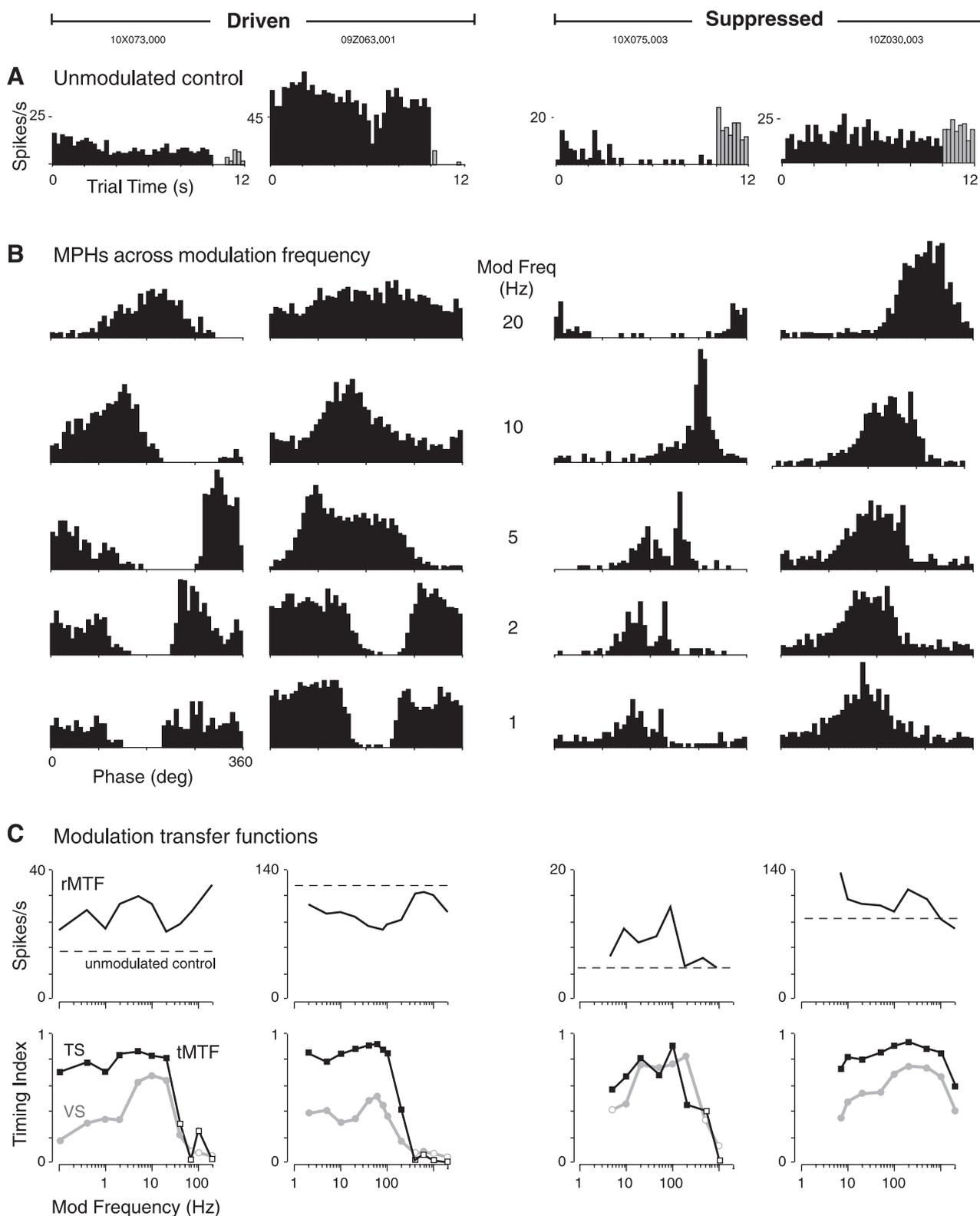


FIG. 12. Responses of 4 neurons exhibiting driven (*left 2 columns*) and suppressed (*right 2 columns*) responses to unmodulated control tone are shown. *A*: PSTHs show neural response during presentation of a 10-s unmodulated control tone at carrier frequency (black) and during the 2-s interstimulus interval thereafter (gray). *B*: MPHs in each column represent response of a neuron to modulation frequencies from 1 (*bottom*) to 20 Hz (*top*). *C*: *top row* depicts rMTFs corresponding to MPHs shown in *B*. Dashed line indicates firing rate elicited by unmodulated control tone. Associated tMTFs are shown in the *bottom row* (filled symbols indicate significance of timing indices at $P < 0.001$).

rBMF (200, 60, 10, and 500 Hz), defined as a maximum response that was significantly greater than that obtained for at least two other modulation frequencies. By this criterion, 83% (295/356) of the units in our sample had a rBMF for 100% modulated stimuli, and 85% (302/356) had a worst modulation frequency (rWMF; i.e., a minimum response significantly smaller than was obtained for at least two other points).

The nonsustained response class was more heterogeneous and included a range of response types for pure tones (onset, offset, and both), as well as a range of spontaneous rates, as the examples chosen for Fig. 13 are intended to suggest. At one extreme, there were some cells which appeared to respond selectively to the rapid rising phase of the amplitude profile (*left column*). In other cells, the response profile was more complex (*right column*). Although the MPH shapes were variable, a response peak coincident with the rising edge of the stimulus envelope was the most consistently observed feature of SAM responses from the nonsustained class. Qualitatively, cells in the nonsustained class exhibited noisier responses, as

would be expected from the fact that the response classes were defined by how effectively the SAM stimulus controlled the firing rate over long intervals.

The robust performance of the phase-only classifiers indicates the importance of spike timing to the cortical representation of SAM stimuli. As previous examples have shown, however, spike trains elicited by a given SAM stimulus will depend on how the carrier level interacts with tuning for sound amplitude. To address the distribution of response phase for a given amplitude at the population level, we created composite MPHs based on all responses ($n = 66$) from a single cylinder placement in the left hemisphere of a single animal (X) to 100% modulated SAM signals presented at the same SPL (60 dB; recordings from this hemisphere were obtained at a standard SPL for this purpose). Because we were primarily interested in response profile diversity for an equivalent sound level, we presented the SAM signals at the best frequency of each neuron rather than using broadband noise, whose effective SPL would depend on each neuron's filter bandwidth. To

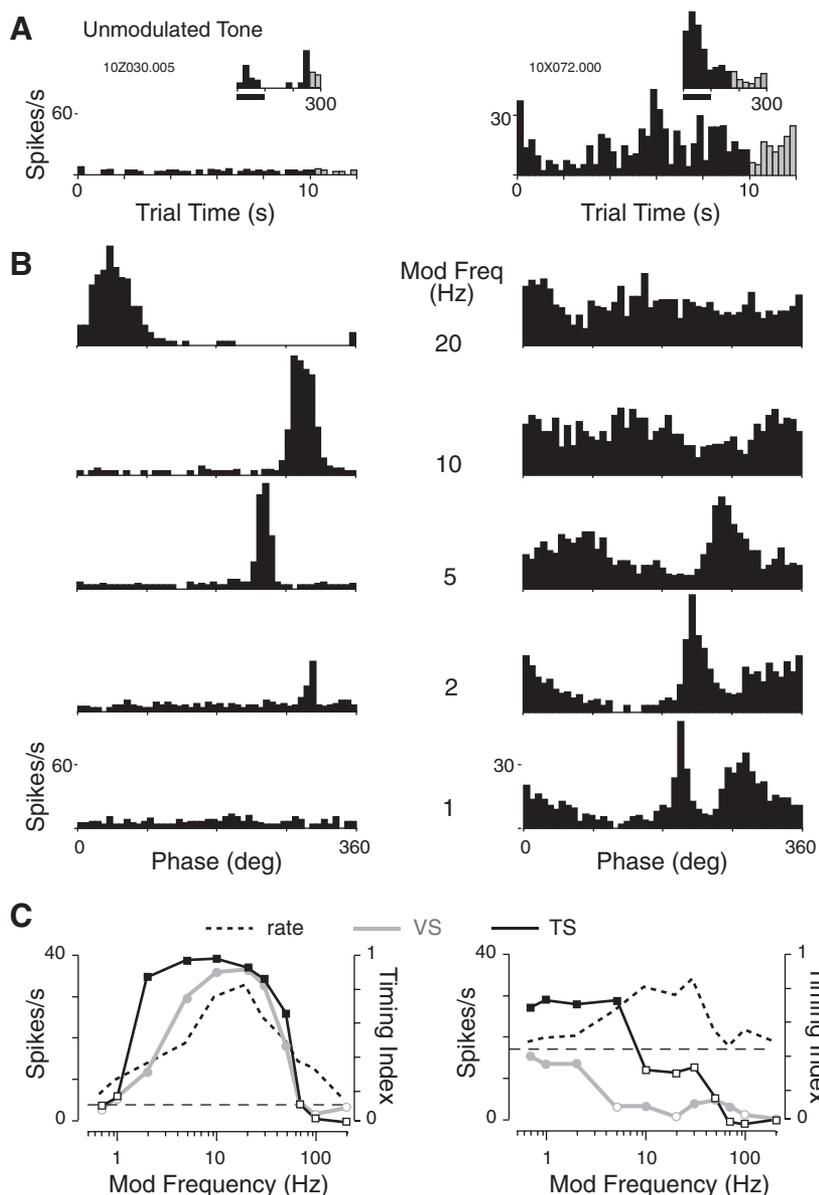


FIG. 13. Responses of 2 neurons exhibiting nonsustained responses to unmodulated control tone are shown. *A*: PSTHs show neural response during presentation of a 10-s unmodulated control tone at carrier frequency (black) and during the 2-s interstimulus interval thereafter (gray). To show that these neurons clearly responded to tonal stimuli, responses to 10 repetitions of short tone pips are shown as insets. Black bar under each indicates tone duration. *B*: MPHs in each column represent response of a neuron to modulation frequencies from 1 (*bottom*) to 20 Hz (*top*). *C*: rMTFs (dashed line) and tMTFs for VS (gray) and TS (black) corresponding to data in *B* are shown (filled symbols indicate significance of timing indices at $P < 0.001$).

preserve changes in firing rate across modulation frequency while preventing the most responsive neurons from dominating the composite MPH, the response profiles for each neuron were normalized by its maximum binned spike count (i.e., if the maximum binned spike count of 50 spikes occurred for the 10-Hz SAM stimulus, all bins would be divided by 50, and only the normalized MPH at 10 Hz will contain a bin valued at 1). The normalized MPHs at each modulation frequency were averaged across cells to generate the composite MPH.

Examination of the left column of Fig. 15 reveals that for modulation frequencies up to ≥ 10 Hz, there is a clear, albeit broad, peak in the composite MPH corresponding to the highest stimulus levels in the modulation period. Clearly, the composite MPH obscures many of the fine details evident in the MPHs of individual neurons. Given the diversity in cortical responses to SAM stimuli and differences in group delay across neurons, it is surprising that even without selecting inputs by response type or correcting for response latency, the population response to 100% SAM at 60 dB SPL is sufficiently coherent to yield a modulated composite MPH for modulation frequencies comprising much of the communication range.

The remaining columns in Fig. 15 show the composite MPHs subdivided by response class. For modulation frequencies < 5 Hz, the differences in the MPH shapes are particularly clear. Driven responses ($n = 22$) follow the amplitude envelope quite well, showing a clear response minimum at the envelope minimum. Conversely, suppressed responses ($n = 11$) respond maximally at the envelope minimum. The nonsustained class proved the most variable in terms of response phase at the lowest modulation frequencies, but became increasingly similar to the driven class above 5 Hz, at which point most MPHs tended to converge on a unimodal response clustered around the rising phase of the envelope. We confirmed that the shapes of the MPHs differed significantly by response class by performing principal components analysis (PCA) on all the 52-bin MPHs used to derive the composites. The eigenvalues associated with the first three principal components were used to map each cell onto a three-dimensional space, and mean values for each response class were computed. The Euclidean distance between the group means serves as a measure of how much MPH shape differed across response classes. To verify that the response class designation captured meaningful differences in the MPH shapes, we repeated the calculation above after randomizing the assignment of cells to response classes (1,000 iterations). A significance value was assigned by counting the fraction of instances when the distances between the group means for the bootstrap procedure exceeded the distances for the real composite MPHs. By this criterion, the driven and suppressed classes were significantly different for all modulation frequencies up to and including 5 Hz ($P < 0.01$), indicating that the response class designation captured important features of the MPH shapes for the sustained responders at very low modulation frequencies.

As can be seen in Fig. 15, the MPH shapes for the driven and suppressed classes are complementary, and when summed, will clearly reduce the apparent modulation of the population response. More generally, it is apparent that the population representation of SAM signals, while variable by response class, is relatively stereotyped within each class, and the composite MPHs of each are readily explicable in terms of the instantaneous stimulus amplitude. Because response class for a

given carrier level (e.g., 60 dB SPL) depends on each neuron's tuning for sound amplitude, the phase coherence of the responses of a given cortical population will depend on their tuning for sound amplitude and not merely the distribution of their synchrony cut-offs in terms of modulation frequency.

Macaque AI does not seem to contain a nonsynchronized population that encodes high modulation frequencies with changes in average firing rate

In this report, we focused on instantaneous rate coding of SAM signals at low modulation frequencies. Lu et al. (2001) have argued that higher modulation frequencies could be encoded by changes in average discharge rates. They identified a "nonsynchronized population" that did not respond synchronously to clicks presented at either long or short interclick intervals, but encoded variations in short interclick intervals (> 20 – 30 ms, or roughly 30–50 Hz) with changes in average discharge rate. In our sample, there were six neurons (6/361, 1.7%) that exhibited significant variations in average discharge rate without also exhibiting significant synchrony to at least one tested modulation frequency. We were also able to identify a group of neurons (56/355; 16%) that exhibited a statistically significant change in average rate for modulation frequencies beyond their synchrony cut-offs ($P < 0.001$ by the Rayleigh test). Although there appears to be no evidence for a "pure" nonsynchronized population, it is still possible that cortical neurons shift from a synchrony code to a rate code as modulation frequency increases. To evaluate this possibility, we compared the discharge rate contrast (defined as the difference across modulation frequency below the synchrony cut-off) to the rate contrast for modulation frequencies above the synchrony cut-off. Our results showed that the mean rate contrast within the synchronized regimen (15.08 spikes/s) was significantly greater (ANOVA, $P < 0.0001$) than that of the nonsynchronized regimen (7.70 spikes/s). The mean rate contrast across all modulation frequencies was 19.51 spikes/s. Thus relatively little of the variation in average spike rate occurs in the range of modulation frequencies where cells no longer synchronize to SAM signals, suggesting that "rate coding" in the nonsynchronized regime is relatively weak in macaque auditory cortex.

It is worth noting in this context that the synchronized regimen extended to much higher modulation frequencies that would be expected based on reports from anesthetized animals (Langner and Schreiner 1988). Thus in a few neurons, the synchronized regimen included all modulation frequencies in the standard testing battery (0.7 to 200 Hz). We calculated the percentage of significantly synchronized responses ($P < 0.001$ by the Rayleigh test) at each of the standard modulation frequencies and found that nearly 80% of all responses followed modulation frequencies ≤ 10 Hz (0.7: 77%; 1: 79%; 2: 84%; 5: 87%; 10: 80%), but progressively fewer were able to follow higher modulation frequencies (20: 55%; 50: 36%; 100: 23%; 200: 9%).

To facilitate comparison with other studies, we generated histograms of the rBMFs (Fig. 14B) and tBMFs (Fig. 14C) for all cells in the population. The joint distribution of rBMFs and tBMFs is indicated by the matrix in Fig. 14A. As we have argued, all neurons effectively default to a synchrony code at a sufficiently high modulation frequency, implying that syn-

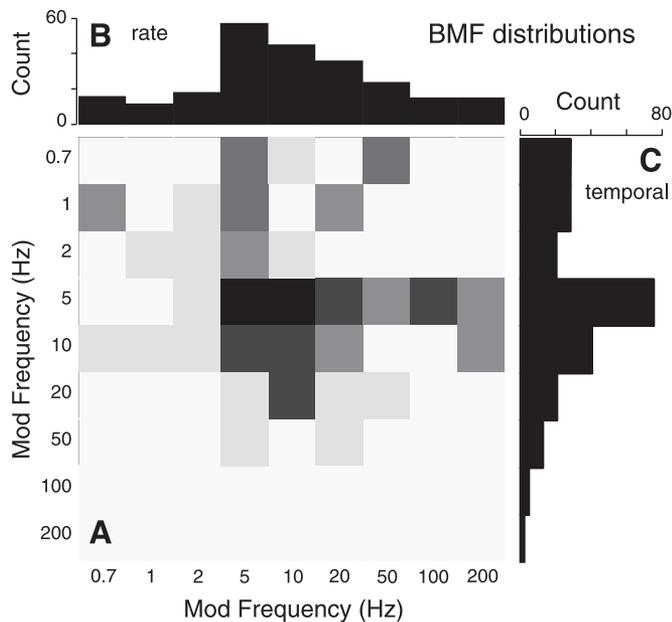


FIG. 14. Distributions of best modulation frequencies for both rate and VS favor SAM stimuli <20 Hz. Distribution of rBMFs is indicated by histogram in *B* and that of tBMFs based on VS in *C*. Matrix plot in *A* shows joint distribution. Only those cells ($n = 238$) tested using the standard battery of modulation frequencies (0.7, 1, 2, 5, 10, 20, 50, 100, and 200 Hz) were included. Each neuron contributed an rBMF and a tBMF to this graph.

chony cut-offs can be reliably measured with VS. The tBMF, however, is problematic because it tends to favor intermediate modulation frequencies (5–10 Hz) over very low modulation frequencies (>5 Hz). Nevertheless, it is clear that the majority of both rBMFs and tBMFs fall below 20 Hz, with peaks distributed around 5 Hz. It is also worth noting that the use of the BMF tends to exaggerate the effectiveness of the 5- and 10-Hz SAM relative to the 1- and 2-Hz SAM. When the discharge rates elicited by each modulation frequency are normalized with respect to the rBMF (i.e., the rBMF has a value of 1) and averaged over the population, the resulting curve describes a low frequency plateau with a relatively shallow peak at 5 Hz (population mean = 0.73). A1 neurons responded relatively poorly to modulation rates at 50 Hz or higher, with a minimum population mean of 0.51 at 200 Hz.

DISCUSSION

AI neurons encode changes in stimulus amplitude

Our results showed that the responses of AI neurons provide a high fidelity representation of amplitude changes. In particular, they showed that cortical neurons robustly encode the parameters defining the SAM stimulus, such as carrier level, modulation depth, and modulation frequency, in the temporal patterns of their responses. As a consequence, changes in any of these parameters typically result in clear and consistent changes in the shape of the MPH. In contrast to studies that have emphasized the ability of cortical neurons to extract a purely temporal modulation from SAM signals (Liang et al. 2002; Lu et al. 2001), our study revealed that cortical neurons instead “multiplex” the stimulus parameters that jointly define the instantaneous amplitude of modulated signals.

There are several reasons why the fidelity of the cortical representation of SAM signals may not have been appreciated

in earlier work. The principal reason is the widespread emphasis on temporal limits on the encoding of SAM, focusing particularly on the neural representation of modulation frequency embodied in the MTF. In more peripheral auditory structures, the shapes of MPHs tend to be sufficiently stereotyped that it is possible to interpret changes in the shapes of MTFs when SAM parameters such as modulation depth are varied (Joris and Yin 1992). The increased heterogeneity in tuning for SPL in central auditory structures, such as the increased prevalence of nonmonotonic RLFs (IC: Semple and Kitzes 1987; AI: Semple and Kitzes 1993a,b), results in increased heterogeneity of MPH response profiles (Moller and Rees 1986), limiting the value of simple response metrics such as vector strength and the tMTFs derived from it. By using detailed examples from individual neurons, we tried to show that the complex relationship between the stimulus envelope and the MPH response profile retains information about instantaneous SPL that a pure synchrony code for modulation frequency would not. We must stress that we have deliberately chosen examples where the demands of a synchrony code for modulation frequency conflict with those of an instantaneous rate code for amplitude. Although these examples were chosen to illustrate, as clearly as possible, the preferential encoding of amplitude, these cells are exemplary in how well they encode amplitude changes, not in the fact that it is amplitude that they encode.

The second reason that direct rate coding of amplitude changes may have been overlooked is the use of anesthesia in all but a few early physiological studies (Bieser and Muller-Preuss 1996; Brugge and Merzenich 1973; De Ribaupierre et al. 1972; Funkenstein and Winter 1973; Goldstein et al. 1959). In the range of modulation frequencies where cortical neurons fire multiple spikes per modulation cycle, MPH shapes often contain a range of response features that clearly reflect the interaction of the instantaneous amplitude of the SAM signal with the amplitude tuning of cortical neurons. At progressively higher modulation frequencies (>20 Hz), where cortical neurons typically fire less than one spike per modulation cycle (Fig. 6), the representation of SAM signals tends to converge on a single MPH peak, commonly corresponding to the rising phase of the stimulus envelope where the amplitude (in dB SPL) changes most rapidly. Under anesthesia, however, the rich representation of low modulation frequency stimuli is compromised by reduced discharge rates. Not only does this dramatically curtail the variety of MPH shapes that can be observed, but the fact that AI neurons are often limited to roughly one spike per cycle implies that rate based measurements such as the BMF will be artificially shifted in favor of higher modulation frequencies (e.g., 10 vs. 2 Hz), because more cycles per second will result in higher firing rates. Consequently, the qualitative differences in the representations of high and low modulation frequencies, which are dependent on differences in average spikes per cycle, may have escaped attention. The increased upper limit of synchronization observed in unanesthetized animals, which could reach several hundred hertz in our data sample, may have drawn attention away from the dramatic increase in spike per cycle firing rates at low modulation frequencies, which is a prerequisite for the form of “direct” amplitude coding we have described here.

Our early figures emphasized the responses of nonmonotonic units to extremely low modulation frequencies because

such examples are readily related to the amplitude waveform without necessitating corrections for group delay. There is no principal reason to believe that what the timing of cortical action potentials encode—a quantity related to instantaneous SPL (Heil and Irvine 1996; Heil and Neubauer 2003)—is different at higher modulation frequencies, when the relationship between the envelope and the MPH is strained by biophysical limits on temporal precision (Blackburn and Sachs 1989; Frisina et al. 1990) and instantaneous discharge rates. Nevertheless, the quality of the cortical representation of amplitude changes certainly degrades with increasing modulation frequency. We argue that the instantaneous rate code evident at the lowest modulation frequencies defaults to an apparent temporal code as the modulation period shortens (a synchrony code also necessarily has a limit, as evidenced by the distribution of synchrony cut-offs we observed).

For the experimenter, it becomes increasingly difficult to relate features of the MPH response profile to features of the stimulus envelope because small errors in the estimate of the group delay represent larger proportions of the modulation period. The brain is faced with a similar problem: a fixed variance in the group delay among a population of AI neurons will reduce the phase coherence of their responses by greater degrees at higher modulation frequencies (Fig. 15). Thus direct, rate-based amplitude coding is probably confined to a very limited range of modulation frequencies.

Lu et al. (2001) have argued that cortical neurons seem to use two distinct coding schemes for modulated signals: a

temporal code based on synchronized responses at low modulation frequencies, and a nonsynchronized rate code at higher modulation frequencies. We contend that the temporal code for modulation frequency described by these authors is fundamentally a rate code for amplitude. The nonsynchronized rate code they describe is identified with a largely distinct subpopulation of neurons that does not exhibit stimulus-synchronized discharges to modulated stimuli. In our data, relatively few neurons showed significant changes in their firing rates for modulation frequencies beyond their synchrony cut-offs, and all of these neurons exhibited synchronized discharges to at least one tested modulation frequency. Thus we were unable to find evidence of a nonsynchronized population of rate-coding neurons analogous to that observed in the marmoset, although we did find a number of cells that showed significant variation in firing rates within the range of modulation frequencies where the neurons no longer synchronized to the envelope. We also note that the rate-only classifier, using only information about average firing rate, performed significantly above chance in many neurons in our sample.

This discrepancy could reflect differences in species (macaque vs. marmoset), but it is more likely that differences in the stimuli used contributed. In our study, for example, all stimuli were presented through earphones, rather than in the free field, and the stimulus duration was extremely long (10 s) by auditory standards. The most likely explanation for the higher incidence of nonsynchronized, rate-coding neurons reported by Lu et al. (2001) is their use of click trains, rather than SAM

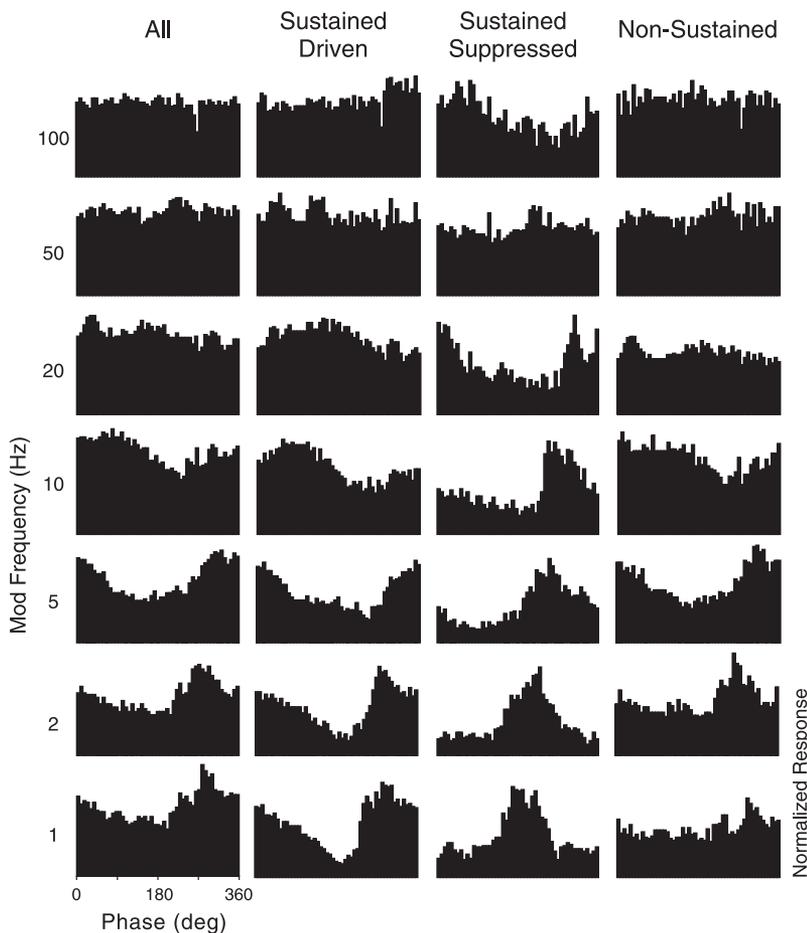


FIG. 15. Series of composite MPHs shows how normalized and summed activity of a population of neurons recorded from the same hemisphere of animal X responds to 60-dB SAM stimuli at best carrier frequency. Numbers at left indicate modulation frequency. *Rightmost 3 columns* of composite MPHs shows how normalized population average within a response class varies with modulation frequency. Sum of these MPHs across rows would reproduce composite MPHs shown in *leftmost column*. Driven, suppressed, and nonsustained composites are based on responses from 22, 11, and 33 single units, respectively.

tones. A Gaussian click train at a given rate (e.g., 10 clicks/s) may approximate the temporal features of SAM at the same modulation frequency (e.g., 10 Hz), but the spectra of the stimuli differ substantially, suggesting that their neural representation may be vastly different, both at the level of the population and that of individual neurons.

Nevertheless, Liang et al. (2002), using SAM stimuli, also reported that nonsynchronized rate-coding was observed in 30–40% of A1 neurons. The cardinal difference between their results and our own, however, is the dramatic difference in the incidence of synchronized responses. At 2 Hz, roughly 10% of the neurons in the marmoset (their Fig. 12C) evinced significant synchrony, compared with 84% in rhesus monkeys. This difference was maintained at higher modulation frequencies, where between 40 and 50% of marmoset A1 cells synchronized to 4-, 8-, and 16-Hz SAM, compared with 87, 80, and 55% of the neurons tested at 5, 10, and 20 Hz, respectively. Once again, differences in stimulus presentation (closed field vs. free field) and stimulus duration apply. The longer stimulus durations we used would seem to work against higher synchronization values, given the demonstration that spike timing dispersion increases with successive events in periodic click trains (Lu and Wang 2004). It is also conceivable that laminar differences could explain part of this discrepancy, because our recordings were distributed relatively evenly across cortical depth (see METHODS) rather than concentrated in the upper layers (Liang et al. 2002). It would be of great interest to know to what extent a transformation from temporal to rate-based coding of stimulus modulations occurs across cortical layers (Wang 2007). Finally, it is important to note that differences in sampling strategy may be relevant, because we presented SAM stimuli only to those neurons that had clear responses to tone pips, which may have biased our sample in favor of neurons with synchronous responses to SAM.

Changes in carrier frequency, level, and modulation depth produced clear and consistent changes in the shapes of the MPH, showing that cortical neurons are extremely sensitive to the details of the stimulus envelope and not just its modulation frequency. Summary measures such as the rBMF and tBMF, however, appeared to be relatively poor descriptors of the envelope coding properties of cortical neurons in the awake monkey. In fact, the full spike train and phase-only classifiers exhibited statistically equivalent performance in discriminating carrier level and modulation frequency, which shows that effectively “flattening” all cortical rMTFs had essentially no impact on the discrimination of stimulus identity for these SAM parameters (see also Wohlgemuth and Ronacher 2007).

Changes in MPH shape often resulted in minimal or even misleading changes in VS, and the distribution of firing rates within the modulation period was generally more informative than the distribution of firing rates across modulation frequency. For example, the phases of the responses to fully modulated stimuli at 20 and 60 dB in Fig. 7 are essentially inverted, but the impact on both rate and VS is minimal. Thus the shapes of the MTFs can be far less sensitive to stimulus changes than are the shapes of the MPHs. Liang et al. (2002) reported that rMTF and tMTF shapes are relatively invariant to changes in sound level and modulation frequency. In part, this could reflect the relative insensitivity of these metrics. It should also be noted that in a majority of the neurons tested in this way, SPL was varied by 20 dB or less, and modulation depth

by 25% or less, which may have made such changes difficult to detect. Nevertheless, even the shapes of the rMTFs and tMTFs evinced dramatic changes in some neurons (e.g., Fig. 8). In summary, our findings indicate that the coding of low modulation frequency SAM signals is more directly concerned with providing high fidelity representations of low frequency envelope features rather than a high synchrony representation of the dominant modulation frequency.

What mechanisms influence the cortical representation of the stimulus envelope?

Our data clearly showed that the transformation relating the stimulus envelope to the MPH response profile is highly nonlinear and often quite complex. Nevertheless, it was often possible to relate the shapes of cortical response profiles to the SPL profile of the modulation given some independent measure of the neuron's tuning for signal amplitude, such as the RLF. Knowledge of the neuron's steady-state response to an unmodulated control tone alone was useful for predicting qualitative features of composite MPHs. Although we have argued that AI codes stimulus amplitude, it is useful to remember that the responses of AI neurons are determined by changes in the afferent firing rates, which are related to the amplitude of acoustic signals by tuning properties operating at multiple levels of the ascending auditory pathway. These signals are certainly quite different from the modulating waveform or the logarithmic transform of the SAM signal in decibels. Simply put, we are not measuring the ability of the cortex to follow a 2-Hz sinusoidal modulation of its inputs when we present SAM at 2 Hz. Even in the auditory nerve, the spectrum of the MPH contains more than a single frequency peak (i.e., the MPH is not sinusoidal), and the peak corresponding to the modulation frequency is not necessarily the largest component of the response (Khanna and Teich 1989). MPH shapes become increasingly complex in the cochlear nucleus (Hirsh and Gibson 1976; Kim et al. 1990) and inferior colliculus (Krishna and Semple 2000; Rees and Palmer 1989). Although the reduction in cortical synchrony to high modulation frequencies has received much comment (Eggermont 1991, 1994; Gaese and Ostwald 1995; Schreiner and Urbas 1988), the high frequency content apparent in cortical responses to very low modulation frequency stimuli in awake animals has received far less attention. This feature of cortical responses is clearly an important component of the representation of sound envelopes in the communication range (<20 Hz). The RLF provides a useful estimate of how the stimulus envelope may be encoded by a given cortical neuron, particularly one with sustained responses to pure tones. Nevertheless, AI responses to SAM cannot be predicted by treating the RLF function as a look-up table that generates a firing rate for the instantaneous amplitude of the SAM signal at each point in the modulation cycle (but see Yates 1987).

Nonlinear tuning to stimulus amplitude must contribute to nonlinear response features prevalent in the shapes of MPHs. In addition, mechanisms sensitive to the time scale of the modulation period further shape the MPHs. In a recent review of SAM processing, Joris et al. (2004) asked “whether specialized neural mechanisms exist to extract AM information.” There may indeed be synaptic specializations that influence the processing of temporally modulated signals in particular cor-

tical regions (Eggermont 1999; Varela et al. 1997), and even particular classes of neurons within a single cortical area (Atzori et al. 2001). More generally, adaptive mechanisms sensitive to recent stimulus history appear to be operating whenever changes in the ongoing pattern of stimulation produce changes in the inputs to auditory neurons on timescales from tens to hundreds of milliseconds (Malone and Semple 2001; Malone et al. 2002; Sanes et al. 1998). Because cortical neurons typically fire many spikes per modulation cycle in the range of modulation frequencies crucial for speech intelligibility (e.g., 3–8 Hz: Drullman et al. 1994), such mechanisms are likely to shape the distribution of those spikes over the course of individual modulation periods (333–125 ms).

If the representation of SAM signals is carried by changes in instantaneous discharge rates, gain control mechanisms sensitive to firing rates will filter the temporal representation of dynamic signals. Using a ripple stimulus that contained energy at multiple modulation frequencies, Elhilali et al. (2004) showed that a simplified model of synaptic depression and facilitation could explain how phase locking to the fine structure of the stimulus could be “gated” by responses to slower modulations of the envelope. Synaptic dynamics (depression and facilitation) and intrinsic mechanisms (pyramidal cell calcium kinetics) have already been incorporated into models attempting to explain the shape of cortical MTFs (Eggermont 1999, 2002). It would be of interest to know whether such models could successfully predict the shape of cortical MPHs for cells where the static input–output response (the RLF) is known. For example, synaptic depression can account for a range of temporal response characteristics of primary visual cortical cells for spatial sinusoidal modulation, including both band-pass frequency-response curves (i.e., MTFs) and contrast-dependent phase shifts in the period histogram (i.e., MPH; Chance et al. 1998). Given the observation that a sizeable minority of cortical neurons can follow modulated rates in excess of 100 Hz, we speculate that the progressive downward shift of rBMFs and tBMFs in the ascending the auditory pathway has less to do with intrinsic synaptic limits on the preservation of temporal synchrony than it does with adaptive gain control mechanisms that instantiate sensitivity to recent stimulus context (Malone et al. 2002). Such mechanisms may enhance the representation of envelope features at the time scales most relevant for the processing of communication sounds.

Implications for perception

The notion that auditory cortex is geared toward the extraction and representation of modulation frequency is weakened by the observation that “AM selectivity [based on tMTFs] varies considerably among cortical neurons, but that overall selectivity is relatively poor” (Joris et al. 2004; cited in support are Eggermont 1999; Liang et al. 2002; Miller et al. 2002; Schreiner and Urbas 1988). If auditory cortex acted as a modulation filterbank (Green and Kay 1974; Kay and Matthews 1972), perfect synchrony would represent an effective neural code for modulation frequency but an extremely poor code for other features of the envelope. Our data revealed that low frequency amplitude modulations, which predominate in natural environments (Attias and Schreiner 1997; Singh and Theunissen 2003), as well as music and speech (Voss and

Clarke 1975), are those least likely to produce high-synchrony, unimodal MPHs invariant to changes in carrier level or modulation depth. Instead, the shapes of cortical MPHs exhibit their greatest diversity in this range of modulation frequency, both across neurons and within single neurons when parameters defining the SAM stimuli are varied.

This diversity endows cortical neurons with the ability to process envelope changes in multiple ways. Using noise-modulated tones, Moller and Rees (1986) argued that IC units “differ in the ways they transmit information about the envelope of a sound, as one might expect from the variety of rate-intensity functions which have been noted in the inferior colliculus.” This statement applies even more forcefully in AI, where many strongly nonmonotonic neurons signal decrements in sound level with sharp increments in firing rate, whereas monotonically tuned sustained responders suddenly cease firing. As our examples of composite MPHs showed, the diversity of cortical tuning for stimulus level presents a challenge for maintaining the phase coherence of responses to envelope changes across multiple neurons. Nevertheless, properly used, this diversity may aid in demarcating both the sudden onsets and offsets that are crucial for the intelligibility of speech and other communication sounds (Drullman et al. 1994; Smith et al. 2002).

The phase coherence of responses to modulated signals must certainly affect how a population of auditory cortical neurons represents the envelopes of complex stimuli. The composite MPHs depicted in Fig. 15 were not corrected for group delay, nor is it clear how the brain could “correct” for differences in group delay across neurons. To generate a phase coherent population representation of the signal, a common facet of the modulation must be selected as the basis for phase locking. If the common feature (e.g., the rising phase of the envelope) is related to amplitude, this suggests that the representation is most directly described as coding for changes in amplitude. In contrast, a synchrony code for modulation frequency is entirely compatible with arbitrary and independent response phases among the neurons in a population.

We observed considerable coherence in AI responses to SAM signals for low modulation frequency signals, particularly within response classes defined by responses to the unmodulated tones at equivalent carrier levels. The dominant response feature to 100% modulated SAM was in fact the response peak corresponding to the rapid rising phase of the envelope (the combination of driven and nonsustained responses outweighed the relatively smaller population of suppressed responses, whose response peaks coincided with the falling phase of the envelope, as in Fig. 15). Responses to modulated noise in the IC are also consistent with preferential responses to the rising phase of the stimulus envelope (Jones et al. 1987). Envelope peaks have been shown to be more important than troughs in psychophysical studies of speech intelligibility (Drullman 1995), perhaps reflecting the enhanced salience of their neural representation.

The phase coherence of responses across the cortical population may contribute to the perceptual boundaries for SAM signals. In human listeners, SAM tones are perceived as “fluctuations” in amplitude at very low modulation frequencies (less than ~5 Hz), remain resolved as individual periods ≤ 20 Hz, and are heard as “roughness” between 20 and 300 Hz (Kay 1982). These perceptual boundaries correspond to the changes

in MPH shapes we observed with increasing modulation frequency—units with sustained responses, in particular, tracked the instantaneous amplitude up to ~5–10 Hz, at which point most MPHs converged on a unimodal response peak. Modulation was apparent in the composite MPH up to ~20 Hz, at which point the modulated signal from units able to follow higher modulation frequencies was lost in the noise of nonsynchronized neurons or compromised as differences in group delay distributed their response peaks more widely in the modulation period.

Central auditory processing of a range of communication sounds, especially speech, depends critically on temporal information at low modulation frequencies. It has been argued that “temporal cues appear to be useful for speech intelligibility only up to about 20 Hz” (Shannon 2002). Experiments using “chimeric sounds,” which combine the envelope of one utterance with the fine structure of another, showed forcefully that envelope cues dominate speech recognition (Smith et al. 2002). Because speech has been shown to remain intelligible even when broken up into time-reversed sections approaching 100 ms, it has been proposed that “ultralow frequency” modulation envelopes of 3–8 Hz are the critical cues to intelligibility (Saberri and Perrott 1999). This notion is consistent with the average syllabic rate of 3–4 Hz, where modulation in speech is most prominent (Houtgast and Steeneken 1985). Ahissar et al. (2001) showed that single trial success for speech comprehension was best predicted by the correlation between the temporal envelopes (0–20 Hz) of the stimulus and the envelope of cortical activity obtained with magnetoencephalography. Normal and dyslexic children have been shown to differ significantly in their ability to detect the onsets of AM, and sensitivity to the shape of AM was a significant predictor of reading and spelling acquisition in both groups (Goswami et al. 2002).

Finally, the idea that cortical responses to very low modulation frequencies are special is consistent with a range of early psychophysical studies of modulation processing performed by Kay (1982). Square-wave AM waveforms were shown to adapt human detectability of sinusoidal AM above—but not below—10 Hz. Our results indicate that 10 Hz is the approximate limit for robust envelope tracking (i.e., for TS to exceed VS; see Fig. 11F) in rhesus macaque cortical neurons. More generally, Kay (1982) concluded that “coincident modulation rate is not of itself a sufficient stimulus for adaptation” (p. 924). These psychophysical studies support our argument, based on cortical physiology, that the processing of slowly modulated signals is better described as envelope shape discrimination than modulation frequency extraction.

ACKNOWLEDGMENTS

Present addresses of B. H. Scott: Laboratory of Neuropsychology, NIMH, 49 Convent Dr., Room 1B80, Bethesda, MD 20892; B. J. Malone, Dept. of Neurobiology, David Geffen School of Medicine, UCLA.

GRANTS

B. J. Malone was supported by National Institute of Mental Health Grant MH-12993-02. M. N. Semple was supported by the W.M. Keck Foundation. B. H. Scott was supported by National Institute of Deafness and Other Communication Disorders Grant DC05287-01 and a James Arthur Fellowship from New York University.

REFERENCES

Ahissar E, Nagarajan S, Ahissar M, Protopoulos A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response

patterns recorded from auditory cortex. *Proc Natl Acad Sci USA* 98: 13367–13372, 2001.

Ahissar M, Protopoulos A, Reid M, Merzenich MM. Auditory processing parallels reading abilities in adults. *Proc Natl Acad Sci USA* 97: 6832–6837, 2000.

Attias H, Schreiner CE. Temporal low-order statistics of natural sounds. In: *Advances in Neural Information Processing Systems 9*. MIT Press, 1997.

Attias H, Schreiner CE. Blind source separation and deconvolution: the dynamic component analysis algorithm. *Neural Comput* 10: 1373–1424, 1998.

Atzori M, Lei S, Evans DI, Kanold PO, Phillips-Tansey E, McIntyre O, McBain CJ. Differential synaptic processing separates stationary from transient inputs to the auditory cortex. *Nat Neurosci* 4: 1230–1237, 2001.

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

Blackburn CC, Sachs MB. Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *J Neurophysiol* 62: 1303–1329, 1989.

Brugge JF, Merzenich MM. Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J Neurophysiol* 36: 1138–1158, 1973.

Chance FS, Nelson SB, Abbott LF. Synaptic depression and the temporal response characteristics of V1 cells. *J Neurosci* 18: 4785–4799, 1998.

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

Creutzfeldt O, Hellweg FC, Schreiner C. Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res* 39: 87–104, 1980.

De Ribaupierre F, Goldstein MH Jr, Yeni-Komshian G. Cortical coding of repetitive acoustic pulses. *Brain Res* 48: 205–225, 1972.

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

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

Eggermont JJ. Rate and synchronization measures of periodicity coding in cat primary auditory cortex. *Hear Res* 56: 153–167, 1991.

Eggermont JJ. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. *Hear Res* 74: 51–66, 1994.

Eggermont JJ. The magnitude and phase of temporal modulation transfer functions in cat auditory cortex. *J Neurosci* 19: 2780–2788, 1999.

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

Elhilali M, Fritz JB, Klein DJ, Simon JZ, Shamma SA. Dynamics of precise spike timing in primary auditory cortex. *J Neurosci* 24: 1159–1172, 2004.

Foffani G, Moxon KA. PSTH-based classification of sensory stimuli using ensembles of single neurons. *J Neurosci Methods* 135: 93–107, 2004.

Frisina RD, Smith RL, Chamberlain SC. Encoding of amplitude modulation in the gerbil cochlear nucleus. I. A hierarchy of enhancement. *Hear Res* 44: 99–122, 1990.

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

Funkenstein HH, Winter P. Responses to acoustic stimuli of units in the auditory cortex of awake squirrel monkeys. *Exp Brain Res* 18: 464–488, 1973.

Gaese BH, Ostwald J. Temporal coding of amplitude and frequency modulation in the rat auditory cortex. *Eur J Neurosci* 7: 438–450, 1995.

Goldberg J, Brown P. Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* 32: 631–636, 1969.

Goldstein M, Kiang N-S, Brown R. Responses of the auditory cortex to repetitive acoustic stimuli. *J Acoust Soc Am* 31: 356–364, 1959.

Goswami U, Thomson J, Richardson U, Stainthorpe R, Hughes D, Rosen S, Scott SK. Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proc Natl Acad Sci USA* 99: 10911–10916, 2002.

Green GG, Kay RH. Channels in the human auditory system concerned with the wave form of the modulation present in amplitude and frequency-modulated tones. *J Physiol* 241: 29P–30P, 1974.

Hackett TA, Stepniewska I, Kaas JH. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475–495, 1998.

- Heil P, Irvine DR.** On determinants of first-spike latency in auditory cortex. *Neuroreport* 7: 3073–3076, 1996.
- Heil P, Neubauer H.** A unifying basis of auditory thresholds based on temporal summation. *Proc Natl Acad Sci USA* 100: 6151–6156, 2003.
- Hirsch HR, Gibson MM.** Responses of single units in the cat cochlear nucleus to sinusoidal amplitude modulation of tones and noise: linearity and relation to speech perception. *J Neurosci Res* 2: 337–356, 1976.
- Houtgast T, Steeneken HJM.** The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica* 28: 66–73, 1973.
- Houtgast T, Steeneken HJM.** A review of the MTF-concept in room acoustics. *J Acoust Soc Am* 77: 1069–1077, 1985.
- Jones JP, Stepnoski A, Palmer LA.** The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1212–1232, 1987.
- Joris PX, Louage DH, Cardoen L, van der Heijden M.** Correlation index: a new metric to quantify temporal coding. *Hear Res* 216–217: 19–30, 2006.
- Joris PX, Schreiner CE, Rees A.** Neural processing of amplitude-modulated sounds. *Physiol Rev* 84: 541–577, 2004.
- Joris PX, Yin TC.** Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am* 91: 215–232, 1992.
- Kajikawa Y, Hackett TA.** Entropy analysis of neuronal spike train synchrony. *J Neurosci Methods* 149: 90–93, 2005.
- Kay RH.** Hearing of modulation in sounds. *Physiol Rev* 62: 894–975, 1982.
- Kay RH, Matthews DR.** On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. *J Physiol* 225: 657–677, 1972.
- Khanna SM, Teich MC.** Spectral characteristics of the responses of primary auditory-nerve fibers to amplitude-modulated signals. *Hear Res* 39: 143–157, 1989.
- Kim DO, Sirianni JG, Chang SO.** 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, 1990.
- Krishna BS, Semple MN.** Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* 84: 255–273, 2000.
- Langner G.** Periodicity coding in the auditory system. *Hear Res* 60: 115–142, 1992.
- Langner G, Schreiner CE.** Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60: 1799–1822, 1988.
- Liang L, Lu T, Wang X.** Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol* 87: 2237–2261, 2002.
- Lu T, Liang L, Wang X.** Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci* 4: 1131–1138, 2001.
- Lu T, Wang X.** Information content of auditory cortical responses to time-varying acoustic stimuli. *J Neurophysiol* 91: 310–313, 2004.
- Malone BJ, Scott BH, Semple MN.** Processing of sinusoidal modulations of amplitude and frequency in the auditory cortex of alert rhesus macaques. *Soc Neurosci Abstr* 26: 957, 2000.
- Malone BJ, Scott BH, Semple MN.** Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques. *J Neurosci* 22: 4625–4638, 2002.
- Malone BJ, Semple MN.** Effects of auditory stimulus context on the representation of frequency in the gerbil inferior colliculus. *J Neurophysiol* 86: 1113–1130, 2001.
- Mardia K, Jupp P.** *Directional Statistics*. New York: Wiley, 2000.
- Middlebrooks JC.** Auditory cortex cheers the overture and listens through the finale. *Nat Neurosci* 8: 851–852, 2005.
- Miller LM, Escabi MA, Read HL, Schreiner CE.** Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87: 516–527, 2002.
- Moller AR, Rees A.** Dynamic properties of the responses of single neurons in the inferior colliculus. *Hear Res* 24: 2039–2158, 1986.
- Rees A, Moller AR.** Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hear Res* 10: 301–330, 1983.
- Rees A, Moller AR.** Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. *Hear Res* 27: 129–143, 1987.
- Rees A, Palmer AR.** 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, 1989.
- Saberi K, Perrott DR.** Cognitive restoration of reversed speech. *Nature* 398: 760, 1999.
- Sanes DH, Malone BJ, Semple MN.** Role of synaptic inhibition in processing of dynamic binaural level stimuli. *J Neurosci* 18: 794–803, 1998.
- Schreiner CE, Urbas JV.** Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 32: 49–63, 1988.
- Scott BH.** *Physiological Organization of Auditory Cortex in Awake Macaque* (PhD dissertation). New York: New York University, Center for Neural Science, 2004.
- Scott BH, Malone BJ, Semple MN.** Physiological delineation of primary auditory cortex in the alert rhesus macaque. *Soc Neurosci Abstr* 26: 957, 2000.
- Semple MN, Kitzes LM.** Binaural processing of sound pressure level in the inferior colliculus. *J Neurophysiol* 57: 1130–1147, 1987.
- Semple MN, Kitzes LM.** Binaural processing of sound pressure level in cat primary auditory cortex: evidence for a representation based on absolute levels rather than interaural level differences. *J Neurophysiol* 69: 449–461, 1993a.
- Semple MN, Kitzes LM.** Focal selectivity for binaural sound pressure level in cat primary auditory cortex: two-way intensity network tuning. *J Neurophysiol* 69: 462–473, 1993b.
- Shannon RV.** The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *Am J Audiol* 11: 124–127, 2002.
- Singh NC, Theunissen FE.** Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114: 3394–3411, 2003.
- Smith ZM, Delgutte B, Oxenham AJ.** Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416: 87–90, 2002.
- Swarbrick L, Whitfield IC.** Auditory cortical units selectively responsive to stimulus ‘shape’. *J Physiol* 224: 68P–69P, 1972.
- Ulanovsky N, Las L, Farkas D, Nelken I.** Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* 24: 10440–10453, 2004.
- Varela JA, Sen K, Gibson J, Fost J, Abbott LF, Nelson SB.** A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci* 17: 7926–7940, 1997.
- Varela JA, Song S, Turrigiano GG, Nelson SB.** Differential depression at excitatory and inhibitory synapses in visual cortex. *J Neurosci* 19: 4293–4304, 1999.
- Voss R, Clarke J.** 1/f noise in music and speech. *Nature* 258: 317–318, 1975.
- Wang X.** Neural coding strategies in auditory cortex. *Hear Res* 229: 81–93, 2007.
- Wang X, Lu T, Snider R, Liang L.** Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435: 341–346, 2005.
- Wohlgemuth S, Ronacher B.** Auditory discrimination of amplitude modulations based on metric distances of spike trains. *J Neurophysiol* 97: 3082–3092, 2007.
- Yates GK.** Dynamic effects in the input/output relationship of auditory nerve. *Hear Res* 27: 221–230, 1987.