

Srikantan S. Nagarajan, Steven W. Cheung, Purvis Bedenbaugh, Ralph E. Beitel, Christoph E. Schreiner and Michael M. Merzenich
J Neurophysiol 87:1723-1737, 2002.

You might find this additional information useful...

This article cites 64 articles, 25 of which you can access free at:
<http://jn.physiology.org/cgi/content/full/87/4/1723#BIBL>

This article has been cited by 11 other HighWire hosted articles, the first 5 are:

Neural Rate and Timing Cues for Detection and Discrimination of Amplitude-Modulated Tones in the Awake Rabbit Inferior Colliculus

P. C. Nelson and L. H. Carney
J Neurophysiol, January 1, 2007; 97 (1): 522-539.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Virtual Vocalization Stimuli for Investigating Neural Representations of Species-Specific Vocalizations

C. DiMattina and X. Wang
J Neurophysiol, February 1, 2006; 95 (2): 1244-1262.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Functional Organization of Squirrel Monkey Primary Auditory Cortex: Responses to Frequency-Modulation Sweeps

B. Godey, C. A. Atencio, B. H. Bonham, C. E. Schreiner and S. W. Cheung
J Neurophysiol, August 1, 2005; 94 (2): 1299-1311.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Effect of the Temporal Pattern of Contralateral Inhibition on Sound Localization Cues

G. Marsat and G. S. Pollack
J. Neurosci., June 29, 2005; 25 (26): 6137-6144.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Plasticity in Primary Auditory Cortex of Monkeys with Altered Vocal Production

S. W. Cheung, S. S. Nagarajan, C. E. Schreiner, P. H. Bedenbaugh and A. Wong
J. Neurosci., March 9, 2005; 25 (10): 2490-2503.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Medline items on this article's topics can be found at <http://highwire.stanford.edu/lists/artbytopic.dtl> on the following topics:

Physiology .. Cortical Neurons
Veterinary Science .. Auditory Cortex
Physiology .. Humans
Physiology .. Monkeys

Updated information and services including high-resolution figures, can be found at:
<http://jn.physiology.org/cgi/content/full/87/4/1723>

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 August 20, 2007 .

Representation of Spectral and Temporal Envelope of Twitter Vocalizations in Common Marmoset Primary Auditory Cortex

SRIKANTAN S. NAGARAJAN,¹ STEVEN W. CHEUNG,² PURVIS BEDENBAUGH,³ RALPH E. BEITEL,² CHRISTOPH E. SCHREINER,² AND MICHAEL M. MERZENICH²

¹Department of Bioengineering, University of Utah, Salt Lake City, Utah 84112-9458; ²Coleman Memorial Laboratory and W. M. Keck Center for Integrative Neuroscience, Department of Otolaryngology, University of California, San Francisco, California 94143-0732; and ³Departments of Neuroscience and Otolaryngology, University of Florida, Gainesville, Florida 32610-0244

Received 31 July 2001; accepted in final form 3 December 2001

Nagarajan, Srikantan S., Steven W. Cheung, Purvis Bedenbaugh, Ralph E. Beitel, Christoph E. Schreiner, and Michael M. Merzenich. Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J Neurophysiol* 87: 1723–1737, 2002; 10.1152/jn.00632.2001. Cortical sensitivity in representations of behaviorally relevant complex input signals was examined in recordings from primary auditory cortical neurons (AI) in adult, barbiturate-anesthetized common marmoset monkeys (*Callithrix jacchus*). We studied the robustness of distributed responses to natural and degraded forms of twitter calls, social contact vocalizations comprising several quasi-periodic phrases of frequency and AM. We recorded neuronal responses to a monkey's own twitter call (MOC), degraded forms of their twitter call, and sinusoidal amplitude modulated (SAM) tones with modulation rates similar to those of twitter calls. In spectral envelope degradation, calls with narrowband channels of varying bandwidths had the same temporal envelope as a natural call. However, the carrier phase was randomized within each narrowband channel. In temporal envelope degradation, the temporal envelope within narrowband channels was filtered while the carrier frequencies and phases remained unchanged. In a third form of degradation, noise was added to the natural calls. Spatiotemporal discharge patterns in AI both within and across frequency bands encoded spectrotemporal acoustic features in the call although the encoded response is an abstract version of the call. The average temporal response pattern in AI, however, was significantly correlated with the average temporal envelope for each phrase of a call. Response entrainment to MOC was significantly correlated with entrainment to SAM stimuli at comparable modulation frequencies. Sensitivity of the response patterns to MOC was substantially greater for temporal envelope than for spectral envelope degradations. The distributed responses in AI were robust to additive continuous noise at signal-to-noise ratios ≥ 10 dB. Neurophysiological data reflecting response sensitivity in AI to these forms of degradation closely parallel human psychophysical results on the intelligibility of degraded speech in quiet and noisy conditions.

INTRODUCTION

The processing of behaviorally relevant species-specific communication sounds such as speech, monkey calls, and birdsong, both under quiet and naturalistic noisy background conditions, is an important aspect of auditory behavior in vocalizing species. Complex species-specific vocalizations can be parametrically de-

composed into acoustic and perceptual features such as intensity, spectral envelope, temporal envelope, carrier frequencies, phase, and pitch. Behavioral and neuronal sensitivities, at different levels of the central auditory pathway, to these features and combinations thereof that are present in complex vocalizations are now beginning to be understood. For example, results from human psychophysical experiments with speech stimuli have indicated that temporal envelope modulation features, rather than spectral envelope features, are extremely critical for identification and recognition of speech under both quiet and noisy background conditions (Drullman 1995a; Greenberg and Arai 2001; Kingsbury et al. 1998; Shannon et al. 1995).

Neurophysiological experiments in bird and bats with complex stimuli have elucidated some organizing principles in the auditory forebrain such as spectral, temporal combination sensitivity in discrete cortical areas and maps. For example, in zebra finches, high-level central neurons in the auditory forebrain responsible for vocal learning and adult vocalization discrimination are sensitive to temporal rather than spectral cues in their stereotypical calls (Margoliash and Fortune 1992; Theunissen and Doupe 1998). In mustached bats, cortical areas that are involved in echolocation also participate in the processing of species-specific communication sounds; neurons are sensitive to a specific combination of features within complex stimuli (Esser et al. 1997; Kanwal et al. 1994; Ohlemiller et al. 1996). This higher-order combination sensitivity to features within complex stimuli emerges hierarchically within the auditory system (Lewicki and Arthur 1996). Generalizations of these organizing principles to other mammals remain to be established (Rauschecker 1997; Rauschecker et al. 1995; Tian et al. 2001).

The basic organizational and functional features of simpler acoustic stimuli in the primary auditory cortex (AI) of ferrets, cats, and primates are now well understood (Calhoun and Schreiner 1998; Cheung et al. 2001; Eggermont 1991; Heil et al. 1992; Imig et al. 1977; Lu et al. 2001; Mendelson et al. 1997; Recanzone et al. 1999; Schreiner 1991; Schwarz and Tomlinson 1990). In these species, the representation and processing of complex spatiotemporal stimuli—especially

Address for reprint requests: S. S. Nagarajan, Dept. of Bioengineering, University of Utah, 20S 2030 East BPRB 506D, Salt Lake City, UT 84112-9458 (E-mail: sri@utah.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.

stimuli that are behaviorally relevant to the animal—are less well understood, especially in primates. Early studies in squirrel monkeys have demonstrated that although AI neurons respond vigorously to species-specific vocalizations, they do not exhibit a high degree of specificity or selectivity to these stimuli (Glass and Wollberg 1979, 1983a,b; Manley and Muller-Preuss 1978; Newman and Wollberg 1973a,b; Pelleg-Toiba and Wollberg 1991; Winter and Funkenstein 1973; Wollberg and Newman 1972).

Other studies suggest that the encoding of complex species-specific vocalizations is represented in discharge patterns of distributed neuronal populations (Creutzfeldt et al. 1980; Gehr et al. 2000; Rauschecker 1997, 1998a,b; Rauschecker et al. 1995; Rotman et al. 2001; Tian et al. 2001; Wang et al. 1995). Evidence in favor of this “distributed-encoding” hypothesis has been demonstrated perhaps most clearly in the common marmoset (Wang et al. 1995), where the spectrotemporal discharge patterns of spatially distributed neuronal populations in the AI was correlated with the spectrotemporal acoustic patterns of complex natural vocalizations. Interestingly, a majority of neurons in AI exhibit a preference for the natural time scale or modulation frequency of complex vocalizations (Wang et al. 1995) and cortical neurons that are highly effectively excited by vocalizations presented in the forward (natural) direction respond very poorly to reversed-direction forms. More recently, direct evidence for distributed encoding of species-specific vocalizations in AI has also been obtained in cats (Gehr et al. 2000; Rotman et al. 2001).

In the current study, we further examine this distributed encoding hypothesis where coherent neuronal subpopulations contribute to the representation of complex stimuli in the marmoset AI. The goals of the current study are to compare quantitatively neural responses to vocalizations with responses to sinusoidal amplitude modulated (SAM) tones: both stimuli have comparable periodicity; to evaluate the robustness of patterns of distributed response profiles to natural and degraded vocalizations, synthesized by changing either the spectral or temporal envelopes; and to measure the response sensitivity of distributed populations of AI neurons to natural and degraded vocalizations in the presence of background noise.

METHODS

Vocalization recordings and degradations

Vocalizations were recorded similar to those described earlier (Wang et al. 1995). Briefly, 1–2 weeks before each experiment, the marmoset under study was placed in a separate individual cage in the colony. A digital tape recorder with a sampling rate of 48 kHz, and a 16-bit A/D converter was used to record 2 h of this monkey’s vocalizations each day. These recordings were then scanned for loud and clear calls, and candidate vocalizations were subsequently transferred to a computer for editing using MATLAB. Twitter calls, the specific vocalizations under study here, were commonly produced by the monkeys in the colony. Calls vocalized by individual marmosets were highly stereotypical with stable spectrotemporal features. A stereotypical twitter call for each animal was chosen from a set of 10–30 vocalizations. The representative twitter call for each animal was referred to as the natural call or the “monkey’s own call” (MOC). For each MOC, the vocalization “phrase-frequency” f_p , was computed using methods similar to those described by Wang and colleagues (1995). Briefly, the Hilbert transform of the call was computed to obtain the analytical signal. The absolute value of the analytical signal

was low-pass filtered with a finite-impulse response (FIR) filter with a cutoff of 100 Hz to obtain an estimate of the envelope of the call. The power spectrum of the envelope was then estimated using multi-taper spectral estimation methods. The frequencies at which multi-tapered spectrum had a maximum was chosen to be the phrase frequency. The MOC was then systematically digitally degraded in two forms using filter-bank analysis-by-synthesis algorithms. A filter bank was constructed from FIR filters designed with >50 dB attenuation in the stop-band and <0.001 dB ripple in the pass-band. The filter bank was designed to give perfect reconstruction of stimuli between frequencies of 4–20 kHz, which covered the spectral energy of all naturally occurring twitter calls (see Fig. 2A). The number of filters in the filter-bank varied from 2 to 16 bands, corresponding to a bandwidth decrease in each filter from 8 to 0.5 kHz, respectively. Each narrowband signal from the filter bank was decomposed into an envelope and a carrier waveform.

In spectral envelope degradation, calls were resynthesized with intact envelopes within narrowband channels of varying bandwidths but with the carrier signals modified to band-limited noise for each channel. Multi-tapered spectrogram examples of such a set of degraded calls is shown in the *left column* of Fig. 1 for 16-band (Fig. 1B), 8-band (Fig. 1C), 4-band (Fig. 1D), and 2-band (Fig. 1E) filters in the filter bank. Stimuli were generated parametrically as a function of the number of bands in the filters. In another form of degradation, within each of 32 narrowband channels, the temporal envelope was modified. However, the carrier frequencies and phases were kept identical to the MOC. This form of MOC manipulation was referred to as “temporal envelope degradation.” Multi-tapered spectrogram examples of these degraded calls are shown in the *right column* of Fig. 1, *F–J*. The temporal envelope was band-pass filtered between 2–30 Hz (BP2-30 call) in Fig. 1F, low-pass filtered at 4 Hz (LP4 call) in Fig. 1H, low-pass filtered at 10 Hz (LP10 call) in Fig. 1G, and high-pass filtered at 60 Hz (HP60) in Fig. 1I.

Stimulus generation and delivery

In addition to natural and degraded calls, the stimulus ensemble in our experiments included tone pips, SAMs, clicks, and calls presented in continuous background noise. Tone pips (50-ms duration, 3-ms linear ON-OFF ramp) were used to derive frequency response areas for each recording site. Tone pips were generated on a digital signal processor (TMS 32010). The characteristic frequency (CF), the frequency of a tone to which the neuron is most sensitive, was first estimated using manually adjusted tonal frequencies and intensities. Subsequently, frequency-intensity “response area” was derived by randomly presenting tones at 15 different intensities and 45 different frequencies (ranging over 2–4 octaves) centered at the estimated CF at an inter-stimulus interval of 400–1,000 ms. SAM tones (500-ms duration, 55 dB-A, modulation depth = 100%) were generated by using the estimated CF of each recorded unit or cluster as the carrier frequency and were presented at 10 modulation rates between 2 and 20 Hz, randomly interleaved. Periodic click train sequences of constant 500-ms duration were also generated and presented to one animal in our study. The number of clicks ranged from a single click to 19 clicks over 500 ms to generate stimulus rates of 2–38 Hz in 4-Hz steps. Click stimuli were biphasic, with 200 μ s per phase. A particular click train sequence was delivered 15 times with a pause of 1–2 s between successive click train sequences.

Vocalization stimuli (55 dB-A) were generated using a Silicon-Graphics workstation. Natural and degraded vocalizations were also presented in continuous white noise. Continuous background noise was generated by a General Radio Model 1390-B noise generator at fixed intensities of 57 to 27 dB-A corresponding to signal-to-noise ratios (SNR) of -2 to 28 dB. All experiments were conducted in a double-walled soundproof room (IAC). Stimuli were presented through a STAX-headphone enclosed in a small chamber that was connected through a sealed tube into the external acoustic meatus of

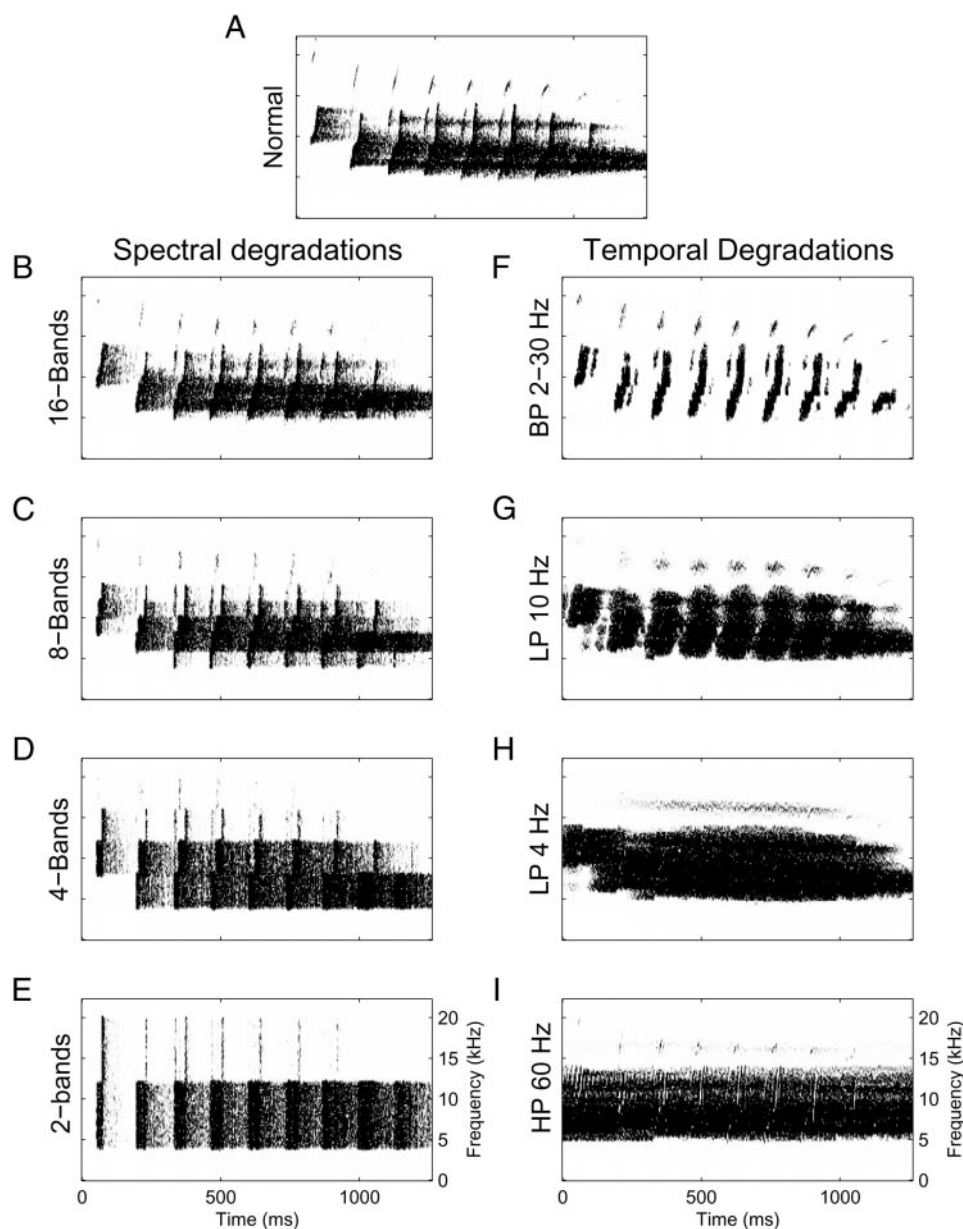


FIG. 1. Spectrograms of a monkey's own natural call (MOC) and degraded twitter calls. *Left*: systematic degradation of the spectral envelope, while preserving the temporal envelope and randomizing the carrier phase of the call within a specific number of frequency bands. *A*: MOC, *B*: 16 band call, *C*: 8 band call, *D*: 4 band call, and *E*: 2 band call. *Right*: systematic degradations of the temporal envelope of a twitter call while preserving the spectral envelope and carrier information. *F*: band-pass filtered between 2 and 30 Hz. *G*: low-pass filtered at 10 Hz. *H*: low-pass filtered at 4 Hz, and *I*: high-pass filtered at 60 Hz.

the contralateral ear. Both vocalizations and SAM tones were randomly interleaved and presented with a silent period of 1 s plus a jitter of ≤ 500 ms that was uniformly distributed.

Surgical preparation

All procedures described in the following text were conducted in accordance to protocols approved at the University of California at San Francisco that followed animal guidelines established by the National Institutes of Health. Adult marmosets were anesthetized with a mixture of halothane (2%)-oxygen (48%)-nitrous oxide (50%) to induce a surgical level of anesthesia. The skin overlying the trachea, stereotaxic pin sites, and the scalp was injected with a local anesthetic (lidocaine 2%). Tracheotomy was performed to secure the airway. An intravenous (iv) cannula was placed into the saphenous vein for delivery of pentobarbital sodium (15–30 mg/kg iv). The level of anesthesia was titrated to effect throughout the experiment. Lactated Ringer solution with 5% dextrose and 20 meq/l KCl was infused (6–8 ml \cdot kg $^{-1}$ \cdot h $^{-1}$) to maintain body hydration, metabolic homeostasis,

and cardiovascular function. A third generation cephalosporin that crosses the blood brain barrier was administered intravenously to prevent infection. A warming blanket with feedback control was used to maintain proper body temperature. During the entire experiment, the animal's level of anesthesia, fluid status, urine output, core temperature, and cardio-pulmonary functions were monitored.

A craniotomy was performed to expose the dura around auditory cortex. The overlapping cortex was then exposed by reflecting the dural flap and was maintained under a thin layer of silicone oil. A video image of the cortical zone was captured and stored in a computer to guide the positioning of microelectrode penetrations relative to the vasculature on the surface of the cortex.

Recordings

Double-barrel tungsten microelectrodes coated with parylene (FHC) with impedances of 1–2 M Ω at 1 kHz were introduced orthogonal to the cortical surface with a hydraulic microdrive (Kopf). In each animal, responses were sampled from ~ 50 –75 cortical sites per

hemisphere. The microelectrodes were separated by 250–300 μm and controlled by a single microdrive. All recordings were obtained from the tonotopically organized marmoset AI. Neuronal activity of small groups of neurons (multi-units) were recorded from each electrode at depths of 700–900 μm from the cortical surface, corresponding to cortical layers IIIb and IV. A window discriminator (BAK DIS-1) was used to isolate action potentials relative to background noise. Spike times were recorded in a computer and raw waveforms were stored on digital tapes for off-line analysis. The goal of these experiments was to obtain responses from a large number of AI neurons in a single animal, both for accurate distributed reconstructions and for efficient usage of animals. Therefore we placed less emphasis on the fact that all our recordings were multi-units, not single units. Off-line spike sorting of our data has also been difficult because of the considerable overlap between spikes from the neurons within our clusters. We attribute this to the fact that neurons in each cluster appear to encode similar information about stimuli.

Data analysis

Data collected from 589 sites recorded in left and right hemispheres in four marmosets (*M121*: left, 101 sites, and right, 64 sites; *M213*: left, 79 sites, and right, 40 sites; *M403*: left, 82 sites, and right, 35 sites; *M379*: left, 102 sites, and right, 86 sites) are presented. No systematic differences were observed between hemispheres and data were pooled across hemispheres in each animal. From the multi-unit response to tone pips at varying intensities and frequencies, FRA were obtained at all recording sites from which the CF, minimum response threshold and tuning bandwidths (Q10 and Q40) of each site were measured (Schreiner and Mendelson 1990). The distribution of CF's sampled from the four animals in this study is shown in Fig. 2B. This sampling distribution is well matched to the average spectral energy of all vocalizations used in the study (Fig. 2A).

Response properties to vocalization stimuli were first analyzed by accumulating the spike counts for each electrode penetration in 1- and 2-ms bins to form peristimulus time histograms (PSTHs). Subsequently, multi-taper spectral estimation techniques were applied to determine spectral energies in the PSTHs. This method offers an

improvement over direct FFT transforms applied to the PSTH data to determine estimates of spectral power (Thomson 1982). An example of this analysis is shown in Fig. 3. Figure 3A, *top*, shows the spike rasters from a multi-unit recording with CF = 6.97 kHz in response to MOC. Figure 3, *bottom*, shows the PSTH, which illustrates the phasic firing response to each phrase of the twitter call. Figure 3B shows a spectral estimate of the PSTH computed for a time-bandwidth factor of 1.5 at frequencies ranging from 0 to 50 Hz, from which a response strength measure (R_s) was calculated by averaging the magnitude spectrum over a small band of frequencies around the vocalization phrase rate, f_v . The vocalization phrase rate indicates the average inter-phrase frequency of each call and was computed using procedure similar to those described in Wang et al. (1995). Other peaks in the spectral estimate appear at harmonic frequencies to the vocalization phrase frequency and were not used in this analysis. The magnitude spectrum is expressed in a dB scale relative to 1 spikes²/s. For this example, the response strength (R_s) to a call is 32.21 dB *re.* 1 spikes²/s. A similar response strength measure was also derived from the PSTHs to SAM and click-train stimuli wherein the response strength (R_s) at each modulation rate was derived from the mean spectral estimate of the PSTH at the modulation rate of the SAM stimulus and click train sequence, respectively. The response synchronization measure used here is similar to the “synchronized discharge rate” measure developed by Wang et al. (1995). Two differences between our analysis and that of Wang and colleagues is first, we use multi-taper spectral estimation methods to estimate the spectral-power at the phrase frequency. Second, we express this measure in dB (log) units, typical for any measure of spectral power.

Two additional measures were used to characterize the entrainment response to vocalizations. First, the mean driven spike rate (MDR) over the entire epoch of the stimulus was determined. Second, the mean driven spike-rate/phrase (MDP) of the stimulus was computed. The MDP was the average number of spikes within a 15- to 40-ms window following the onset of each phrase of the twitter call. To compare these response measures across conditions, one-way ANOVA statistical tests were performed with stimulus condition as a factor and the response measures as the dependent variable. Data for these analyses were pooled across animals and electrode penetrations.

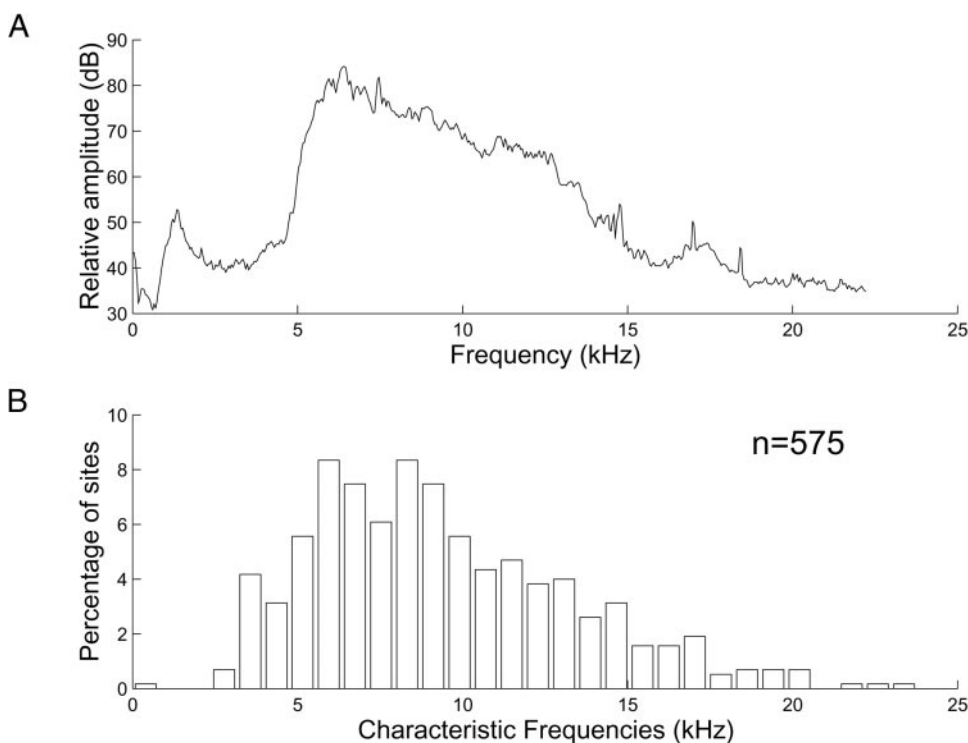


FIG. 2. Magnitude spectrum of twitter calls and sampling density of characteristic frequencies. *A*: the average spectrum of the twitter calls from 4 marmosets used in this study. *B*: a distribution of characteristic frequencies recorded from primary auditory cortical neurons (AI) in the 4 studied marmosets.

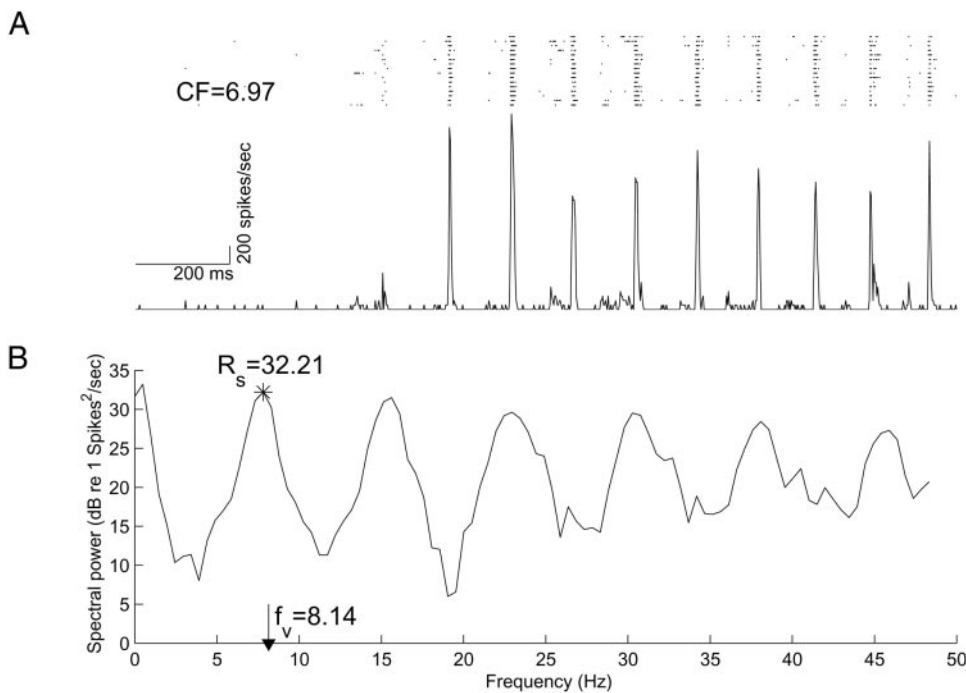


FIG. 3. Representative response of an AI neuron to MOC. *A*: spike rasters of a representative multi-unit recording in response to MOC and the peri-stimulus time histogram (PSTH) corresponding to these spike rasters. The characteristic frequency of this unit response was 6.97 kHz. *B*: multi-tapered spectral estimate of the PSTH. *, the entrainment strength at the modulation frequency of the stimulus.

Population PSTHs, neurograms, and mean-spectral representations were also computed by integrating the responses across electrode penetrations. Population PSTHs were computed by averaging the PSTHs from individual multi-unit recordings across the entire set of recordings obtained from a single animal. Neurograms are linear time-frequency maps of the distributed responses to a particular stimulus. The vertical axis is characteristic frequency, the horizontal axis is time, and the gray scale level encodes the response strength. Neurograms were computed by ordering PSTHs from individual electrode penetrations according to CF. Linear interpolation followed by a one-step three-tap median filtering was then performed at each time bin along the CF dimension to obtain a response map with uniform time-frequency tiling. No smoothing was performed along the time axis. Mean-spectral representation was computed by averaging the neurograms across time. Examples of population PSTHs, neurograms and mean-spectral representations for MOC are shown in Figs. 5 and 6. Finally, in addition to computing a neurogram, we also compute correlation-matrices of the responses in AI to MOC. This matrix comprises the correlation-coefficients between the responses of different neurons with different characteristic frequencies (Fig. 7, *A–D*). Analogous calculations were also performed on the multi-tapered spectrogram of the stimuli to compute the cross-frequencies correlations present in a call (see Fig. 7, *E–H*).

RESULTS

Comparison of responses to SAM stimuli, click-trains sequences, and MOC

First, we address the relationship between responses to MOC and responses to SAM stimuli of comparable modulation rates. The vocalization phrase frequencies were comparable across animals at 7.7 ± 0.53 (SE) Hz. Figure 4*B* shows the distribution of strength of response synchronization to MOC (mean $R_s = 13.2$ dB), and Fig. 4*A* shows a corresponding distribution for the strength of response synchronization to SAM stimuli with a modulation frequency of 8 Hz (mean $R_s = 7.7$ dB). A scatter plot between the responses to MOC and the responses to SAM stimuli across the population is shown in

Fig. 4*D*. Responses to vocalizations and to SAM stimuli were correlated ($r = 0.45$, $P < 0.001$). At the same time, these data indicate that only 22% of the variance in the responses to MOC across the population of neurons recorded in AI can be accounted for by the temporal modulation transfer function characteristics of the neurons. Responses to SAM stimuli of comparable modulation frequencies typically underestimate the strength of the response to vocalizations. Interestingly, the correlation between the best modulation frequency to SAM and the strength of its response to vocalizations was not significant ($r = -0.03$, data not shown). To account for whether the lack of spectral content in SAM tones when compared with MOC could account for this moderate correlation between SAM stimuli and MOC, the correlation between the responses to click trains at comparable periodicities and the responses to MOC were computed. Figure 4*C* shows the entrainment response strength to click trains (mean $R_s = -0.938$). This data indicated that only 13% of the variance in MOC responses could be accounted for by the responses to clicks (Fig. 4*E*). Furthermore, it is seen that the responses to click trains significantly underestimate the responses to MOC.

Distributed responses to MOC

Several features of the distributed response to a MOC in individual animals were elucidated in population PSTHs, neurograms, and mean-spectral-representations derived from >100 multi-unit recordings in each individual marmoset. Examples of data each from two different monkeys (*M213* and *M403*) are shown in Figs. 5 and 6. Population PSTHs, neurograms, and the mean-spectral-representations are plotted in *A* (for a complete MOC) and *C* (for the 1st 2 phrases of the MOC). The time-aligned spectrograms, mean temporal envelope and spectral envelopes of the MOC are shown in *B* (complete call) and *D* (1st 2 phrases). An examination of the neurograms reveals that the distributed cortical responses rep-

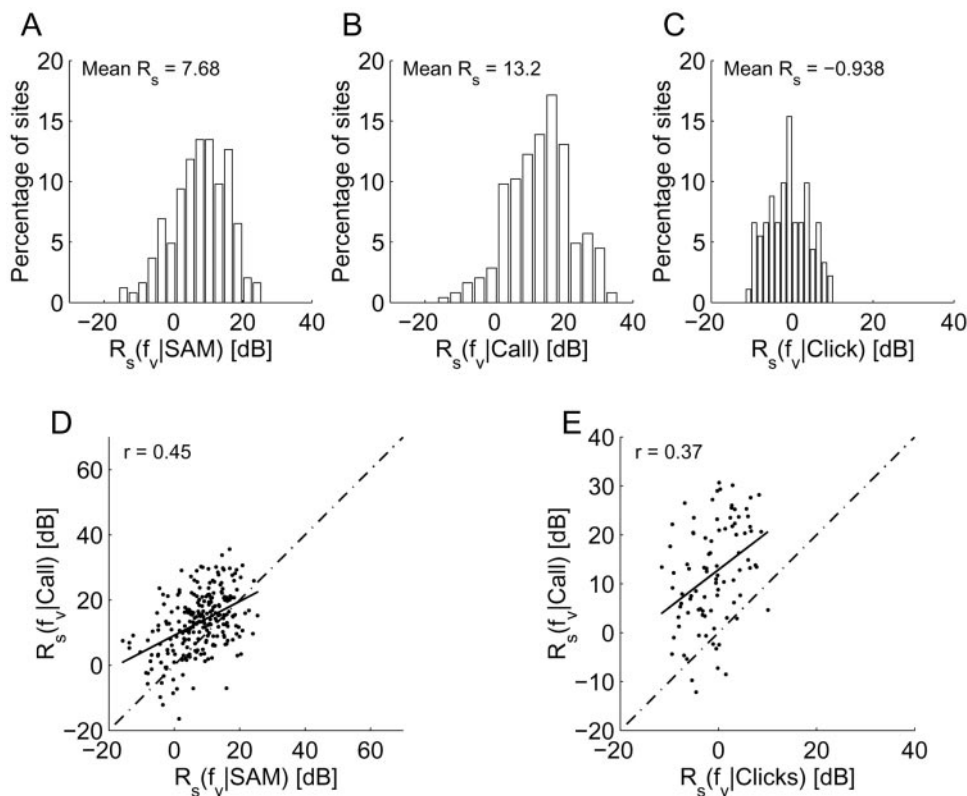


FIG. 4. Responses to a MOC, SAM, and click-train stimuli. *A*: the distributions of entrainment response strengths to twitter calls at the phrase frequency of the call. *B*: the distribution of entrainment response strengths to SAM stimuli at the vocalization phrase frequency, f_v . *C*: the distribution of entrainment response strengths to click-train sequences at the vocalization phrase frequency, f_v . *D*: scatter plot of each units response to SAM and its response to twitter calls. The thick line is the regression fit. The dashed line is the line that represents diagonal. *E*: scatter plot of each units response to click-train sequences and its response to twitter calls. The dashed line is the diagonal and the thin line is the regression fit.

resent the main spectrotemporal features of the calls albeit in an abstracted manner (Figs. 5A and 6A). In particular, frequency shifts of peak energy regions and details of intra-phrase and inter-phrase timing are reflected in the neurograms. A more detailed examination indicates that some of the features of the call spectra are distorted in the distributed cortical encoding through long- and short-range response synchronization. For example, the first phrase of the spectrogram of the

MOC in Fig. 5D exhibits a chirp. In the neurograms, the timing of this chirp is encoded in the latency of the peak firing-rate across corresponding recording sites (see Fig. 5C). However, the chirp rate representation is compressed in time, an indication of increased short-range synchronization. In general, the dispersion in firing latencies is smaller than the duration of FM sweeps (Figs. 5 and 6). Similarly, piecewise-linear segments of the FM sweep in MOCs appear to be encoded in the synchro-

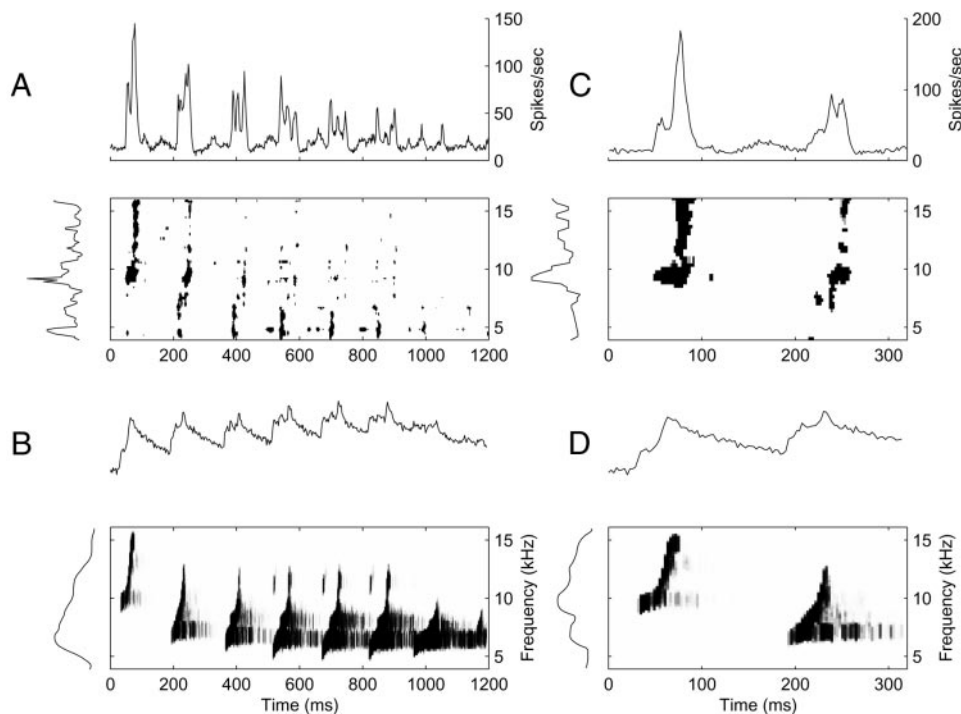


FIG. 5. Responses of M213 to MOC. *A*, *top*: population peristimulus time histogram (PSTH) to entire call. *A*, *bottom left*: mean-spectral representation of entire call. *A*, *bottom right*: neurogram of entire call. *B*, *top*: average temporal envelope of entire call. *B*, *bottom left*: average spectral envelope of entire call. *B*, *bottom right*: average spectrogram of the entire call. *C*, *top*: population PSTH to 1st 2 phrases of MOC. *C*, *bottom left*: mean-spectral representation of 1st 2 phrases of MOC. *C*, *bottom right*: neurogram of 1st 2 phrases. *D*, *top*: average temporal envelope of 1st 2 phrases. *D*, *bottom left*: average spectral envelope of 1st 2 phrases. *D*, *bottom right*: average spectrogram of the 1st 2 phrases.

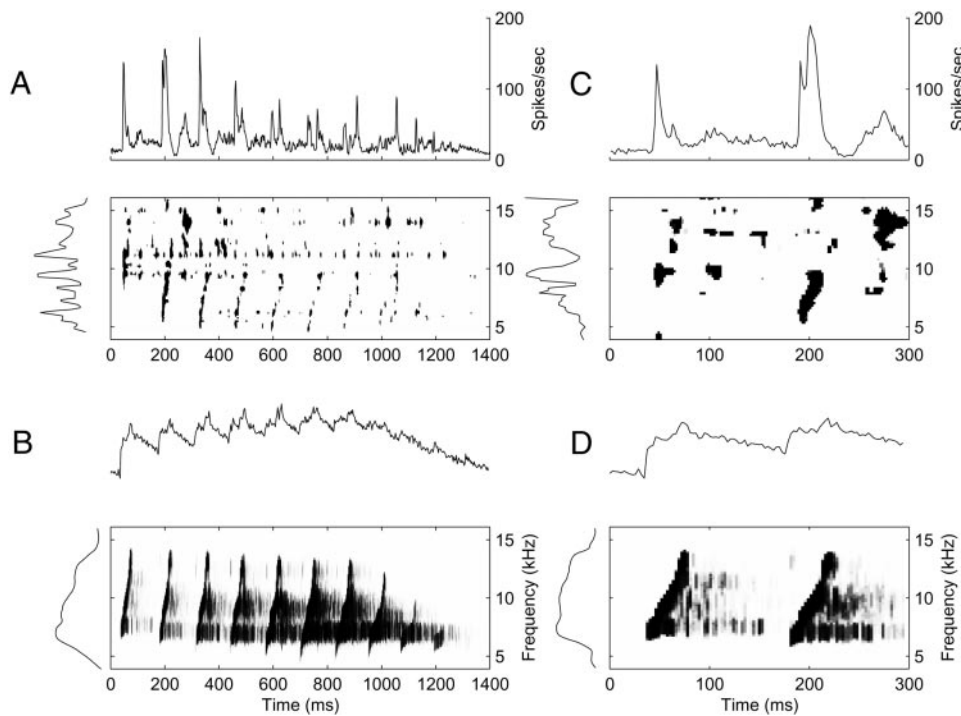


FIG. 6. Responses of *M403* to MOC. Legend as in Fig. 5.

nous firing of different populations of neurons in AI (Figs. 5C and 6C).

We also compute the correlation between the responses to MOC of different neurons with different characteristic frequencies. This is shown for the four animals of our study in Fig. 7, A–D. From this analysis, it is clear that long-range synchronizations (>1 octave in CF difference) and short-range correlations are observed across the distributed population of AI in response to a call. The short-range correlations are expressed by the widening of the diagonal in the correlation comparison. Long-range correlations are seen as separate regions off the diagonal in Fig. 7, A–D. Interestingly, both the short-range and long-range correlations between the responses are rather dif-

ferent from across-frequency stimulus-based correlations for three of the four animals studied (see Fig. 7, E–H).

In addition to comparisons between the spectrogram and neurogram and to examining the cross-frequency correlations, we also compared the average temporal envelope of the calls with the population PSTH and the average spectral envelope with the mean-spectral representation. The goal here was to examine the encoding of the stimuli in the average temporal and spatial firing patterns across AI, quantified by examining the correlation coefficients as outlined in the following text. For each phrase of a MOC, correlation-coefficients were computed between the population PSTH and the average temporal envelope, and the mean-

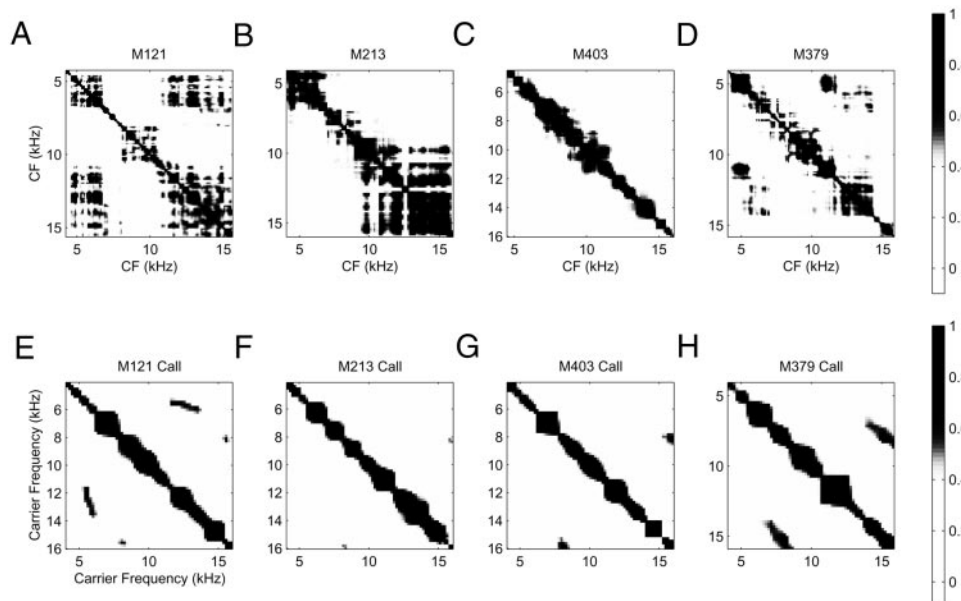


FIG. 7. Synchronization of responses in AI. A–D: correlation-coefficients between the responses to MOC of AI neurons with different characteristic frequencies (CFs). Data for each of the 4 animals in our study is shown (A, *M121*; B, *M213*; C, *M403*; D, *M379*). The off-diagonal terms indicate the presence of short- and long-range synchronization in the responses. E–H: correlation coefficients between the envelope of MOC for different frequencies of the envelope of MOC for each of the 4 animals. The synchronization of neurons in AI does not merely reflect the cross-frequency correlations in AI.

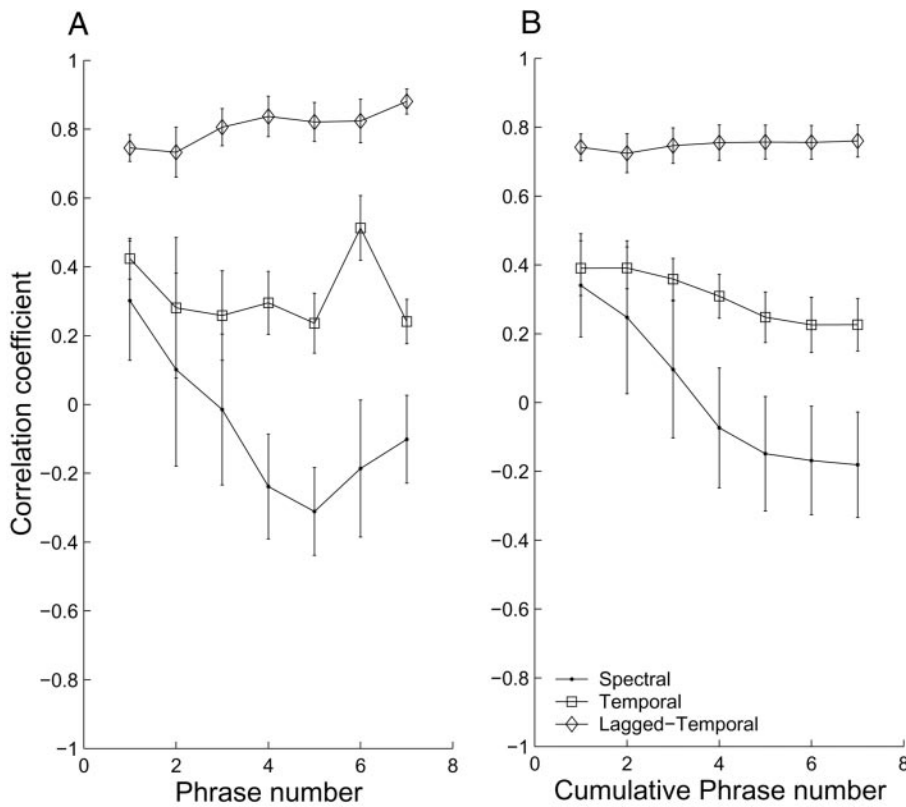


FIG. 8. Mean temporal and spectral representations of MOC. *A*: per-phrase analysis. Correlation coefficient between the temporal envelope of each phrase of a call and the population PSTH response for each phrase (□, nonlagged; ◇, lagged) and between the spectral envelope of each phrase and the mean-spectral representation (●). *B*: cumulated phrase analysis. Correlation coefficients computed same for calls of increasing duration to cumulatively include the effect of each additional phrase in a call. Conventions same as in *A*.

spectral representation and the spectral envelope of the call, respectively. Given that the correlation-coefficient represents only zero-time-lagged correlations and the fact that a time lag exists for the temporal response to stimulus due to

propagation of information to auditory cortex, the maximum of nonzero time lagged correlations between the population PSTHs and the temporal envelope were also computed (see Fig. 8). Typical lag times at which the correlations were

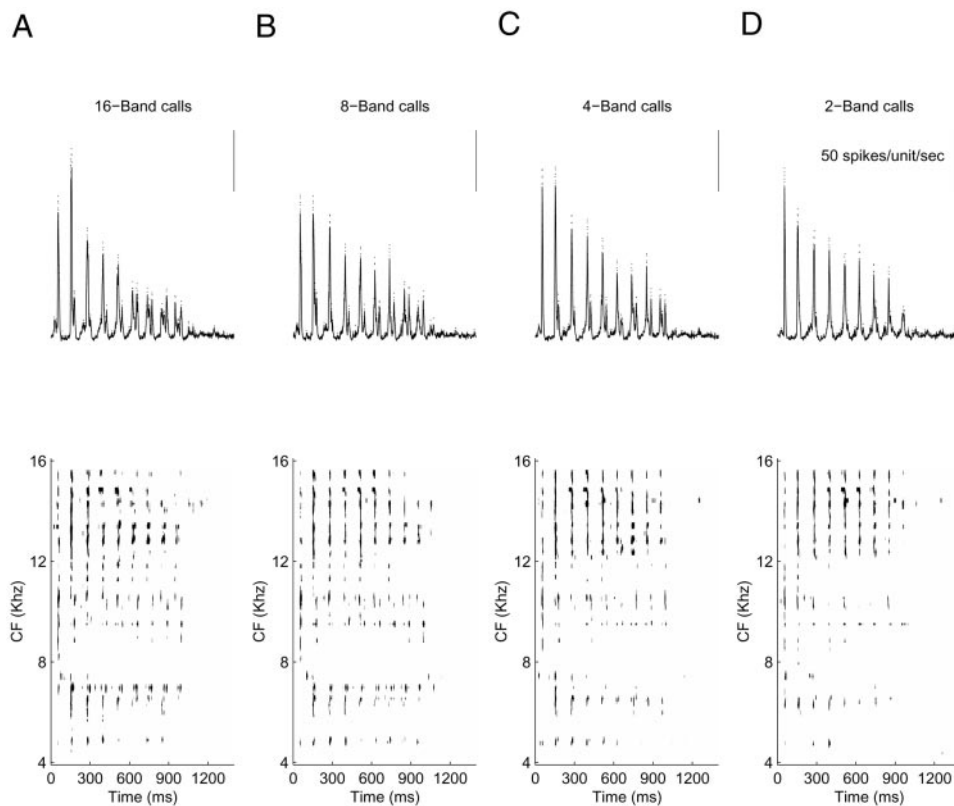


FIG. 9. Population responses to spectral envelope degradations from *M121*. *A*, top: the population PSTH response across units to a spectrally degraded twitter call with carrier frequencies randomized across 16 bands. *Bottom*: the corresponding neurogram response to the same degraded twitter call. *B-D*: conventions the same as in *A*. Responses corresponding to degraded calls with carrier frequencies randomized across 8, 4, and 2 bands.

maximum were ~ 20 ms, consistent with conduction and propagation delays in the auditory system. The correlation analyses were conducted on a per-phrase basis as well as on a cumulative-phrase basis. Individual phrase analysis was conducted for a fixed duration of the call and the evoked responses, corresponding to each phrase of the call (Fig. 8A). For cumulative-phrase analysis, the duration of the calls and the responses were incremented to include additional phrases (Fig. 8B). In all animals, the lagged-temporal correlations between the temporal envelope and the population PSTH are significantly higher than the nonlagged correlation both for the individual per phrase and the cumulative phrase analysis conditions. The correlation between the spectral envelope and mean spectral representation (spectral profile of the neurograms) showed low correlation coefficients. For later phrases or when including later phrases into the cumulative analysis, the correlation coefficients were either near zero or even negative (Fig. 8). Therefore over the time scale of the call, the temporal envelope of the call is better represented by the population response of AI neurons than the spectral envelope.

Responses to spectral envelope degradations

The relevance of details in the spectral envelope for the coding of complex sounds was further tested by gradually degrading the spectral information present in the calls. The synchronous firing of responses in AI to marmoset calls did not decrease when spectral envelopes were degraded by reducing the number of bands from 16 to 2 and by randomization of the carrier phase. Figure 9 illustrates examples of population PSTHs and corresponding neurograms from the entire sample of neurons recorded in one animal (*M121*). Qualitatively, despite variations in the responses of individual neurons across the population, the spectral pattern of the neurograms and in particular, across the population the synchronized temporal discharge pattern did not differ considerably as the number of bands was reduced (Fig. 9, A–D).

Three measures were used to quantify the responses—MDR across the entire duration of the call, response strength (R_s), and mean driven-rate/phrase (MDP; Fig. 10). These measures are shown both for the MOC and when the number of bands of the spectral envelope degradations was reduced from 16 to 2. Although two of the three measures, the MDR and R_s , were significantly different across these stimulus conditions ($P < 0.01$ and $P < 0.05$, respectively), mean differences across these conditions were small. For example the mean difference in MDR was only 2–3 spikes/s, and the mean difference in R_s was ~ 2 –3 spikes²/s.

Responses to temporal envelope degradations

In contrast to responses to calls with spectral envelope degradations, the neural representations was altered dramatically by degradations in the temporal envelope when the carrier and phase structure is unchanged. An example of the effects of temporal envelope degradations is shown in Fig. 11 (*M379*). Qualitatively, the synchronized responses was significantly diminished for calls where the temporal envelope is high-pass filtered at 60 Hz (HP60, Fig. 11D), low-pass filtered at 10 Hz (LP4, Fig. 11C), and low-pass filtered at 4 Hz (LP10, Fig.

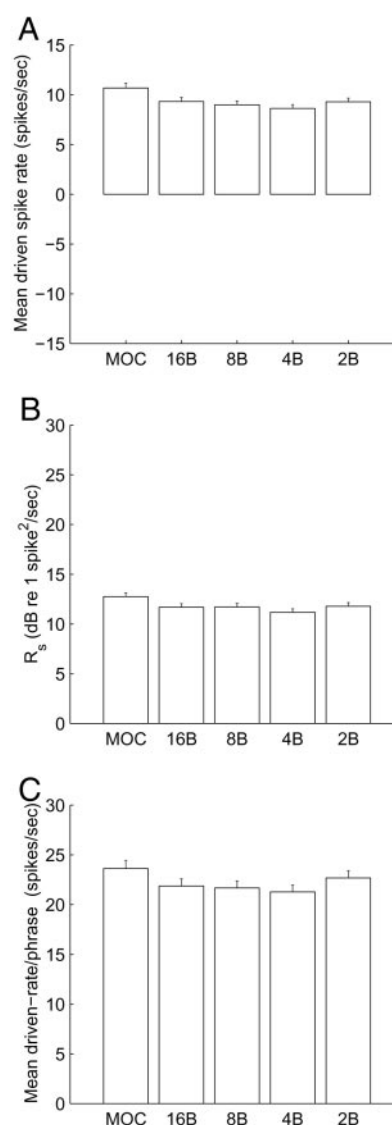


FIG. 10. Quantification of responses to degraded calls. A–C: measures of response strength to MOC and the spectral envelope degradation calls 16, 8, 4, and 2 bands. A: driven spike rate during a call. B: response synchronization. C: driven spike rate per phrase.

11B). However, there is no reduction in the synchronous response profile to each phrase of the call when the temporal envelope is band-pass filtered between 2 and 30 Hz (BP2-30, Fig. 11A), compared with MOC.

To quantify the responses across these different conditions, the three measures MDR (Fig. 12A), R_s (Fig. 12B) and MDP (Fig. 12C) were computed. MOC responses are greater than the responses to LP04, LP10, and HP60 calls ($P < 0.00001$ for all 3 measures). However, the responses to BP2-30 calls were found to be slightly greater than the responses to MOC ($P < 0.0001$ for all 3 measures).

Response to MOC in the presence of background noise

Responses to MOC under continuous noise background conditions ($\text{SNR} \geq 10$ dB) does not degrade the spectrotemporal relationships across units. Population data from the four animals studied are illustrated in Figs. 13 and 14 for calls pre-

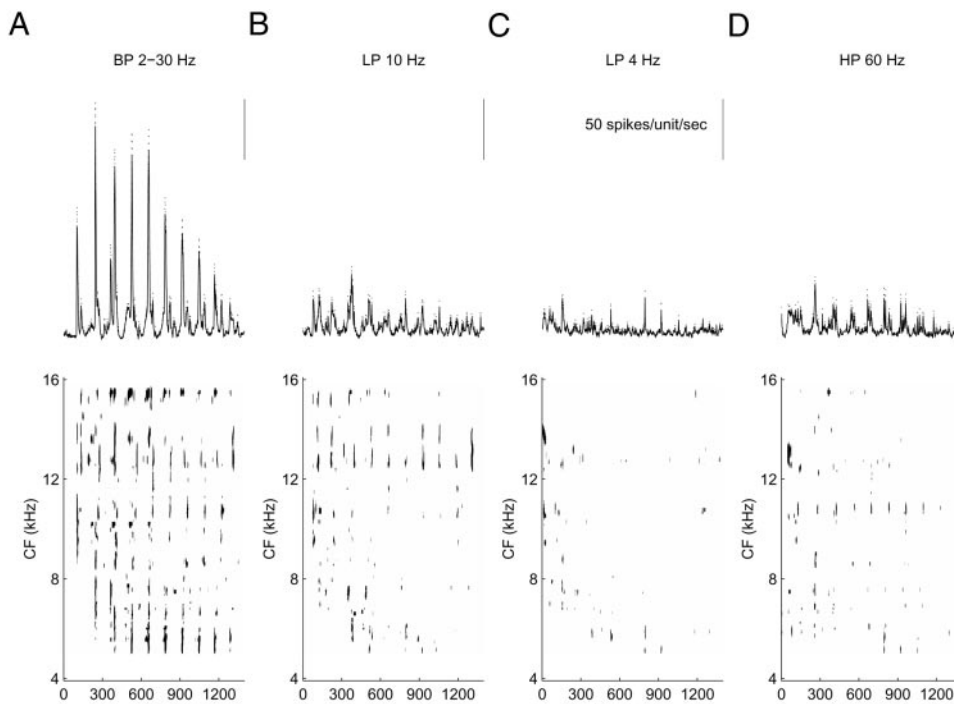


FIG. 11. Population responses to temporal envelope degradation from M379. *A, top*: the population PSTH response across units to a degraded call where the temporal envelope within narrowband channels were filtered between 2 and 30 Hz. *Bottom*: the corresponding neurogram to the same degraded twitter call. *B*: responses to a degraded twitter call where the temporal envelope was high pass filtered at 60 Hz (HP60). *C*: responses to a twitter call where the temporal envelope was low-pass filtered at 10 Hz (LP10). *D*: responses to a twitter call where the temporal envelope was low-pass filtered at 4 Hz (LP4).

sented in the presence of background noise at 20 (Figs. 13, *A* and *C*, and 14, *A* and *C*), 10 (Figs. 13, *B* and *D*, and 14*B*), and 0 dB SNR (Fig. 14*D*). Qualitatively, the population PSTHs do not show a significant reduction under noise levels of 20 or 10 dB SNR (see Figs. 13, *B* and *D*, and 14*B*). However, the responses at 0 dB SNR clearly indicate a marked and significant reduction ($P < 0.001$) in the synchronous response to each phrase of the call (Fig. 14*D*). However, even at the lowest tested SNRs, the across-frequency short- and long-range correlation structure is only moderately altered (Fig. 15).

A closer examination of the neurograms revealed that individual differences exist in the responses of the AI neurons to MOC under background noise. Surprisingly, some neurons respond stronger to calls under noisy conditions when compared with silence as illustrated by units with CFs >10 kHz (compare Fig. 5 with Fig. 13, *C* and *D*). About 10–15% of units in every animal have greater synchronous firing rates evoked by calls presented in noise than by calls in silence.

Comparisons of responses to spectral and temporal envelope degradations and additive noise

Comparisons among responses to degraded forms of vocalizations and additive noise to MOC (SNR 20 dB) are quantified using normalized variant of the three measures, MDR, MDP, and R_s . These summary measures of parametrically varied experimental conditions are normalized to the MOC responses in silence. Figure 16 illustrates the differences. Three observations are revealed from this normalized data analysis. First, there is a small but significant reduction (Fig. 16, *A–C*, $P < 0.00001$), for all three measures in the response within each recording site to degradations of the spectral envelope when compared with the response to MOC. Second, the reduction in responses to temporal-envelope degraded calls (LP04, LP10, and HP60) are similar to corresponding reductions in R_s , and MDP, to MOC presented at SNR = 20 dB ($P > 0.05$ for both

measures, Fig. 16, *B* and *C*). Third, the response to the BP2-30 Hz call is consistently and significantly greater than the response to MOC ($P < 0.001$ for all measures, Fig. 16, *A–C*).

DISCUSSION

Six major findings of this paper clarify and extend our understanding of the distributed cortical encoding of complex species-specific vocalizations within AI of barbiturate-anesthetized marmosets. These findings can be summarized as follows. First, responses to twitter vocalizations are correlated with responses to amplitude-modulated tonal stimuli of comparable modulation rates. Second, both long- and short-range synchronization across the frequency axis are observed in the spectrotemporal responses to complex vocalizations. Third, the temporal envelope of complex vocalizations is significantly correlated with the temporal response patterns of AI neurons. In contrast, the spectral envelope of a call is poorly represented in the mean-spectral representation of AI responses. Fourth, response magnitudes in AI neurons are relatively insensitive to the fidelity of the spectral-envelope characteristics of the call. Fifth, responses of AI neurons and their distributed synchronizations are significantly reduced when the temporal envelope is degraded. These data indicate that AI neurons in the barbiturate-anesthetized marmoset monkeys mainly respond to, and represent by their synchronized discharges, the temporal cues in complex stimuli. Finally, responses of AI neurons to MOC and con-specific calls are robust in the presence of additive background noise (SNR ≥ 0 dB).

Neurophysiological data in this study parallel human behavioral data on speech intelligibility with degraded forms of speech. Several studies of human speech perception have revealed that slow fluctuations in speech spectrum for intelligibility manifested as low-frequency temporal modulations (<16 Hz) in narrowband are sufficient (Drullman 1995b; Drullman et al. 1994a,b; Shannon et al. 1995). Use of relatively slow

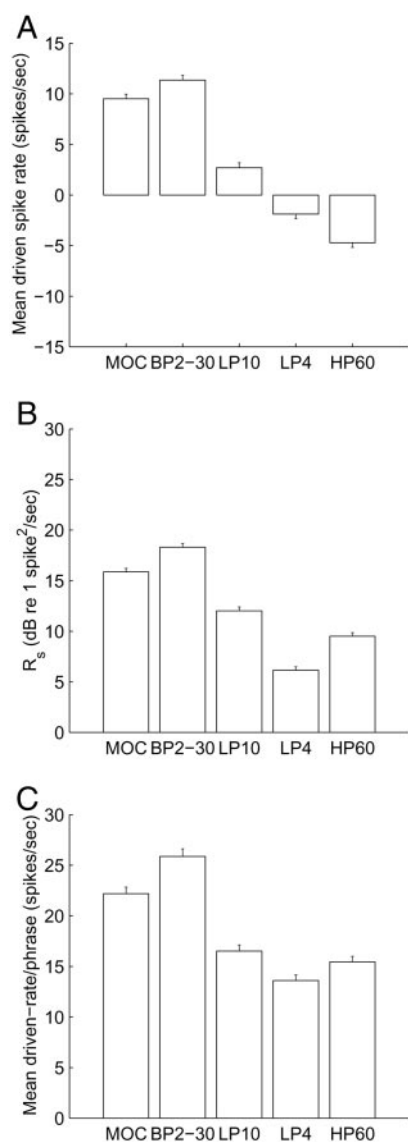


FIG. 12. Quantification of responses to degraded calls. A–C: measures of response strength to MOC and temporal envelope degradations low-pass at 4 Hz (LP4), low-pass at 10 Hz (LP10), band-pass at 2–30 Hz (BP2-30), and high-pass at 60 Hz (HP60). A: driven spike rate. B: response synchronization. C: driven spike rate per phrase.

modulations reduces both the sensitivity of listeners and artificial speech recognition devices to background noise, reverberation, and transient interference (Kingsbury et al. 1998). Such human behavioral data are similar to the sensitivity of marmoset AI neurons to slow modulation of the temporal envelope of the twitter call.

Shannon et al. (1995) performed degradations in speech similar to our spectral envelope degradations. Their studies have shown that speech intelligibility is preserved as long as temporal envelope cues are present despite gross spectral degradation to as few as four broad frequency bands. A significant reduction in speech intelligibility was observed when only two bands were used. This behavioral correlate is not directly evident in our recordings from AI neurons potentially due to differences in the acoustics and perception between monkey calls and speech. However, a small but consistent tuning in the response measures is observed across the spectral degradation

continuum, with a nadir for a four-band stimulus form as indicated in the normalized analysis of Fig. 15. The temporal structure of the twitter call is overwhelmingly dominant for perception. Whether this tuning is sufficient to account for behavioral data of speech intelligibility in humans remains to be further investigated. It should be kept in mind, though, that the spectral degradation has other effects than response magnitude reduction with potential perceptual consequences. Most notably (see Fig. 9), fine details of the onset timing of phrases across CF are eliminated by limiting the spectral information to a small number of bands. This loss of fine timing distinctions between information in different frequency channels may have a greater perceptual impact than differential changes in firing rate.

From Fig. 16 it is clear that, although the response magnitudes show moderate sensitivity to spectral envelope features, the dominant feature of sensitivity within the twitter call is the temporal envelope. A reduction in the number of channels reduces temporal differentiation among different subpopulations in accordance with human behavioral data (Shannon et al. 1995) that suggest nonsynchronous temporal details in more than two channels are necessary for intelligibility. The lack of sensitivity to spectral envelope features reported here might lead to a misconception that these results are in conflict with ripple frequency tuning observed in auditory cortex of cats and ferrets (Calhoun and Schreiner 1998; Kowalski et al. 1996a,b; Versnel and Shamma 1998). Although the dynamic range of ripple tuning remains to be established in monkeys, it should be noted that spectral envelope degradations can be conceived as re-sampling the spectral envelope. Such a spectral envelope resampling will reduce higher ripple densities in the spectral envelope by aliasing to low-ripple densities while not changing the ripple phase or peak frequencies. Therefore our results do not reflect aspects of ripple-frequency tuning of AI neurons.

The observed temporal envelope sensitivity exhibited by AI neurons is analogous to Drullman et al.'s (1994) psychophysical results on speech intelligibility to degraded temporal envelopes. For example, the reduction in cortical responses for LP4 and HP60 calls are consistent with the behavioral data indicating decrement in speech intelligibility when similar degradations are performed with speech. Furthermore, a significant increase in the response to BP2-30 calls in comparison to MOC suggest that these envelope frequencies are indeed most relevant for the robustness and fidelity of the representation of phrase synchronous responses. Similar band-pass filtering of the temporal envelope of speech has been found to contribute to significant learning and enhanced of speech comprehension in language-learning-impaired children (Tallal et al. 1996).

However, there are interesting differences between physiological and psychophysical observations as well. Drullman and colleagues report no significant reduction in speech intelligibility to LP10 and BP2-30 modifications, whereas we observe a significant reduction in AI responses for the LP10 condition. It was indeed surprising that the LP10 calls contribute to significant reductions in AI responses because the call has energy well within the modulation transfer function characteristics of most AI neurons. These observations, in combination with our findings with SAM stimuli, further underscore the sensitivity of the AI responses to higher-frequency spectral components and non-

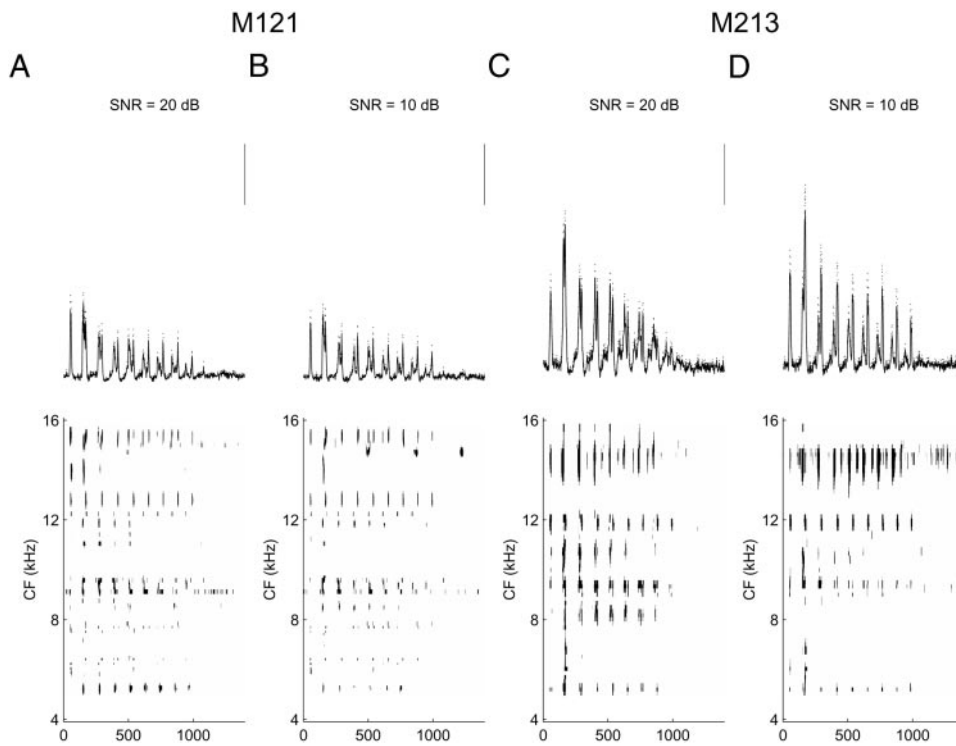


FIG. 13. Population responses to noise. Responses from *M121* to MOC at SNR of 20 (A) and responses to MOC at SNR = 10 dB (B). C: responses from *M213* to MOC at SNR = 20 dB. D: responses from *M213* to MOC at SNR = 10 dB.

linear onset characteristics of species-specific vocalizations.

Responses to SAM stimuli account for only 22% of the variance of the responses to calls, and responses to click stimuli only accounts for 13% of the variance in the response to calls. Further, the fact that responses to MOC were significantly greater than to SAM stimuli and to click-train sequences at comparable periodicities indicate that repetition rate

coding and onset spectrum effects are separable aspects of temporal envelope processing.

Twitter vocalizations appear to have two time scales in their structure. A fast time scale (10–30 ms) marks individual phrases or syllables; a slower time scale marks inter-phrase intervals. The data from this study indicate that AI neurons are sensitive to the slower time scales in twitter calls. However, the representation of faster time scales appears to be in the form of

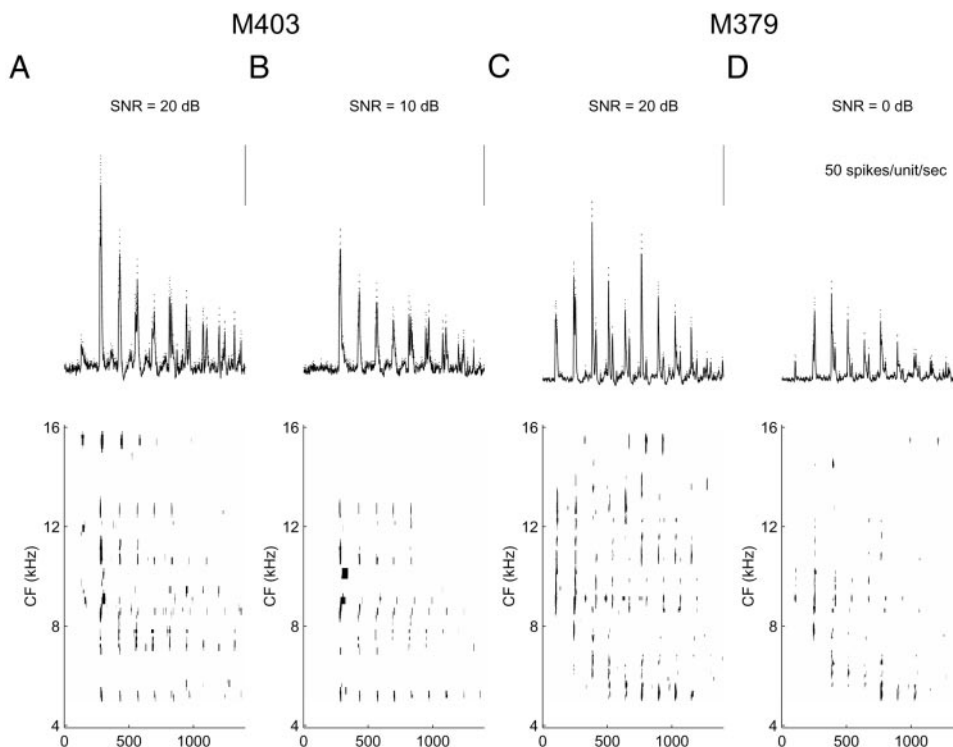


FIG. 14. Population responses to noise. Responses from *M403* to MOC at SNR of 20 (A) and responses to MOC at SNR = 10 dB (B). C: responses from *M379* to MOC at SNR = 20 dB. D: responses from *M379* to MOC at SNR = 0 dB.

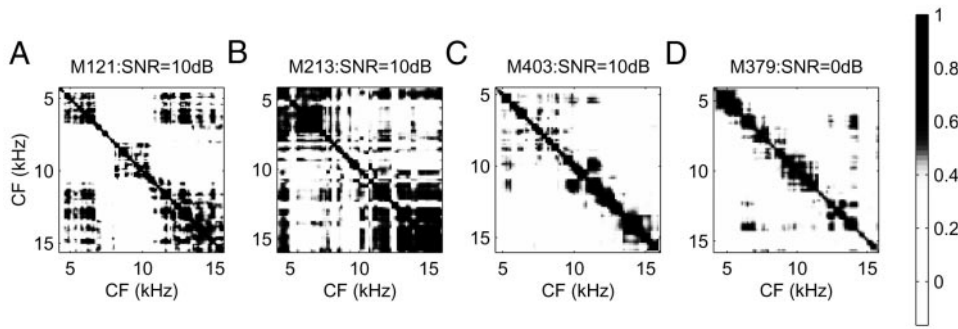


FIG. 15. Synchronization in AI in the presence of background noise. Correlation coefficients between the responses to MOC of AI neurons with different CFs are plotted for the least SNR tested for each animal. Data for each of the 4 animals in our study are shown (A: *M121*, 10 dB; B: *M213*, 10 dB; C: *M403*, 10 dB; D: *M379*, 0 dB). The continued presence of off-diagonal terms indicate that the short- and long-range synchronization in the responses are only moderately destroyed at 0 dB SNR.

short and long-range synchronizations between neuronal populations within AI.

The robustness of the AI representations of vocalizations to additive background noise ≤ 0 dB SNR is consistent with psychophysical observations of speech comprehension under these low signal-to-noise conditions. Furthermore, the additive

noise data serve as a calibration to determine the sensitivity of responses to spectral and temporal envelope degradations. Response synchronization was generally attenuated due to additive noise, although response patterns were preserved. Surprisingly, subpopulations of neurons did not change their responses to additive noise and some neurons (10–15% in each

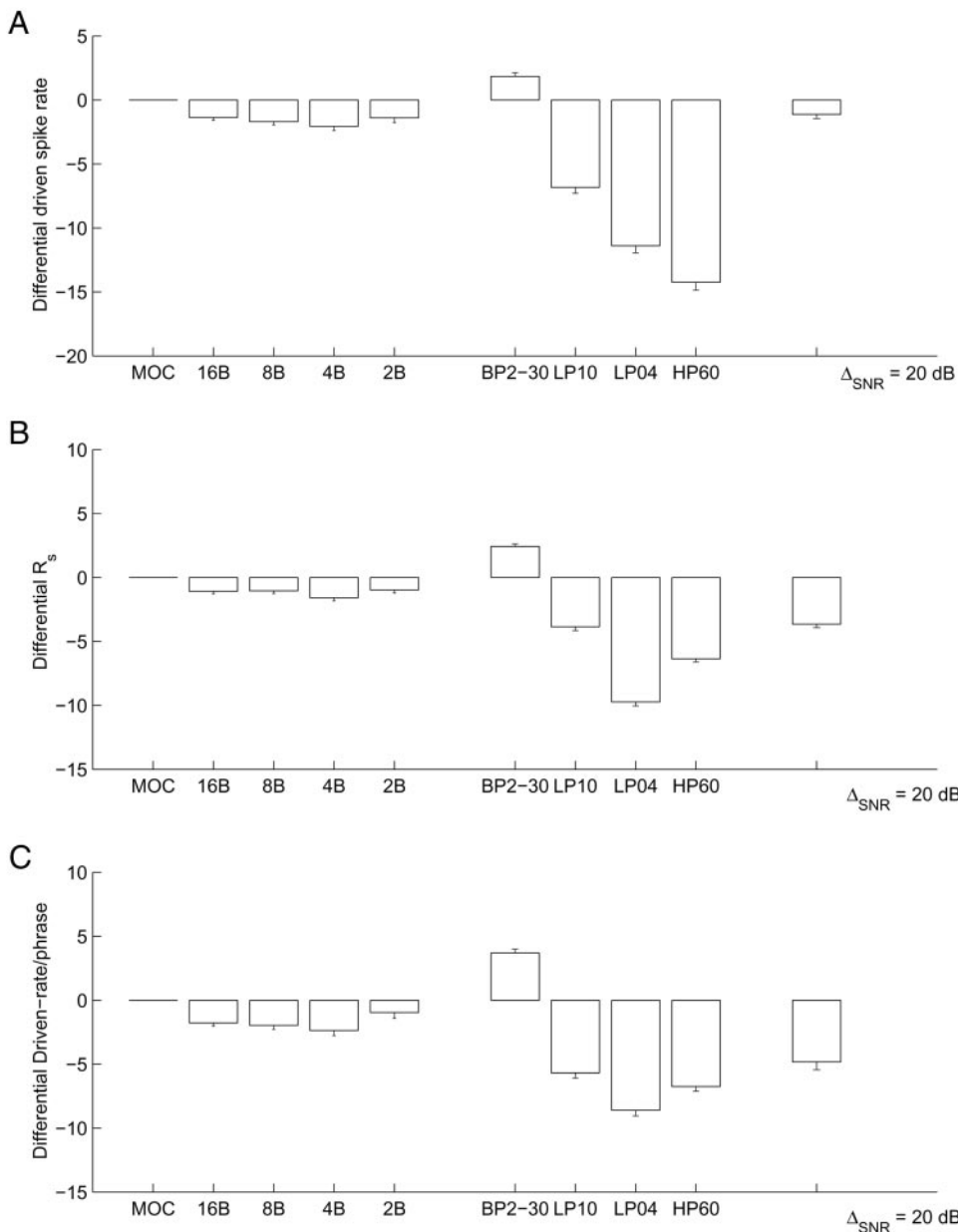


FIG. 16. Normalized comparisons of responses of MOC to calls in silence and in the presence of background noise with spectral and temporal envelope degradations. Responses are normalized for each unit to the response to MOC. A: driven spike rate during duration of call. B: response synchronization. C: driven spike rate per phrase.

animal) increased their response synchronization to the call during increased background noise. The mechanisms underlying this surprising robustness to additive noise remain poorly understood and will be the focus of future investigations.

The data reported here were obtained from animals anesthetized with pentobarbital, and one must be cautious in generalizing these results to cortical activity in awake marmosets (Gaese and Ostwald 2001; Zurita et al. 1994). The evidence so far indicates that the spatial organization of auditory cortex is similar in anesthetized and awake animals (Brugge and Merzenich 1973; Lu et al. 2001; Pelleg-Toiba and Wollberg 1989; Recanzone 2000; Recanzone et al. 1999, 2000). In the temporal domain, the frequencies at which responses of AI neurons can follow AM stimuli were found to be similar in anesthetized and awake animals, although awake animals had a higher cutoff for synchronous responses (Creutzfeldt et al. 1980; Lu and Wang 2000; Schreiner and Urbas 1988). Recordings of SAM responses from awake squirrel monkeys show similar range of following although upper cutoff frequencies of the modulation transfer functions (MTFs) may be shifted to higher values (Bieser 1998; Bieser and Müller-Preuss 1996). Our findings that the best-modulation frequency of neurons was not correlated with the responses to vocalizations further suggest that similar findings could be obtained in awake recordings. Also, in awake animals, the spontaneous firing rate of AI neurons are generally higher and sustained responses are more apparent (deCharms et al. 1998; Lu and Wang 2000; Recanzone 2000; Recanzone et al. 2000), but the synchronous and coherent distributed responses shown here should still be evident in both awake and barbiturate-anesthetized animals (De Ribaupierre et al. 1972; Goldstein et al. 1968; Schwarz and Tomlinson 1990). These findings further support the conclusion that in AI, single neurons do not exhibit specificity to the monkey's own call; rather, they indicate sensitivity to specific acoustic features of the temporal envelope within the call. It is clear that, across the AI neuronal population, stimulus features such as FM sweeps are abstracted in representation by synchronous, discharging cell assemblies. The degree of synchronization depends on the idiosyncratic spectrotemporal acoustic features of the vocalization. These abstractions constitute a form of the "neuronal" translation of the acoustic features of a call. Presumably these abstracted representations may lead to downstream hierarchical organization of "call-specific" detectors or "cell assemblies," analogous to higher-order neurons in birds and bats and monkeys (Doupe and Konishi 1991; Fitzpatrick et al. 1993; Margoliash and Fortune 1992; Suga 1972, 1989; Suga et al. 1983, 1990; Theunissen and Doupe 1998; Tian et al. 2001).

In summary, this study further elucidates the specific aspects of the encoding in AI and forms of transmission of such information downstream from AI. Specifically, we provide direct evidence that AI response profiles to complex species-specific vocalizations are robust to additive noise, sensitive to temporal envelope features, and relatively insensitive to the details of spectral envelope features. Interestingly in humans, speech encoding with primarily temporal envelope features are also robust at moderate levels of background noise (Drullman 1995a; Kingsbury et al. 1998). These data suggest that cortical responses to vocal stimuli may contribute to the recognition of complex stimuli, including speech, under naturalistic noisy conditions.

We gratefully acknowledge support from the National Institutes of Health (NS-10414, NS-34835, NSF-SBR 9720398, and National Research Service Award fellowship F32-DC00285), Veterans Affairs Medical Research (S. W. Cheung), Deafness Research Foundation (S. S. Nagarajan), the Coleman Fund, and Hearing Research Inc.

REFERENCES

- BIESER A. Processing of twitter-call fundamental frequencies in insula and auditory cortex of squirrel monkeys. *Exp Brain Res* 122: 139–148, 1998.
- BIESER A AND MÜLLER-PREUSS P. Auditory responsive cortex in the squirrel monkey—neural responses to amplitude-modulated sounds. *Exp Brain Res* 108: 273–284, 1996.
- BRUGGE JF AND MERZENICH MM. Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J Neurophysiol* 36: 1138–1158, 1973.
- CALHOUN BM AND SCHREINER CE. Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. *Eur J Neurosci* 10: 926–940, 1998.
- CHEUNG SW, BEDENBAUGH PH, NAGARAJAN SS, AND SCHREINER CE. Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. *J Neurophysiol* 85: 1732–1749, 2001.
- CREUTZFELDT O, HELLWEG FC, AND SCHREINER C. Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res* 39: 87–104, 1980.
- DE RIBAUPIERRE F, GOLDSTEIN MH JR, AND YENI-KOMSHIAN G. Cortical coding of repetitive acoustic pulses. *Brain Res* 48: 205–225, 1972.
- DECHARMS RC, BLAKE DT, AND MERZENICH MM. Optimizing sound features for cortical neurons. *Science* 280: 1439–1443, 1998.
- DOUPE AJ AND KONISHI M. Song-selective auditory circuits in the vocal control system of the zebra finch. *Proc Natl Acad Sci USA* 88: 11339–11343, 1991.
- DRULLMAN R. Speech intelligibility in noise: relative contribution of speech elements above and below the noise level. *J Acoust Soc Am* 98: 1796–1798, 1995a.
- DRULLMAN R. Temporal envelope and fine structure cues for speech intelligibility. *J Acoust Soc Am* 97: 585–592, 1995b.
- DRULLMAN R, FESTEN JM, AND PLOMP R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 95: 1053–1064, 1994a.
- DRULLMAN R, FESTEN JM, AND PLOMP R. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am* 95: 2670–2680, 1994b.
- EGGERMONT JJ. Rate and synchronization measures of periodicity coding in cat primary auditory cortex. *Hear Res* 56: 153–167, 1991.
- ESSER KH, CONDON CJ, SUGA N, AND KANWAL JS. Syntax processing by auditory cortical neurons in the FM-FM area of the mustached bat *Pteronotus parnellii*. *Proc Natl Acad Sci USA* 94: 14019–14024, 1997.
- FITZPATRICK DC, KANWAL JS, BUTMAN JA, AND SUGA N. Combination-sensitive neurons in the primary auditory cortex of the mustached bat. *J Neurosci* 13: 931–940, 1993.
- GAESE BH AND OSTWALD J. Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex. *J Neurophysiol* 86: 1062–1066, 2001.
- GEHR DD, KOMIYA H, AND EGGERMONT JJ. Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. *Hear Res* 150: 27–42, 2000.
- GLASS I AND WOLLBERG Z. Lability in the responses of cells in the auditory cortex of squirrel monkeys to species-specific vocalizations. *Exp Brain Res* 34: 489–498, 1979.
- GLASS I AND WOLLBERG Z. Auditory cortex responses to sequences of normal and reversed squirrel monkey vocalizations. *Brain Behav Evol* 22: 13–21, 1983a.
- GLASS I AND WOLLBERG Z. Responses of cells in the auditory cortex of awake squirrel monkeys to normal and reversed species-specific vocalizations. *Hear Res* 9: 27–33, 1983b.
- GOLDSTEIN MH JR, HALL JLD, AND BUTTERFIELD BO. Single-unit activity in the primary auditory cortex of unanesthetized cats. *J Acoust Soc Am* 43: 444–455, 1968.
- GREENBERG S AND ARAI T. The relationship between speech intelligibility and the complex modulation spectrum. *Proc Eurospeech* 473–476, 2001.
- HEIL P, RAJAN R, AND IRVINE DR. Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I. Effects of variation of stimulus parameters. *Hear Res* 63: 108–134, 1992.
- IMIG TJ, RUGGERO MA, KITZES LM, JAVEL E, AND BRUGGE JF. Organization of auditory cortex in the owl monkey (*Aotus trivirgatus*). *J Comp Neurol* 171: 111–128, 1977.

- KANWAL JS, MATSUMURA S, OHLEMILLER K, AND SUGA N. Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *J Acoust Soc Am* 96: 1229–1254, 1994.
- KINGSBURY BED, MORGAN N, AND GREENBERG S. Robust speech recognition using the modulation spectrogram. *Speech Commun* 25: 117–132, 1998.
- KOWALSKI N, DEPIREUX DA, AND SHAMMA SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76: 3503–3523, 1996a.
- KOWALSKI N, DEPIREUX DA, AND SHAMMA SA. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *J Neurophysiol* 76: 3524–3534, 1996b.
- LEWICKI MS AND ARTHUR BJ. Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16: 6987–6998, 1996.
- LU T, LIANG L, AND WANG X. Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol* 85: 2364–2380, 2001.
- LU T AND WANG X. Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat. *J Neurophysiol* 84: 236–246, 2000.
- MANLEY JA AND MULLER-PREUSS P. Response variability of auditory cortex cells in the squirrel monkey to constant acoustic stimuli. *Exp Brain Res* 32: 171–180, 1978.
- MARGOLIASH D AND FORTUNE ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 12: 4309–4326, 1992.
- MENDELSON JR, SCHREINER CE, AND SUTTER ML. Functional topography of cat primary auditory cortex: response latencies. *J Comp Physiol [A]* 181: 615–633, 1997.
- NEWMAN JD AND WOLLBERG Z. Responses of single neurons in the auditory cortex of squirrel monkeys to variants of a single call type. *Exp Neurol* 40: 821–824, 1973a.
- NEWMAN JD AND WOLLBERG Z. Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res* 54: 287–304, 1973b.
- OHLEMILLER KK, KANWAL JS, AND SUGA N. Facilitative responses to species-specific calls in cortical FM-FM neurons of the mustached bat. *Neuroreport* 7: 1749–1755, 1996.
- PELLEG-TOIBA R AND WOLLBERG Z. Tuning properties of auditory cortex cells in the awake squirrel monkey. *Exp Brain Res* 74: 353–364, 1989.
- PELLEG-TOIBA R AND WOLLBERG Z. Discrimination of communication calls in the squirrel monkey: “call detectors” or “cell ensembles”? *J Basic Clin Physiol Pharmacol* 2: 257–272, 1991.
- RAUSCHCKER JP. Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl* 532: 34–38, 1997.
- RAUSCHCKER JP. Cortical processing of complex sounds. *Curr Opin Neurobiol* 8: 516–521, 1998a.
- RAUSCHCKER JP. Parallel processing in the auditory cortex of primates. *Audiol Neurootol* 3: 86–103, 1998b.
- RAUSCHCKER JP, TIAN B, AND HAUSER M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111–114, 1995.
- RECANZONE GH. Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. *Hear Res* 150: 104–111, 2000.
- RECANZONE GH, GUARD DC, AND PHAN ML. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J Neurophysiol* 83: 2315–2331, 2000.
- RECANZONE GH, SCHREINER CE, SUTTER ML, BEITEL RE, AND MERZENICH MM. Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *J Comp Neurol* 415: 460–481, 1999.
- ROTMAN Y, BAR-YOSEF O, AND NELKEN I. Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. *Hear Res* 152: 110–127, 2001.
- SCHREINER CE. Functional topographies in the primary auditory cortex of the cat. *Acta Otolaryngol Suppl* 491: 7–15, 1991.
- SCHREINER CE AND MENDELSON JR. Functional topography of cat primary auditory cortex: distribution of integrated excitation. *J Neurophysiol* 64: 1442–1459, 1990.
- SCHREINER CE AND URBAS JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 32: 49–63, 1988.
- SCHWARZ DWF AND TOMLINSON RWW. Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J Neurophysiol* 64: 282–298, 1990.
- SHANNON RV, ZENG F-G, KAMATH V, AND WYGONSKI J. Speech recognition with primarily temporal cues. *Science* 270: 303–304, 1995.
- SUGA N. Analysis of information-bearing elements in complex sounds by auditory neurons of bats. *Audiology* 11: 58–72, 1972.
- SUGA N. Principles of auditory information-processing derived from neuroethology. *J Exp Biol* 146: 277–286, 1989.
- SUGA N, OLSEN JF, AND BUTMAN JA. Specialized subsystems for processing biologically important complex sounds: cross-correlation analysis for ranging in the bat's brain. *Cold Spring Harb Symp Quant Biol* 55: 585–597, 1990.
- SUGA N, O'NEILL WE, KUJIRAI K, AND MANABE T. Specificity of combination-sensitive neurons for processing of complex biosonar signals in auditory cortex of the mustached bat. *J Neurophysiol* 49: 1573–1626, 1983.
- TALLAL P, MILLER SL, BEDI G, BYMA G, WANG X, NAGARAJAN SS, SCHREINER C, JENKINS WM, AND MERZENICH MM. Language comprehension in language-learning impaired children improved with acoustically modified speech [see comments]. *Science* 271: 81–84, 1996.
- THEUNISSEN FE AND DOUPE AJ. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J Neurosci* 18: 3786–3802, 1998.
- THOMSON DJ. Spectrum estimation and harmonic analysis. *Proc IEEE* 70: 1055–1096, 1982.
- TIAN B, RESER D, DURHAM A, KUSTOV A, AND RAUSCHCKER JP. Functional specialization in rhesus monkey auditory cortex. *Science* 292: 290–293, 2001.
- VERSNEL H AND SHAMMA SA. Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J Acoust Soc Am* 103: 2502–2514, 1998.
- WANG X, MERZENICH MM, BEITEL R, AND SCHREINER CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685–2706, 1995.
- WINTER P AND FUNKENSTEIN HH. The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey (*Saimiri sciureus*). *Exp Brain Res* 18: 489–504, 1973.
- WOLLBERG Z AND NEWMAN JD. Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations. *Science* 175: 212–214, 1972.
- ZURITA P, VILLA AE, DE RIBAUPIERRE Y, DE RIBAUPIERRE F, AND ROUILLER EM. Changes of single-unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. *Neurosci Res* 19: 303–316, 1994.