JOURNAL OF NEUROPHYSIOLOGY
Vol. 74, No. 6, December 1995. *Printed in U.S.A.*

Representation of a Species-Specific Vocalization in the Primary Auditory Cortex of the Common Marmoset: Temporal and Spectral Characteristics

XIAOQIN WANG, MICHAEL M. MERZENICH, RALPH BEITEL, AND CHRISTOPH E. SCHREINER
Coleman Laboratory and W. M. Keck Center for Integrative Neuroscience, University of California, San Francisco, California 94143-0732

SUMMARY AND CONCLUSIONS

1. The temporal and spectral characteristics of neural representations of a behaviorally important species-specific vocalization were studied in neuronal populations of the primary auditory cortex (A1) of barbiturate-anesthetized adult common marmosets (*Callithrix jacchus*), using both natural and synthetic vocalizations. The natural vocalizations used in electrophysiological experiments were recorded from the animals under study or from their conspecifics. These calls were frequently produced in vocal exchanges between members of our marmoset colony and are part of the well-defined and highly stereotyped vocal repertoire of this species.

2. The spectrotemporal discharge pattern of spatially distributed neuron populations in cortical field A1 was found to be correlated with the spectrotemporal acoustic pattern of a complex natural vocalization. However, the A1 discharge pattern was not a faithful replication of the acoustic parameters of a vocalization stimulus, but had been transformed into a more abstract representation than that in the auditory periphery.

3. Subpopulations of A1 neurons were found to respond selectively to natural vocalizations as compared with synthetic variations that had the same spectral but different temporal characteristics. A subpopulation responding selectively to a given monkey's call shared some but not all of its neuronal memberships with other individual-call-specific neuronal subpopulations.

4. In the time domain, responses of individual A1 units were phase-locked to the envelope of a portion of a complex vocalization, which was centered around a unit's characteristic frequency (CF). As a whole, discharges of A1 neuronal populations were phase-locked to discrete stimulus events but not to their rapidly changing spectral contents. The consequence was a reduction in temporal complexity and an increase in cross-population response synchronization.

5. In the frequency domain, major features of the stimulus spectrum were reflected in rate-CF profiles. The spectral features of a natural call were equally or more strongly represented by a subpopulation of A1 neurons that responded selectively to that call as compared with the entire responding A1 population.

6. Neuronal responses to a complex call were distributed very widely across cortical field A1. At the same time, the responses evoked by a vocalization scattered in discrete cortical patches were strongly synchronized to stimulus events and to each other. As a result, at any given time during the course of a vocalization, a coherent representation of the integrated spectrotemporal characteristics of a particular vocalization was present in a specific neuronal population.

7. These results suggest that the representation of behaviorally important and spectrotemporally complex species-specific vocalizations in A1 is *1*) temporally integrated and *2*) spectrally distributed in nature, and that the representation is carried by spatially dispersed and synchronized cortical cell assemblies that correspond to each individual's vocalizations in a specific and abstracted way.

INTRODUCTION

The reception of communication sounds, i.e., species-specific vocalizations, is an important aspect of the auditory behavior of primates, crucial for their social interactions, reproductive success, and survival (Altmann 1967; Andrew 1963; Fossey 1972; Gautier and Gautier 1977; Green 1975; Petersen 1982; Seyfarth et al. 1980; Smith et al. 1982; Snowdon 1982; Struhsaker 1967; van Lawick-Goodall 1968; Winter et al. 1966). Biologically important communication sounds of primates and many other mammals are often complex sounds with time-varying spectral features. Understanding the neural mechanisms underlying the representations of vocal communication sounds in the auditory system is a prerequisite for understanding the neural bases of primate auditory perception and for understanding the neural bases of representing spectrotemporally complex stimuli (including human speech) in general.

In analyzing the neural representations of complex sounds such as speech or animal vocalizations in the auditory system, two fundamental questions are often asked. What acoustic features are represented at each stage of the auditory pathway? In what form? At the level of the auditory nerve in mammals, it has been shown that spectral and temporal features of speech or speechlike complex sounds are represented by discharge patterns of fibers in the form of rate-place or temporal-place codes (Delgutte and Kiang 1984a–c; Miller and Sachs 1983; Sachs and Young 1979; Sinex and Geisler 1983; Young and Sachs 1979). The representations of steady-state vowels in the auditory nerve are further processed at the level of the cochlear nucleus, where each form of the representation (rate or temporal) has been found to be preserved or enhanced by specific neuronal subpopulations (Blackburn and Sachs 1990). However, although we have a growing understanding of how simple acoustic stimuli engage the primary auditory cortex (A1) (e.g., Eggermont 1991; Heil et al. 1992a; Imig et al. 1977; Merzenich and Brugge 1973; Merzenich et al. 1975; Nelken et al. 1994; Phillips et al. 1985; Schreiner and Mendelson 1990; Schreiner et al. 1992; Shamma et al. 1993a), it is virtually unknown how complex communication sounds are represented at this level of the auditory system, except in the echolocating bat, where significant progress has been made

in neural representation of biosonar signals (see Suga 1990 for review).

A1 plays an important role in processing species-specific vocalizations in primates (Heffner and Heffner 1986a,b). An understanding of neural mechanisms of encoding complex stimuli by A1 neurons is essential for understanding representational schemes in other cortical areas that further integrate A1 outputs as well as outputs from other cortical and extracortical areas. It should also shed some light on the significance of specific aspects of complex sound processing in the auditory brain stem. Furthermore, mechanisms responsible for representing vocalizations are important for understanding how cortical networks process spatially and temporally distributed sensory inputs in general. The results reported here represent our initial effort directed toward establishing a new conceptual and experimental framework for studying the neuronal representation of such complex stimuli in the auditory forebrain.

In studies of subcortical hearing systems, the behavioral relevance of an experimental stimulus has usually been ignored under the assumption that stimulus representations at these lower system levels are relatively static in adult animals, not subject to significant influences of the auditory environment and experience. This assumption clearly does not hold at the cortical level. Studies of cortical plasticity, especially in the past decade, have shown that the functional structures of even the primary sensory cortical areas, the primary somatosensory cortex (S1), A1, and the primary visual cortex (V1), are continuously modified by an animal's sensory experience, e.g., in S1 (Calford and Tweedale 1988; Clark et al. 1988; Jenkins et al. 1990; Merzenich et al. 1983, 1984; Pons et al. 1991; Recanzone et al. 1992a; Wall et al. 1983; Wang et al. 1994; Xerri et al. 1994), in A1 (Rajan et al. 1993; Recanzone et al. 1992b; Robertson and Irvine 1989; Weinberger and Diamond 1987), and in V1 (Gilbert and Wiesel 1992; Kaas et al. 1990). A dynamic, experience-modified cortex imposes an important constraint for experimental design: the familiarity and behavioral significance of stimuli must be taken into consideration. Given such a plastic nature of the cortex, a sound that an animal hears many times in its life and must continually identify (e.g., to recognize conspecifics or to respond by specific behavior) can be expected to be represented differently in A1 than would other similarly complex but rarely encountered stimuli. For primates in their natural environments, species-specific vocalizations are among the most behaviorally important and familiar sounds in life. In the present study we developed techniques to quantitatively evaluate physiological responses to natural vocalizations and to synthetic vocalization-like sounds.

Over the past several decades, a number of experimental attempts have been made to elucidate the forms of cortical representations of species-specific vocalizations in A1 of primates (see review by Newman 1988). The results of these studies have been mixed, with no clear or consistent picture emerging as to how behaviorally relevant complex sounds are "coded" or "represented" in the auditory cortex. This lack of success in earlier attempts may be accounted for in part, retrospectively, by earlier expectations on the form of cortical coding or representation of behaviorally important stimuli. For a time it was thought that primate vocalizations were encoded by individual "call detectors" (Newman and Wollberg 1973b; Winter and Funkenstein 1973). However, individual neurons in the auditory cortex were often found to respond to more than one call or to various "features" of calls (Glass and Wollberg 1979; Manley and Müller-Preuss 1978; Newman and Wollberg 1973a,b; Winter and Funkenstein 1973; Wollberg and Newman 1972). An alternative strategy of encoding complex sounds is by the discharge patterns of distributed neuronal populations (Creutzfeldt et al. 1980; Pelleg-Toiba and Wollberg 1991). There has been increasing evidence that distributed coding schemes operate in other sensory or motor cortical regions (Di Lorenzo 1989; Georgopoulos et al. 1989; Gochin et al. 1994; Merzenich et al. 1990a,b). Questions examined in this study relate to the representation of complex vocalization by the distributed populations of A1 neurons and to how these distributed responses might be linked together to serve as a basis for the perception of a specific vocalization in its entirety.

Our experimental model, the common marmoset (*Callithrix jacchus*), is a small, highly vocal New World monkey that has a well-described series of communication calls with dominant spectral components centered around 7–8 kHz, where their audiograms exhibit the lowest thresholds (Epple 1968; Seiden 1957). Vocalizations appear to be of great importance in the social behavior of this species. The vocalizations of adult marmosets are relatively stereotyped, but with dialects marking the calls of separate bands and tribes. The auditory fields in the marmoset are largely on the surface of the temporal lobe and thus can be easily accessed in electrophysiological recording experiments (Aitkin and Park 1993; Aitkin et al. 1986, 1988). Marmosets also have the important experimental advantage of being easily bred in captivity, making this species an attractive model for studying the ontogeny of their vocal repertoires and communication behaviors. The common marmoset is therefore an excellent primate model for studying the cortical coding of communication sounds.

In present report we focus on cortical responses to both natural and synthetic forms of one specific social call, designated the twitter call (Epple 1968), which is frequently involved in marmoset vocal exchanges and has interesting general features that have facilitated our study of how spectral and temporal features of complex calls are represented in A1. In particular, the spectral and temporal characteristics of the cortical representation of the twitter call are examined in detail.

METHODS

Vocalization recording and synthesis

Vocalizations from monkeys under study and their companions were recorded with the use of a digital tape recorder (Panasonic SV3700), with a sampling rate of 48 kHz and a 16-bit A-D conversion, and were transferred to a computer for editing using Sound Designer II software (Digidesign). Further analysis on individual vocalizations was performed on computer workstations (Silicon Graphic). The monkeys were housed in individual cages in a small ($N = 2–6$) colony at the University of California at San Francisco. In the present study, we focused on a social call designated as "twitter" (Epple 1968; see Fig. 1*A*). Twitter calls were frequently and loudly vocalized by marmoset monkeys in our colony and were a common element of vocal exchanges between marmoset pairs. An example of twitter calls recorded as part of a vocal exchange between a pair of marmosets is shown in Fig. 1*D*. Twitter

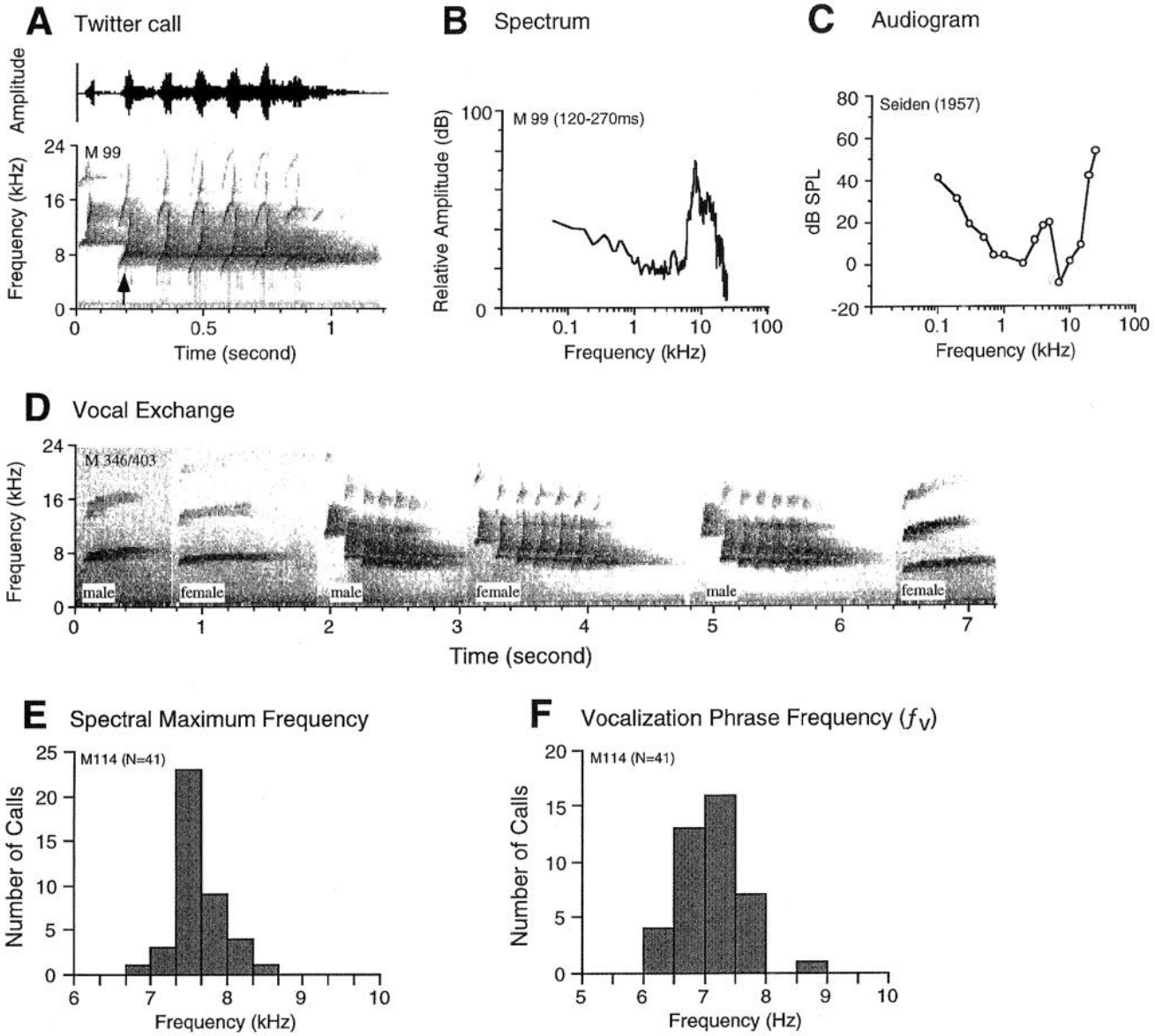

FIG. 1. *A*: example of a typical twitter call vocalized by an adult marmoset monkey in our colony. Both the waveform (*top*) and the spectrogram (*bottom*) of the call are shown. Eight call phrases are clearly visible in this call. Arrow: 2nd phrase. *B*: magnitude spectrum of the 2nd phrase of the twitter call shown in *A*. The spectrum has a maximum at ~7–8 kHz, typical of twitter calls recorded in our colony. *C*: audiogram of the common marmosets averaged from 5 animals (adopted from Seiden 1957). The lowest threshold is within a frequency range where twitter and other types of marmoset calls have the strongest energy. *D*: real-time recording of a vocal exchange between a pair of marmosets (1 male, 1 female), shown in the form of spectrogram. This typical vocal exchange, recorded while 2 marmosets were housed in separate cages in a room isolated from other marmosets and had no visual contact from each other, involved both twitter and other types of calls. *E*: distribution of the frequency at the spectral maximum of the 2nd call phrase for 41 twitter calls vocalized by a representative marmoset. The mean and SD of the spectral maximum frequency are 7.61 and 0.31 kHz, respectively. *F*: distribution of the vocalization phrase frequency (f_v) for the same group of twitter calls shown in *E* (mean, 7.16 Hz; SD, 0.48 Hz).

calls vocalized by an individual marmoset are highly stereotyped and have stable spectral features.

A typical twitter call consists of seven to nine phrases. By the word "phrase" (or event), we refer to a portion in a twitter call where acoustic energy clusters in time and is separated from other energy clusters of the call. The first phrase usually starts at a higher frequency than the subsequent ones. Each phrase is made of two to three frequency modulation (FM) sweeps at different frequencies. Each FM sweep within a phrase typically consists of two segments (see Fig. 3*F* for an enlarged view). The sweeping rates and shapes of these segments are different from animal to animal. The intervals between each phrase in a twitter call are relatively constant, with small variations. For the second and subsequent call phrases, the maximum of the spectrum is usually located at 7–8 kHz (Fig. 1*B*). In this frequency region, the audiogram of the marmoset has the lowest threshold (Seiden 1957; see Fig. 1*C*). An example of the distribution of the frequency at the spectral maximum from 41 twitter calls vocalized by one marmoset is given in Fig. 1*E*. The call's envelope, obtained using the Hilbert transform (Oppenheim and Schafer 1975), exhibited a local maximum reflecting the repetition frequency of call phrases in its spectrum. The frequency at this maximum is defined as the vocalization

phrase frequency (f_v). Twitter calls vocalized by marmosets in our colony typically have f_v values of 6–9 Hz. Twitter calls vocalized by each marmoset have a stable f_v value. A representative distribution of f_v measured from 41 twitter calls produced by one marmoset is shown in Fig. 1*F*. Twitter calls produced by different marmosets have the same spectrotemporal pattern but differ considerably in their idiosyncratic details.

A basic approach used to define response selectivity for call features was to parameterize calls in either the temporal or spectral domain. In this report we focus on manipulations in the temporal domain. Time-compressed or expanded calls were synthesized using the "phase vocoder" technique (Flanagan and Golden 1966), which allows the independent manipulation of signals in the time or frequency domain. The signal is modeled as a sum of sine waves with time-varying amplitudes and phases, as determined by a succession of overlapping short-time Fourier transforms (time window: 5.33 ms, 99.2% overlap). Variations of the original signal with altered spectral or temporal structure can be synthesized by summing sine waves with changed frequencies or amplitudes as a function of time. A sound analysis-synthesis software (SoundView, Peabody Computer Music Department, Johns Hopkins University) was used to generate synthetic vocalizations. In addition, time-reversed calls were generated by reversing a natural call in the time domain, i.e., playing the call backward.

A set of such parameterized stimuli was synthesized for each individual marmoset's vocalizations. Examples of temporally parameterized calls are shown in Fig. 2. The stimuli used included natural calls and a series of time-compressed, expanded, and reversed calls associated with each natural call. The series also included a call synthesized without altering temporal or spectral parameters that was used as the control stimulus for our stimulus synthesis procedure. Time-compressed or expanded calls have higher or lower f_v values, respectively, inversely proportional to the temporal alteration ratio, whereas the f_v of a time-reversed call is identical to that of a natural call. When a twitter call is compressed (expanded), the interphrase intervals are shortened (elongated) and the sweeping rates of FM segments within each phrase become faster (slower).

Surgical preparation

Adult marmoset monkeys were initially anesthetized with a mixture of 3% halothane-25% oxygen-72% nitrous oxide to induce a surgical level of anesthesia. An intravenous cannula was introduced into the brachial vein, and anesthesia was maintained throughout the experiment by intravenous injections of pentobarbital sodium diluted in lactated Ringer solution (1:5) as needed. Lactated Ringer solution with 5% dextrose was also continuously infused (2–5 ml/h) to maintain body hydration. Atropine sulfate (0.1 mg per 12 h) and penicillin-G (30,000 U per 24 h) were administered intramuscularly. Core temperature was monitored with a rectal probe and maintained at ~38°C. Heart and respiration rate were also monitored.

All recordings reported here were made from left hemispheres. To expose A1, the monkey was placed in a head holder. The lateral skull was exposed via a large skin flap, the left temporalis muscle was retracted, and the dorsoposterior aspect of the temporal lobe was then exposed via a wide craniotomy. The dura was resected to directly expose the A1 zone of the auditory cortex, which was maintained under a layer of viscous silicone oil. A magnified video image of the A1 zone was captured with a video camera (Cohu) and stored in a computer for use in positioning microelectrode penetrations relative to the cortical surface microvasculature.

Stimulus generation and delivery

All experiments were conducted in a double-walled soundproof room (IAC). Auditory stimuli were presented through a STAX-54 headphone enclosed in a small chamber that was connected via a sealed tube into the external acoustic meatus of the contralateral ear (Sokolich, US Patent 4251686; 1981). The sound delivery

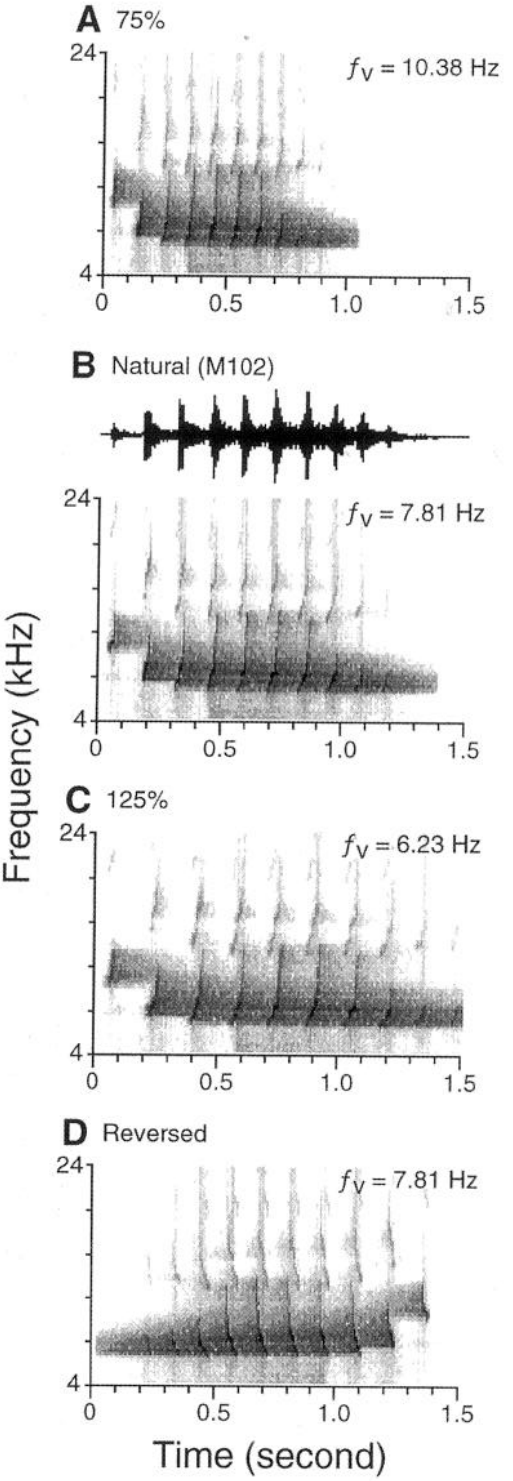

FIG. 2. Examples of a natural twitter call and its temporally altered variations (see METHODS). *B*: waveform and spectrogram of the natural twitter call. *A* and *C*: spectrograms of a compressed (75%) and an expanded (125%) version of this natural call (plotted up to 1.5 s), respectively. *D*: spectrogram of the time-reversed call. All spectrograms were computed using the same analysis parameters (time window, 5.33 ms; 90% overlap). Note that the temporally altered calls (*A*, *C*, and *D*) have the same spectral contents as does the natural call, but differ in their temporal characteristics. f_v value (Hz) computed for each call is shown on individual panels.

system was calibrated with a sound meter (Bruel and Kjaer 2209) and waveform analyzer (General Radio 1521-B). The frequency response of the system was nearly flat (within ±6 dB) up to 14 kHz, which spanned the frequency range for the main stimulus components of the twitter call. Above 14 kHz, the output rolled off at a rate of 10 dB/octave.

Experimental stimuli included *1*) short tone bursts used to derive frequency-intensity response areas and *2*) natural and parameterized synthetic vocalizations. Tone pips (duration 50 ms, onset time 3 ms, interstimulus interval 400–1,000 ms) were generated by a microprocessor (TMS32010, 16-bit D-A converter at 120 kHz). We used tonal stimuli to produce a complete frequency response area for each recorded unit by randomly presenting 675 tone bursts at different frequencies and intensities, which included 15 intensity levels in 5-dB steps and 45 frequencies in a 2- to 4-octave range (dependent on the estimated tuning curve width) centered around the estimated unit characteristic frequency (CF) (Schreiner and Mendelson 1990).

Vocalization stimuli were stored on the hard disk of a computer and delivered via a sound processing interface (Audiomedia II, Digidesign). They were presented to animals at various sound intensities (20, 40, and 60 dB SPL) set by a passive attenuator. The duration of natural and synthetic vocalizations ranged from ~0.5 to 2 s. Each vocalization stimulus was presented 10 times at each stimulus level, with an interstimulus interval of 2–3 s. Because the main objective of these experiments was to sample from as many cortical sites as possible in each animal, not every sound level was tested at each recording site because of the time limit. Complete data sets were obtained at 60 dB SPL for all recording sites in every animal. Vocalization stimuli were also applied at other sounds levels for a subset of sampled cortical sites. In some experiments, more than one natural twitter call and the variations were studied at each sampled site.

Recording procedure

Parylene-coated tungsten microelectrodes (Microprobe) with impedances of 1–2 MΩ at 1 kHz were used to record neuronal discharges from the A1 sector that covered a CF domain (1–2 octaves wide) representing the frequency range of the major spectral components of twitter calls. As in other primates (e.g., Imig et al. 1977; Merzenich and Brugge 1973), the marmoset's A1 is tonotopically organized (Aitkin et al. 1986). In general, more than two thirds of the relevant A1 zone was exposed on the lateral cortical surface of the temporal lobe in these monkeys. Most electrode penetrations were made in this portion of A1. An electrode was introduced into the cortex perpendicular to the cortical surface by use of a hydraulic microdrive. Neuronal activities of single or small groups of neurons (multiunits) were recorded in depths ranging from 700 to 900 μm, corresponding to cortical layers III and IV, and were isolated from background noise by an on-line window discriminator (BAK DIS-1); spike times were recorded by a data acquisition computer (PDP 11/73). A part of A1 extended into the lateral (sylvian) fissure. To record from this A1 sector, deep penetrations were introduced and responses recorded in the middle layers of the ventral bank of the lateral sulcus, with responses sampled at 100- to 200-μm intervals. Tracks of the electrode penetrations were confirmed by subsequent histology and used to extend the reconstruction of response maps into this A1 sector. In each animal we sampled unit responses from 35–140 cortical sites. In some experiments simultaneous recordings were made from a pair of microelectrodes. In that case each electrode was independently advanced into the cortex by use of two separate hydraulic microdrives.

Data analysis

Results presented here are based on the data recorded from seven adult marmosets. Unless specified, all vocalization responses reported in this paper were obtained at a sound level of 60 dB SPL. The analysis for this report was performed in both the spectral and the temporal domains and included the following procedures.

1) A frequency response area to tone pips was reconstructed for units recorded at all sampled sites (Schreiner and Mendelson 1990). Basic parameters of tuning characteristics were objectively derived from this frequency response area (CF, threshold, tuning bandwidth, maximum discharge rate, etc.).

2) Temporal properties of the responses to vocalization stimuli were analyzed in the form of poststimulus histograms (PSTHs), which were further subject to spectral analysis. f_v was computed for each vocalization stimulus. The amplitude of the spectral component of the response at f_v, obtained from the spectrum of a PSTH computed using a discrete Fourier transform, was used as a measure to quantify stimulus-following responses and is referred to as the "synchronized discharge rate" or R_s. The R_s measure used here is similar to a "vector strength" measure (e.g., Goldberg and Brown 1969; Kim and Molnar 1979) except that the latter is normalized by the average discharge rate of the histogram. Therefore R_s (in spikes/s) measures a portion of the total discharges that are synchronized to the repetitive events in a vocalization stimulus. To assure that our R_s measure was not significantly affected by small variations in interphrase intervals of a call, we also tested in some samples a second measure that summed the number of spikes within a narrow time window after the onset of each call phrase. This alternative measure of stimulus-following response gave similar results to those obtained with the use of the R_s measure (data not shown). Because R_s was more convenient to compute, it was used in this report.

3) Time-varying and spectrally distributed cortical responses evoked by complex vocalization stimuli were reconstructed in three-dimensional graphs to form spectrotemporal discharge patterns in which discharges from all sampled cortical units were aligned along one axis according to their CF, and along another axis of time. The spectral properties of responses were further quantified by profiles of discharge rate versus unit CF, computed over variable time windows.

4) To evaluate the selective effectiveness of a natural call for driving A1 neurons, the response to a natural call was compared with those to other parameterized synthetic variations. Two types of "selectivity" were defined. *1*) Units were called type F (for "forward selective") neurons if they responded more strongly (measured by R_s) to a natural call (forward) than to its time-reversed variation (backward). Units that did not meet type F criterion were called F^- (for non-type F). The entire sample was divided into F type and F^- types. *2*) Units were called type S (for "selective to time course") neurons if they responded with a greater R_s to a natural call than to all other time-compressed, expanded, and reversed variations. The type S units were thus a subset of type F units.

RESULTS

Spectrotemporal discharge patterns

An important observation of the present study is that the spectrotemporal neuronal discharge pattern of A1 units was correlated with the spectrotemporal acoustic pattern of a twitter call. A1 units usually responded to a vocalization with a discharge rate above the background firing rate if the stimulus' spectrum extended into a unit's excitatory receptive field. Figure 3*A* displays two representative PSTHs from a high-CF unit (11.4 kHz) and a low-CF unit (7.0 kHz). The low-CF unit (Fig. 3*A*, *bottom*) did not respond to the first call phrase, which had no energy below ~10.5 kHz. This unit had strong responses to phrases 2–5. The high-CF unit (Fig. 3*A*, *top*), on the other hand, responded

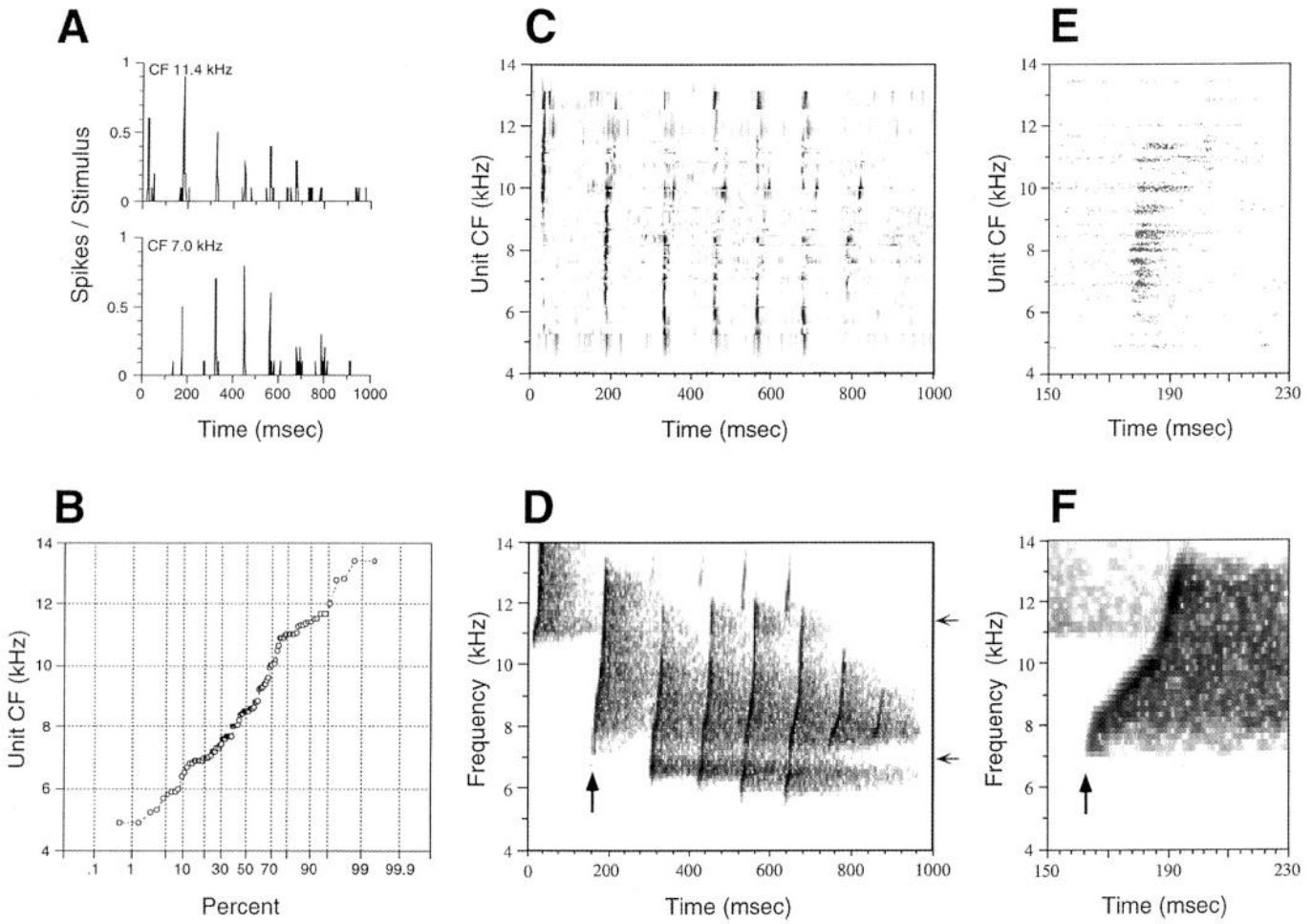

FIG. 3. Comparison between the spectrotemporal acoustic pattern of a twitter call and the corresponding spectrotemporal discharge pattern recorded in the primary auditory cortex (A1) of 2 marmosets. This call was vocalized by a male marmoset (*M346*) housed with a female marmoset (*M403*) in separate cages in a room isolated from other marmosets. Cortical responses to the same twitter call were recorded from A1 of both the male and female marmoset in 2 separate experiments. *A*: poststimulus histograms (PSTHs) of a high-characteristic-frequency (CF) unit (*top: unit M403-U4,* CF 11.4 kHz) and a low-CF unit (*bottom: unit M346-U11,* CF 7.0 kHz) in response to the twitter call presented at 60 dB SPL. The positions of the CFs of these 2 units, relative to the spectrogram of the call, are marked by the 2 horizontal arrows on the *right* side of *D*. *B*: CF distribution of the cortical units ($N = 100$) recorded in 2 experiments, in the form of cumulative percentages. The frequency range of major stimulus components in this twitter call was extensively and relatively evenly sampled. The responses of these units to the male marmoset's twitter call are shown in *C* and *E*. *C*: population responses to the male marmoset's twitter call. Discharges as they occurred in time (abscissa) from individual cortical units, computed in PSTHs, are aligned according to their objectively defined CF (ordinate). This plot was made using a graphic software (Spyglass) with interpolation between neighboring elements. The gray level in the plot is proportional to the number of spikes in each bin (binwidth = 2.0 ms). The sound level was 60 dB SPL. *D*: spectrogram of the male marmoset's twitter call used to produce the neural responses shown in *A, C,* and *E* is shown with the same frequency and time scales as in *C*. Vertical arrow: 2nd call phrase. The 2 horizontal arrows at *right* mark CFs of the 2 units shown in *A*. *E*: expanded view of cortical responses to the 2nd phrase of the twitter call shown in *D*. Responses of the same group of cortical units shown in *C* are included, but displayed in the form of dot roster. Each recorded spike occurrence within the time period shown (150–230 ms) is marked as a dot. Spike times from 10 repetitions are aligned on this figure along 10 lines centered at the CF of the unit, shifted by 10 Hz for each repetition (i.e., positioned from CF − 50 Hz to CF + 40 Hz in 10-Hz steps). *F*: expanded view of the spectrogram of the 2nd call phrase, with a time mark indicated by a vertical arrow as in *D*.

strongly to the first phrase and more weakly to phrases 3–5 than did the low-CF unit, the latter a consequence of the weaker stimulus energy in the frequency range around its CF. The frequency sensitivity of A1 units that gives rise to underlying tonotopic organization in this cortical area predicts an orderly relationship between population discharges and the stimulus spectral features.

In Fig. 3, *C* and *D,* spectrotemporal patterns of both the stimulus and the distributed neuronal responses are compared for a twitter call. This twitter call was vocalized by a male marmoset housed together with a familiar female marmoset in separate cages but away from other monkeys for a period of several weeks before a physiological mapping experiment. There are eight phrases in this twitter call, as shown by its spectrogram (Fig. 3*D*). Each of these phrases consisted of an FM sweep that extended across a range of a few kHz in ~30 ms (Fig. 3*F*). Note that the first FM sweep in the first call phrase began at ~10.5 kHz, whereas later sweeps were initiated at lower frequencies. Population responses evoked by this twitter call were recorded from A1

of both the male and female marmosets. PSTHs of discharges recorded from all 100 sampled A1 units in response to this call delivered at 60 dB SPL are aligned along the vertical axis according to each unit's CF, to form a spectrotemporal firing pattern display (Fig. 3*C*). The gray level of the figure corresponds to the strength of the discharge rate at each histogram bin. The distribution of the CFs of units included in Fig. 3*C* is plotted in Fig. 3*B*, which shows that the frequency range across which the twitter call had significant energy was completely and densely covered by our neuronal response recordings in this (and other) experiment(s). Units with CFs <4 kHz were not studied because there was little response to the twitter call played at 60 dB SPL in that CF range.

Each call phrase is marked by discharges across A1 by units with appropriate frequency tuning properties. For instance, the first phrase primarily evoked responses from units with CFs greater than ~9 kHz, whereas the second phrase evoked responses from units with CFs greater than ~5 kHz. Because most tuning curves were fairly broad at the intensity of the vocalization used to produce the responses shown in Fig. 3*C*, the spread of discharges as marked by unit CF was modestly wider than the range of spectral energy.

Although many units with different CFs were activated by each call phrase, cortical discharges in these units were highly synchronized to each other, as indicated by the vertical alignment of discharges in Fig. 3*C*. This property is more clearly illustrated in Fig. 3, *E* and *F*, where an enlarged view of a single call phrase and the corresponding response patterns are shown. The FM sweep in this typical call phrase had two segments, one sweeping from ~7 to ~10 kHz and the second from ~10 to ~13 kHz. The timing of responses was restricted to a narrow window of ~10 ms and was largely independent of unit CF across a large neuronal population. Implications of this response synchrony for the overall representation of this wide-band complex call are discussed in the DISCUSSION. Notice that the majority of the responses were produced by the initial segment of the FM sweep. For units with CFs >10 kHz, there was also a later and weaker response component, apparently evoked by the second FM segment. Cortical responses shown in Fig. 3, *C* and *E*, were recorded sequentially from a number of units. The fact that they were all synchronized to stimulus events implies that they were synchronized to each other as well, as verified by simultaneous recordings from pairs of cortical units in some of our experiments (data not shown).

The response display in Fig. 3*C* emphasizes clusters of neural activities evoked by individual call phrases. A1 neurons also discharged, although with a lower probability, at other times during the course of a twitter call. We document the synchronized discharges shown in Fig. 3*C* in a different form in Fig. 4. Figure 4 has the same horizontal and vertical axes and includes the same population of A1 units as does Fig. 3*C*, but the responses of individual A1 units in the form of histograms aligned along the vertical axis are now shown by dots rather than by gray level, so that weaker firing activities between clusters of strong discharges are more evident. Each of the panels in Fig. 4 displays responses above a certain discharge rate threshold, i.e., only those histogram bins in which discharge rate exceeds the indicated threshold appear as dots. As the threshold becomes higher from Fig. 4, *A* to *E*, fewer dots representing stronger responses remain

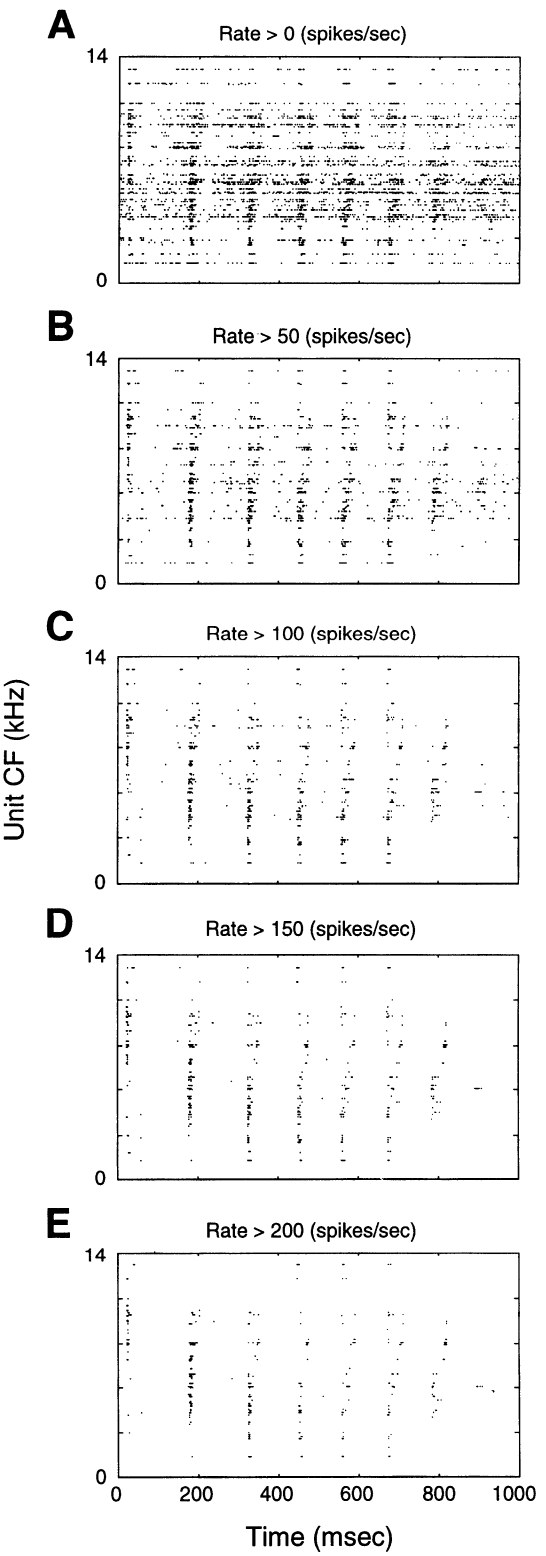

FIG. 4. Population responses shown in Fig. 3*C* are replotted in a different form. Abscissa and ordinate are the same as in Fig. 3*C*. The occurrences of discharges (accumulated in PSTHs with a binwidth of 2.0 ms) are marked by dots for individual bins. Each panel represents responses greater than a given driven discharge rate (total discharge rate minus spontaneous firing rate), i.e., only the bins that have the driven discharge rate above a display threshold (indicated above each panel) are marked.

on the plot. As illustrated by Fig. 4*A*, the characteristic response features we see in Fig. 3*C* are "buried" in neuronal firings across the entire responding populations and through-

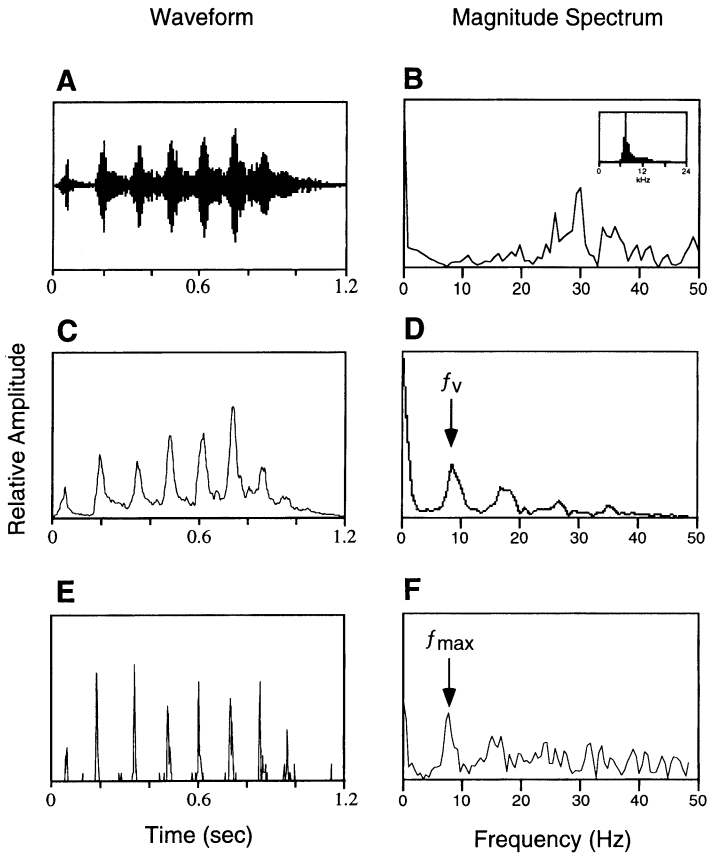

FIG. 5. Example illustrating the relationship between temporal aspects of twitter calls and the responses of individual A1 units. *A* and *C*: waveform and envelope, respectively, of a typical twitter call. *B* and *D*: corresponding magnitude spectra for *A* and *C* in the frequency range of 0–50 Hz. *B, inset*: entire spectrum (0–24 kHz) of the twitter call. *E*: representative PSTH from a cortical unit (*M99-U4,* CF 14.4 kHz) in response to the twitter call. *F*: magnitude spectrum (0–50 Hz) of the response in *E*. The spectrum of the envelope exhibits a local maximum at ~8 Hz, typical of the many twitter calls we recorded and reflecting the repetitive phrases in the call. The spectrum of the PSTH shown also had a local maximum at ~8 Hz. The frequencies corresponding to these 2 local spectral maxima (*D* and *F*, ↓) were defined as f_v and the maximum response frequency (f_{max}), respectively (see METHODS). The stimulus itself does not exhibit a local maximum at or near 8 Hz; its main energy is centered around 7–8 kHz (*B, inset*).

out the entire course of the call. The discharges between call phrases appear to be nearly random, whereas those occurring shortly after each phrase are fairly robust. This stability and strength of phrase-triggered discharges is evident in Fig. 4, *B–E,* where successively stronger responses are shown. The properties of these stimulus related firing patterns thus suggest specific integration models for neurons at higher cortical stages that read such outputs from A1.

Temporal characteristics of stimulus and response

The data shown in Figs. 3 and 4 clearly demonstrate the correspondence between the acoustical structure of a vocalization and the discharge pattern of distributed populations of A1 units. At the same time, the response patterns to twitter calls show that A1 neurons did not follow the fine details of the stimulus waveform, but have their responses phase-locked to call phrases (and to each other). The time course of call phrases is best described by the envelope of a vocalization. The temporal characteristics of a twitter call and its corresponding cortical responses are analyzed in Fig. 5, in which the stimulus, its envelope, and the response of a typical A1 unit are compared. The A1 unit shown responded strongly to each call phrase (Fig. 5*E*). This repetitive firing was reflected as a local maximum in the spectrum of the PSTH at ~8 Hz, which was the largest spectral peak outside the 0 Hz (dc) region (Fig. 5*F*). The frequency at this local maximum was defined as the maximum response frequency (f_{max}). The spectrum of the PSTH has no large components at >50 Hz. The spectrum of the stimulus itself does not show any significant energy at ~8 Hz (Fig. 5*B*), but had large components in the 7- to 10-kHz range (Fig. 5*B, inset*). The spectrum of the stimulus envelope, on the other hand, had a local maximum (the largest spectral peak outside the dc region) in its spectrum at a frequency of ~8Hz (Fig. 5*D*), reflecting the repetition rate of the phrases in the stimulus (Fig. 5*A*). The frequency at this maximum is the f_v as defined in METHODS.

The relationship between f_{max} and CF of each unit is analyzed in Fig. 6*A* for responses to a natural twitter call and several of its synthetic variations, all recorded in one marmoset. f_{max} values were higher for the compressed call (△) and lower for the expanded call (□) compared with those for the natural call, reflecting temporal alterations in these synthetic stimuli. The time-reversed call (×) resulted in f_{max} values similar to those of the natural call (●). Thus the discharges evoked by individual phrases in a reversed call had similar repetitive frequency as compared with the natural call, but the details of firing patterns evoked by these two calls are very different (see Fig. 7). Figure 6*A* also shows that f_{max} is not correlated with unit CF. f_{max} of the units sampled across a wide range of CFs scattered roughly evenly around its mean value (dashed line) in each stimulus condition. The mean f_{max} computed from cortical responses under each stimulus condition was found to be approximately equal to f_v of the stimulus (Fig. 6*B*). This response-stimulus relationship applied to all units for which this analysis was conducted. To simplify our analysis, f_v of each call stimulus was used to compute R_s for all cortical units regardless of their CFs (see METHODS). The analysis in Fig. 6 illustrates that our parametric manipulation of a complex natural twitter call resulted in systematic changes in some temporal aspects

of discharge patterns in A1 neurons. However, there are important differences in the responses evoked by natural and temporally altered calls, as described below.

Natural versus synthetic vocalizations

One of the most important observations from the present study is that natural vocalizations and their temporally altered variations produced differential responses: a subset of units sampled in each marmoset responded more strongly and more coherently to a natural vocalization. Compound PSTHs from 14 such A1 units in response to a natural twitter call and several of its synthetic variations are shown in Fig. 7 for a representative marmoset. These cortical units exhibited

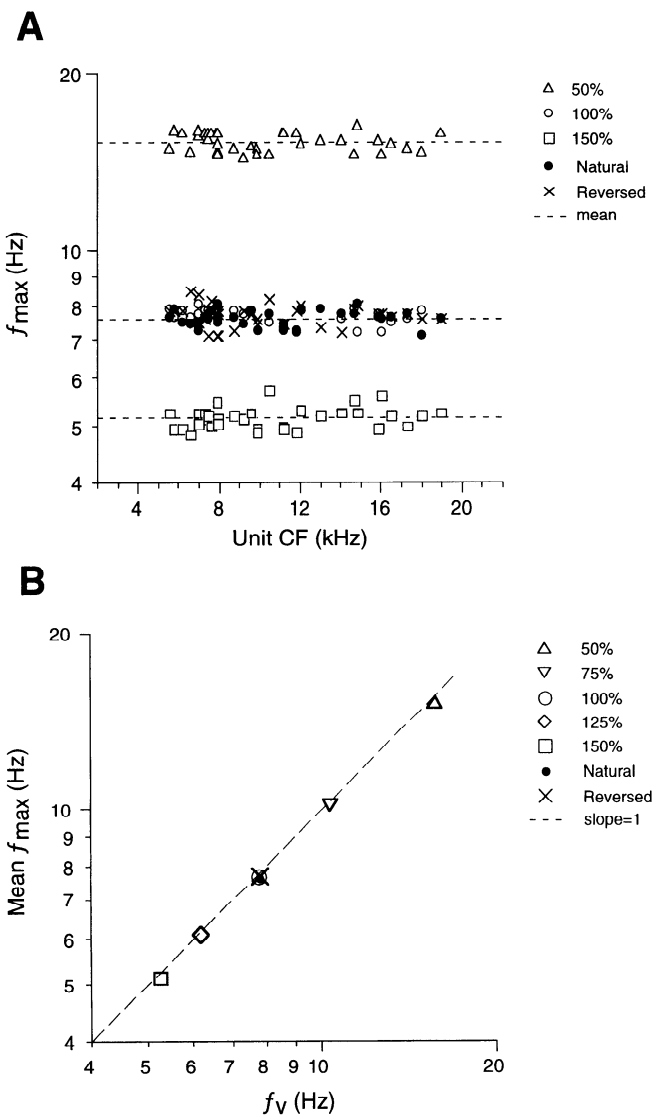

FIG. 6. *A*: f_{max} plotted on a logarithmic scale vs. unit CF for a group of cortical units ($N = 35$) recorded in 1 marmoset (*M102*) in response to a natural twitter call and several of its temporally altered variations. Dashed lines: mean values for 50%, natural, and 150% calls. Symbols are explained at *right*. *B*: mean f_{max} is computed from the same group of cortical units shown in *A* and plotted vs. f_v for each vocalization stimulus, both on a logarithmic scale. The natural, 100% (control stimulus), and time-reversed calls have nearly identical mean f_{max} values (7.59, 7.64, and 7.69 Hz, respectively). Dashed line has a slope of 1 and passes through (0, 0) point. Mean f_{max} and corresponding f_v have approximately equal values. Symbols are explained at *right*.

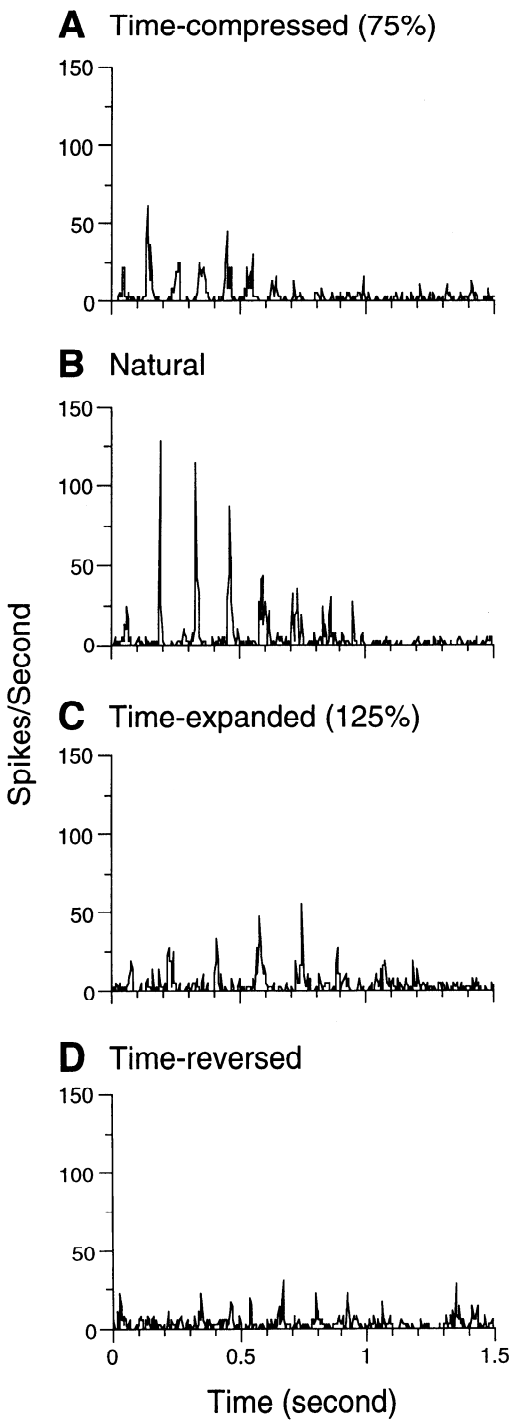

FIG. 7. Examples of compound PSTHs generated by summing individual PSTHs derived from 14 type S units recorded in 1 marmoset (*M102*) in response to a natural twitter call and to several temporally altered variations of this call. The spectrograms of the vocalization stimuli that were used to derive the response histograms in *A–D* are shown in Fig. 2, *A–D*. Stimulus onset was at *time 0*.

selective responses to the natural call. Their responses to a time-compressed (75%, Fig. 7*A*) and a time-expanded call (125%, Fig. 7*C*) were weaker than were responses to the natural call (Fig. 7*B*). The time-reversed call produced the poorest response (Fig. 7*D*). Cortical responses to a synthetic control stimulus, i.e., a 100% twitter call synthesized without altering any spectral or temporal structure, did not show significant differences as compared with the responses to a

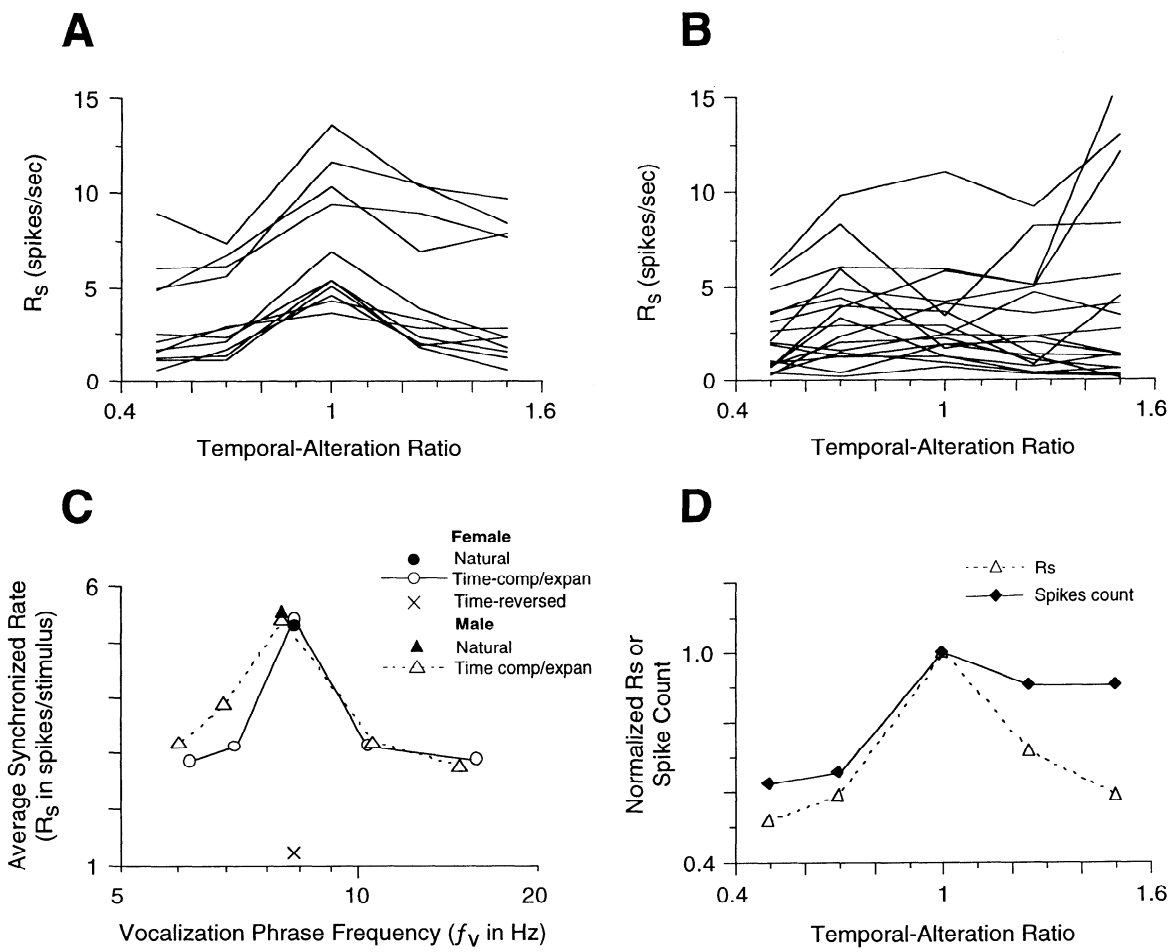

FIG. 8. *A*: synchronized discharge rate (R_s) of responses to a twitter call and its variations plotted vs. temporal alteration ratio (equal to 1.0 for a natural call) from a group of type S units recorded in a marmoset (*M102*). *B*: R_s vs. temporal alteration ratio functions are shown for cortical units that were recorded from the same marmoset as in *A* but were not selective for the natural call. *C*: averaged R_s, computed for 2 subpopulations of type S units recorded in 1 female marmoset (*M102*), plotted vs. f_v on a logarithmic scale. One subpopulation was selective for the female's own call (○——○, $N = 9$), the 2nd subpopulation was selective for the male companion's call (△ - - △, $N = 14$). Ordinate has the unit of spikes per stimulus for the purpose of comparison, because the durations of 2 twitter calls are different. Note that the 2 natural calls had different f_v values. Time-compressed or expanded calls had larger or smaller f_v values compared with that of the natural call. A total of 35 cortical units was sampled in this marmoset at 35 separated recording sites in a zone of A1 representing major spectral elements of the calls (also see Fig. 11). *D*: comparison between measures of vocalization-evoked responses based on R_s (∧ - - ∧) or total spikes counts (♦——♦). Average R_s and spike count ($N = 14$) are normalized with respect to those of the natural call and plotted vs. the temporal alteration ratio for the type S subpopulation selective for the marmoset mate's call. The differences between the natural and other altered calls are larger for R_s measure than for spike count measure.

natural twitter call (see Fig. 9). To quantify the relative strength of the evoked response for each stimulus condition, R_s was computed for every cortical unit.

Examples of quantified selective responses are given in Fig. 8*A*, in which R_s is plotted as a function of the temporal alteration ratio. These R_s functions reached a maximum for the natural call and had various magnitudes. Units that exhibited such selective R_s functions had their CFs across the call's spectrum. As will be seen later, these properties are important for the selective subpopulation in representing spectral information of a call. Examples of responses from units recorded in the same experiment that were not selective for the natural call are shown in Fig. 8*B*. The selective response defined on the basis of individual units can be found for more than one natural call. The averaged selective R_s functions for two natural calls tested in a marmoset are shown in Fig. 8*C*. In this case, responses to a female's own call and to the male companion's call were recorded from the cortex of the female marmoset. Note that the two calls had different f_v values. One subpopulation of recorded units was found to be selective for the female's call and a second subpopulation for the male's call.

R_s is a measure of how closely in time the response of an A1 unit follows call phrases. If a measure of total spike counts other than R_s was used, the differences between responses to natural and altered calls were equivalently great for time-compressed calls and less great for time-expanded calls (Fig. 8*D*). This is because when the total number of spikes was counted during the stimulus period, discharges not closely associated with the ongoing stimulus activity were also included. Such nonsynchronized, nearly random discharges did not contribute to our R_s measure.

The CFs of our sampled units were within the frequency range subtended by call spectra. Two types of selective subpopulation were analyzed. About 30% of the units studied in each animal were classified as "type S" for a particular

natural call, i.e., their R_s to this natural call was greater than R_s for all time-compressed, time-expanded, and reversed calls. A much higher percentage of the sampled units (75%) responded more strongly to natural calls than to reversed calls as measured by R_s. These units were defined as "type F." The means and standard deviations for the occurrences of these two types of units in all experiments are summarized in Fig. 9*A*. By definition, type S units were a subset of type F units. Because type S units were defined on a much more restricted condition than type F units, they were more sensitive to changes in temporal course of a twitter call. The statistics of relative R_s magnitude in different stimulus conditions computed for the type S subpopulation are shown in Fig. 9*B*. As a temporally altered call was modified away from the natural call, R_s declined systematically. The responses evoked by control stimuli (100%) had nearly the same mean amplitude as the natural call. The comparison of response magnitudes between natural calls and their time-reversed variations in type F units is shown in Fig. 9*C*. A1 units in this subpopulation responded to the reversed call at less than half-strength compared with their responses to the natural calls as measured by R_s.

The differences in the strength of R_s produced by various stimuli as seen above on the basis of individual units resulted in different population responses. In Fig. 10 we analyze A1 responses to a natural call (*left*) as well as its reversed version (*right*) for two classes of units recorded in one experiment. Both the mean PSTH (*top*) and the population discharge pattern (*bottom*) are shown in Fig. 10, *A*–*F*. Individual phrases in the natural call produced strong synchronous discharges in units across a wide range of CFs (Fig. 10*A*). The synchronous responses were contributed largely by type F units, indicated by a clearer pattern of synchronization in the population discharges and stronger responses to each call phrase in the mean PSTH (Fig. 10*B*). When this subpopulation was removed, the rest of the sampled units (i.e., type F^-) showed only weak responses and relatively poor discharge synchronization to the natural call (Fig. 10*C*).

Population responses to time-reversed calls were generally much weaker and not well synchronized to call phrases, as illustrated by an example in Fig. 10*D*. If we separate the sampled units into type F and type F^-, the mean response for type F^- units was slightly stronger than were the responses for type F units (Fig. 10, *E* and *F*). But even for type F^- subpopulation whose members exhibited stronger responses for the time-reversed call on an individual basis, the population responses were not as well synchronized across frequency channels as were the responses to the natural call in the type F subpopulation. Furthermore, the difference in response between type F and F^- units for the time-reversed call (Fig. 10, *E* and *F*) was far smaller than that for the natural call (Fig. 10, *B* and *C*), suggesting that the two types of units are not equivalent with respect to their relationships with the natural and reversed calls. That is, the responses of type F units are more "preferential" to the natural call than were the responses of type F^- units to the reversed call.

Different type S (or type F) subpopulations can be defined for different natural calls in each animal. The distribution of a selective subpopulation did not appear to be random. In Fig. 11 we compare the spectral distribution

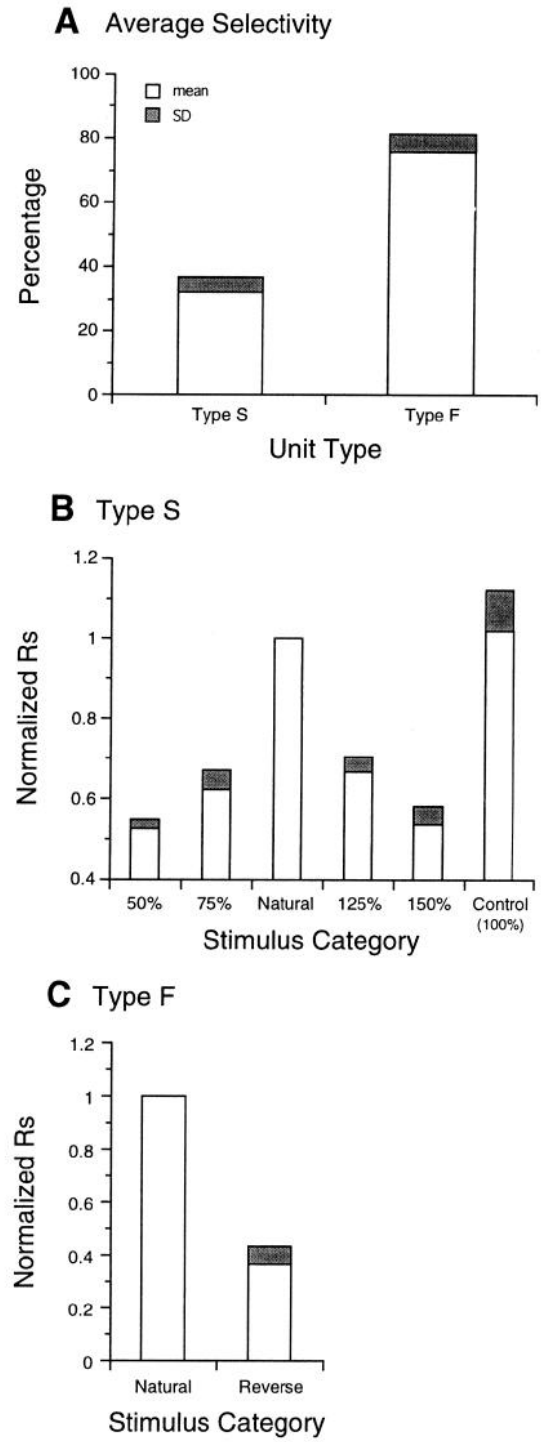

FIG. 9. *A*: overall distribution of A1 units with 2 types of selectivity for all experiments. The mean and SD of the percentage of each unit type are shown. On average, 32.23 ± 4.75% of the sampled units were type S in each experiment (229 of 728 samples studied for 8 natural calls); 75.88 ± 5.76% were type F in each experiment (493 of 655 samples studied for 6 natural calls). See METHODS for the definition of the selectivity. *B*: relative strength of R_s in different stimulus conditions for type S units averaged from all experiments, normalized by R_s for the natural call. The control stimulus (100%) was a synthetic "natural call" produced using a temporal alteration ratio of 1.0. Normalized R_s values are 50% (0.53 ± 0.02), 75% (0.62 ± 0.05), 100% (1.02 ± 0.10), 125% (0.67 ± 0.04), and 150% (0.54 ± 0.04). *C*: relative strength of R_s in 2 stimulus conditions for type F units averaged from all experiments, normalized by R_s for the natural call. Normalized R_s values for time-reversed calls are 0.37 ± 0.07.

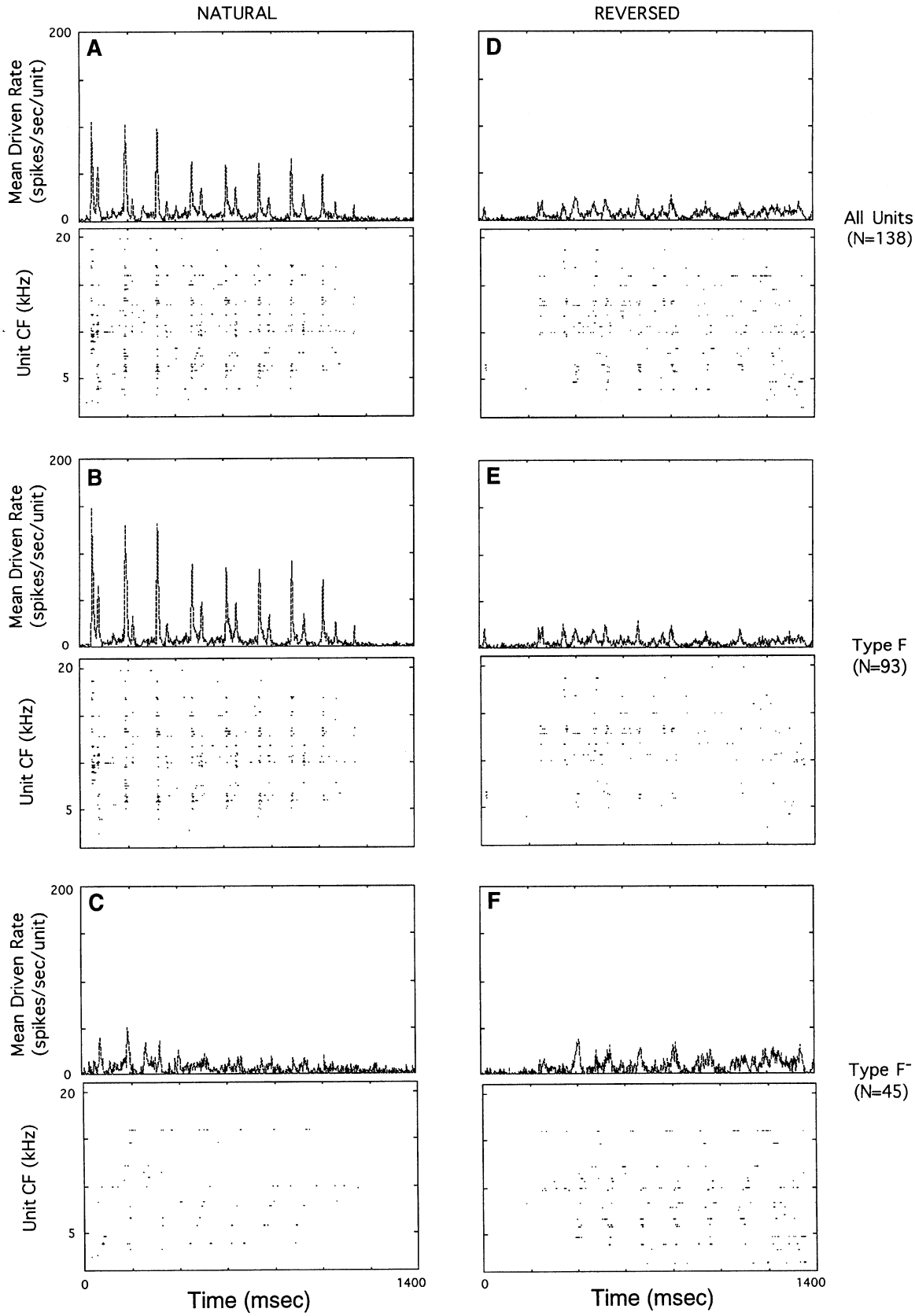

FIG. 10. Population responses to the natural twitter call of a conspecific monkey and its time-reversed version for several unit types. A total of 138 units recorded from this marmoset (*M115*) were included. In *A–F,* the *top* is the mean PSTH (binwidth = 2.0 ms) averaged over all sampled units, the *bottom* shows the population discharge patterns plotted in the same format as in Fig. 4 (display threshold, 200 spikes/s). Responses to the natural call are plotted at *left,* those to the time-reversed call at *right. A* and *D*: all units ($N = 138$). *B* and *E*: type F units ($N = 93$). *C* and *F*: type F^- units ($N = 45$).

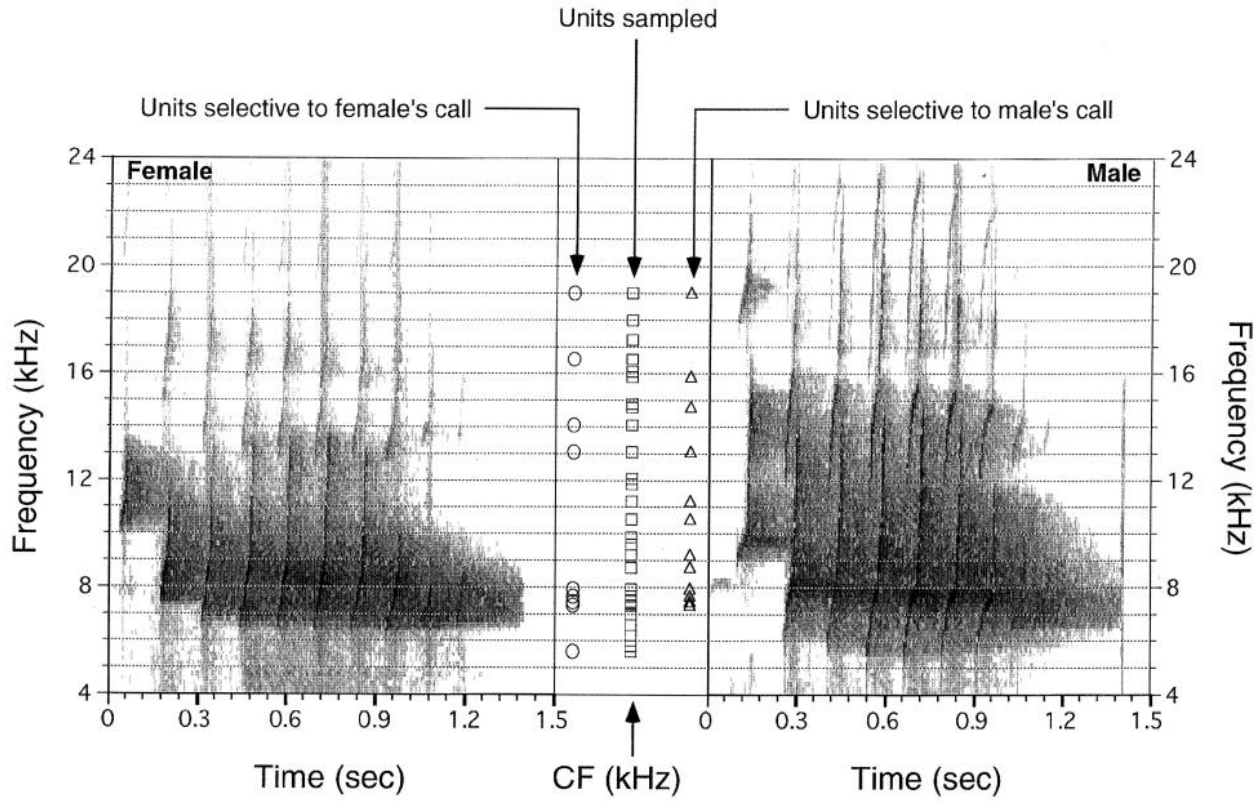

FIG. 11. Illustration of 2 partially overlapping type S subpopulations that were selective for a female monkey's own call (*M102*) and the male companion's call (*M99*). Spectrograms of the distinct twitter calls from these 2 monkeys are shown side by side. All the cortical units ($N = 35$) sampled in the female marmoset are marked by their CF values in the middle column (□), aligned with the frequency axes of the 2 spectrograms by dotted lines. These units were sampled in separate sites across A1. The type S subpopulation selective for the female monkey's call ($N = 9$) is marked by their CFs at *left* (open circles) and the type S subpopulation selective for the male's call ($N = 14$) is marked at *right* (open triangles).

of two type S subpopulations recorded in one marmoset that were found to be selective for two twitter calls: a monkey's own call and its colony companion's call. Locations of all units are marked in terms of their CFs in the middle of the figure (□). Units from the two type S subpopulations are marked by open circles and triangles, respectively. The selective units were clustered where the spectrum of a given call stimulus has high energy. Figure 11 also serves to illustrate that subpopulations that are selective for different natural calls are partially overlapping in their distributions in A1, i.e., they share some but not all of their neuronal memberships with each other. Thus any given A1 unit can belong to none, one, or more than one of such subpopulations.

Spectral characteristics

A twitter call was composed of several phrases. As illustrated in Fig. 3*C*, A1 units integrate stimulus energy within each call phrase while individually and collectively marking the timing of each phrase. To analyze the neuronal representations of spectral features, comparisons between the stimulus spectrum and its evoked response must therefore be made within short but significant periods of time. In Fig. 12 we analyze the spectral representation of a segment of a natural twitter call. The magnitude spectrum of this first call phrase is plotted in Fig. 12*A*, which has a main peak at ~8 kHz and a second peak near 16 kHz, corresponding to the two FM sweeps in this phrase. Driven discharge rates of all recorded A1 units in response to this phrase of the twitter call are shown in Fig. 12*B* (○). A rate-CF profile (solid line) was obtained by using a triangular weighting window that was 0.25 octaves wide at the base and moved in 0.125-octave steps, and that included at least three units in each window (Sachs and Young 1979). This profile shows a large peak between 8 and 9 kHz and a smaller peak at ~16 kHz, signaling the corresponding spectral peaks of the two FM sweeps in the stimulus.

We separated A1 units into different types according to the response criteria for selectivity defined in METHODS. Figure 12*C* shows the rate-CF profiles of all units (solid line) and type S units (solid line with ●). The rate-CF profile for type S units gives just as clear an indication of the main spectral peak as does the rate-CF profile for all units. There were few type S units in the CF range >12 kHz. Figure 12*D* compares rate-CF profiles for all units (solid line), type F units (solid line with △) and type F^- units (○ - - ○). The rate-CF profile for type F units closely resembles that for all units, whereas the rate-CF profile for type F^- units did not strongly mark the main spectral peak. The second FM sweep resulted in a clear peak in the type F profile, but not in the type F^- profile.

The same analysis applied to the first call phrase in Fig. 12 was also applied to the second call phrase of the same twitter call, with the results shown in Fig. 13. Major spectral features were again reflected in rate-CF profiles of both type S and type F units (Fig. 13, *C* and *D*), but in this case were

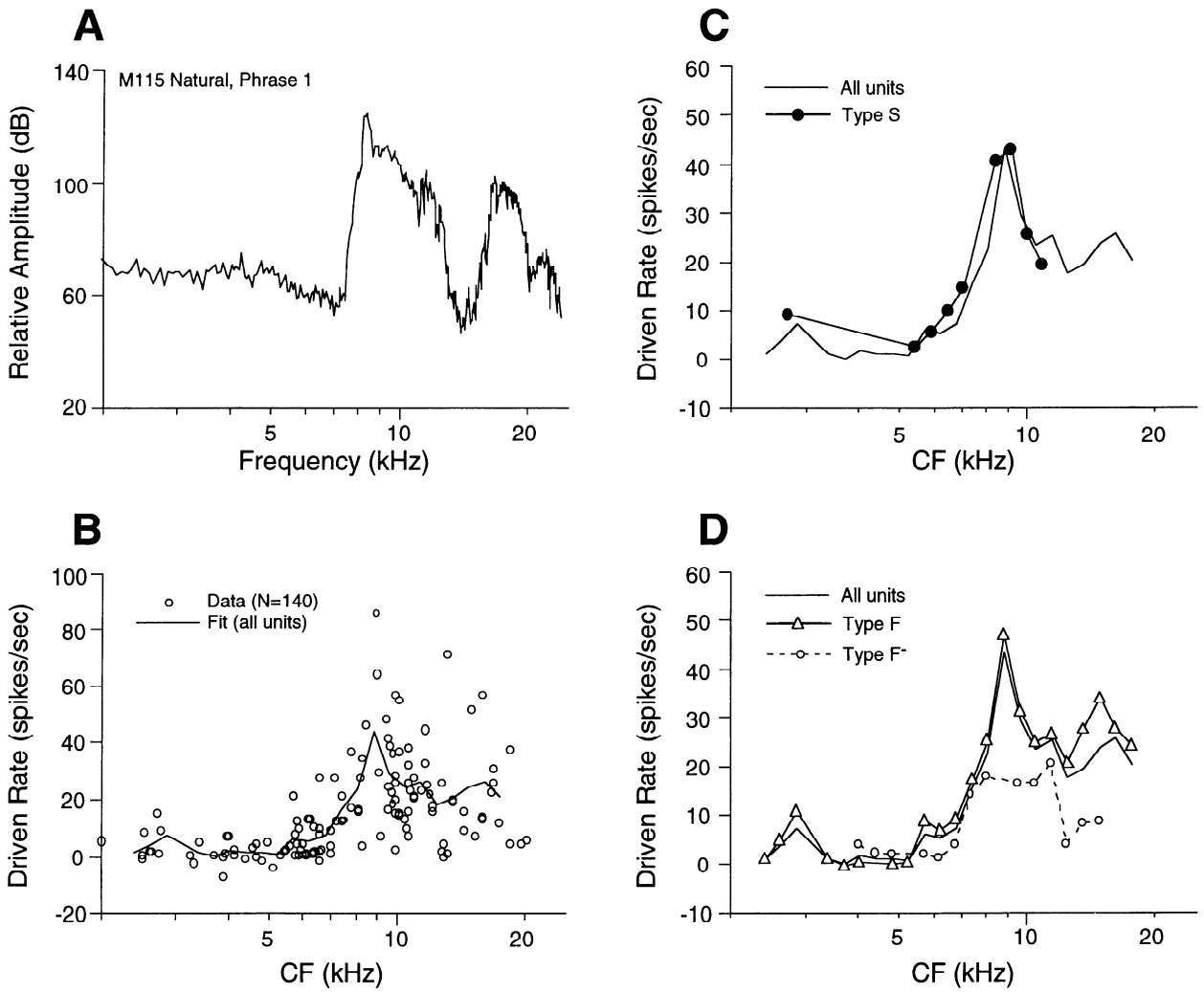

FIG. 12. Comparison of short-term call spectrum and rate-CF profiles computed over a period of time that includes the 1st call phrase. Data shown were obtained from 1 marmoset (*M115*) in responses to its own twitter call. *A*: magnitude spectrum of the 1st call phrase of the natural twitter call. *B*: driven discharge rates of responses plotted vs. unit CF for all sampled A1 units (N = 140, ○). Solid line: rate-CF profile computed using a triangular weighting window whose base was 0.25 octaves wide. The centers of adjacent windows were 0.125 octaves apart. Only averages that had ≥3 units in the window were included. *C*: rate-CF profiles are shown for all units (N = 140, solid line) and type S units (N = 37, solid line with ●). *D*: rate-CF profiles are shown for all units (N = 140, solid line), type F units (N = 102, solid line with △), and type F⁻ units (N = 38, ○ - - ○).

only weakly indicated when all units were considered (Fig. 13*B*) and were represented almost not at all by type F⁻ units (Fig. 13*D*). For this call phrase, both FM sweeps in the phrase evoked strong responses.

To investigate the time course of the spectral representation, we carried out the rate-CF profile analysis for all call phrases of the same natural twitter call analyzed in Figs. 12 and 13. This call had nine phrases, whose spectra are shown in Fig. 14*A*, *bottom* to *top* (phrase 1–9). The rate-CF profiles for all units, type S units, and type F⁻ units are shown in Fig. 14, *B–D*, respectively. The main spectral peak was at ~8–9 kHz in phrase 1 and ~7 kHz in phrase 2, and gradually decreased to ~6 kHz in phrase 9. This trend was approximately represented by the rate-CF profiles of type S units (Fig. 14*C*) throughout the time course of this call (~1 s long). However, when all recorded units were considered, only the first call phrase was clearly represented in the rate-CF profile (Fig. 14*B*). Rate-CF profiles based on type F⁻ units revealed no consistent representation of spectral features across call phrases (Fig. 14*D*).

Taken together, the data shown in Figs. 12–14 suggest that the critical information (i.e., spectral peaks and troughs) about the short-term spectrum of a complex natural vocalization is *1*) preserved by the type S or type F subpopulation and *2*) not represented by the type F⁻ subpopulation. Moreover, although correlation was found between rate-CF profiles and a stimulus spectrum, there are also some striking differences. Although spectral peaks and troughs were clearly reflected in rate-CF profiles, details of the spectrum did not seem to be represented. Peaks in rate-CF profiles were generally sharper and narrower in width than corresponding spectral peaks in all but type F⁻ subpopulations (Figs. 12–14).

Finally, what would happen to the spectral representation by distributed cortical responses if the same call phrase we analyzed in Fig. 12 was presented in the time-reversed version? This question is examined in Fig. 15. Note that the magnitude spectrum of the time-reversed call (Fig. 15*A*) is the same as that of the natural call, because the manipulation only changed the phase spectrum. The main spectral peak seems to be clearly marked in all rate-CF profiles, except

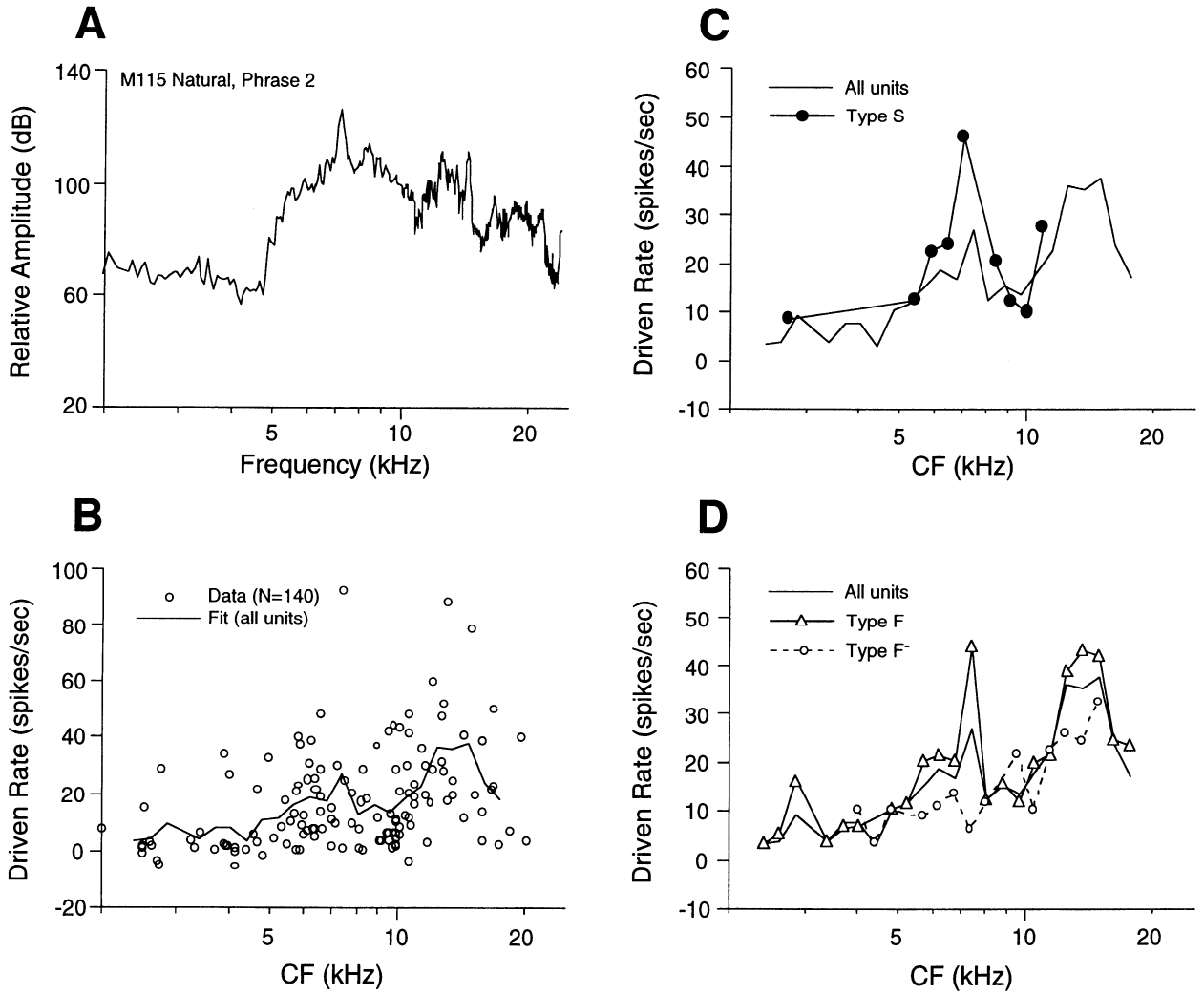

FIG. 13. Comparison of short-term call spectrum and rate-CF profiles for the 2nd call phrase of the same twitter call analyzed in Fig. 12. The display format is the same as in Fig. 12.

by those of type S units (Fig. 15, *C* and *D*). However, there is a distinct difference in A1's representation of the natural and time-reversed versions of this call phrase: the main spectral peak was equally reflected in rate-CF profiles of both type F and F⁻ units for the time-reversed call (Fig. 15*D*), whereas it was differentially represented when played in its natural form (Fig. 12*D*). This is consistent with our conclusion from temporal firing patterns (e.g., Fig. 10) that these two subpopulations, although symmetrically defined, do not represent the natural and time-reversed versions of a call in a symmetrical way. Furthermore, it should be pointed out that the rate-CF profiles in Fig. 15*D* are likely to be attenuated more when passing through the next stage of cortical processing because the discharges on which they are based were temporally far less synchronized across frequency channels when compared with those produced by the natural version of the same call phrase (see Fig. 10).

DISCUSSION

Importance of using natural stimuli in cortical studies

Electrophysiological studies of the auditory cortex of mammals and primates during the past few decades have been devoted largely to the neuronal responses of artificial sounds such as pure and modulated tones and noises, except in echolocating bats (see reviews by Aitkin 1990; Clarey et al. 1992; Phillips et al. 1991). Although progress has been made in understanding how the auditory cortex processes these simple sounds, it is still unclear how behaviorally important complex communication sounds that animals and humans hear in daily life, such as species-specific vocalizations and speech, are represented in the auditory cortex. The complexity of natural sounds and of the auditory pathways leading to the cortex makes it difficult to predict the cortical responses to complex natural sounds from responses to simpler, artificial stimuli. Furthermore, cortical plasticity implies that cortical responses to sounds that do not bear any behavioral relevance to the animal under study will differ from responses produced by behaviorally relevant sounds. As the studies by Suga and colleagues on the mustached bat have clearly demonstrated (Suga 1988, 1990), the selection of behaviorally relevant stimuli can be decisive in revealing the functional organization of the auditory cortex.

Simple stimuli provide opportunities to study basic cortical structure, such as spectral tuning characteristics (Heil et al. 1992b; Imig et al. 1977; Merzenich and Brugge 1973; Merzenich et al. 1975; Schreiner and Mendelson 1990) and

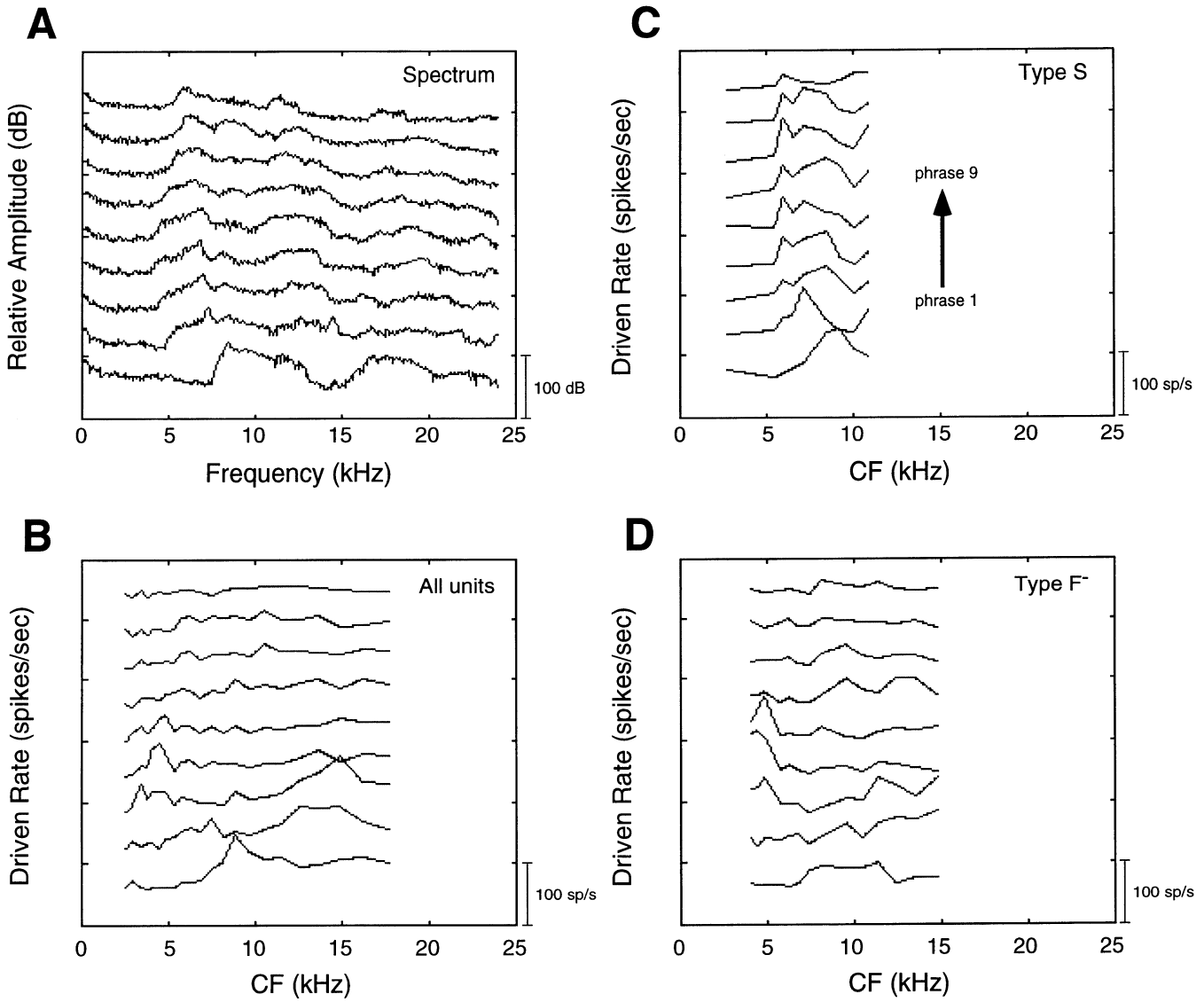

FIG. 14. Analysis of spectral representation of individual call phrases in a natural twitter call (*M115*) that has 9 phrases. In each panel, the 1st phrase is plotted as the *bottom line,* the last (9th) phrase as the *top line. A*: magnitude spectra of individual call phrases of the natural twitter call. One division on the ordinate represents 100 dB. Each successive spectrum is shifted vertically by 100 dB in the plot. *B–D*: rate-CF profiles for call phrases 1–9 computed from all units (*B, N* = 140), type S units (*C, N* = 37), and type F^- units (*D, N* = 38), respectively. Data were recorded from 1 marmoset in response to its own twitter call. One division on the ordinate represents 100 spikes/s. Each successive profile is shifted vertically by 100 spikes/s in the plots.

amplitude modulation (AM) or FM selectivity (Heil et al. 1992a; Mendelson et al. 1993; Schreiner and Urbas 1988; Shamma et al. 1993a). Properly designed complex stimuli such as vowellike ripple spectra (Calhoun and Schreiner 1993; Shamma 1993b) or the synthetic vocalizations used in this study could lead to a better understanding of the functional structure of the cortex beyond that revealed by the use of tonal stimuli. However, by our view, behavioral relevance of a stimulus and prior behavioral experience of an animal must be brought into consideration in cortical studies as a rule. Using acoustically altered vocalizations as comparisons, results of the present study provided evidence to support the notion that the behavioral experience plays an important role in shaping A1 responses to natural vocalizations. A direct verification of the notion needs to be conducted in future studies in which the behavioral relevance of a sound to the animal under study can be manipulated.

There are both conceptual and technical difficulties associated with the use of natural stimuli. Nevertheless, with the advance in digital signal processing techniques and in our improved understanding of the cortex in general, we are now in a much stronger position to attack the problem of cortical representations of communication sounds than were earlier investigators (Newman and Wollberg 1973b; Winter and Funkenstein 1973). The current study represents an effort to study the problem of neural encoding of complex sounds using a "top-down" approach, i.e., starting from natural sounds and their systematically and parametrically manipulated variations to less complex stimuli. Our long-range goal is to study not only how A1 responds to a complex natural sound as a whole, but also how it responds to various features of calls in the vocal repertoire of this species, and ultimately to relate these properties to other basic properties of A1 neurons revealed using simpler stimuli.

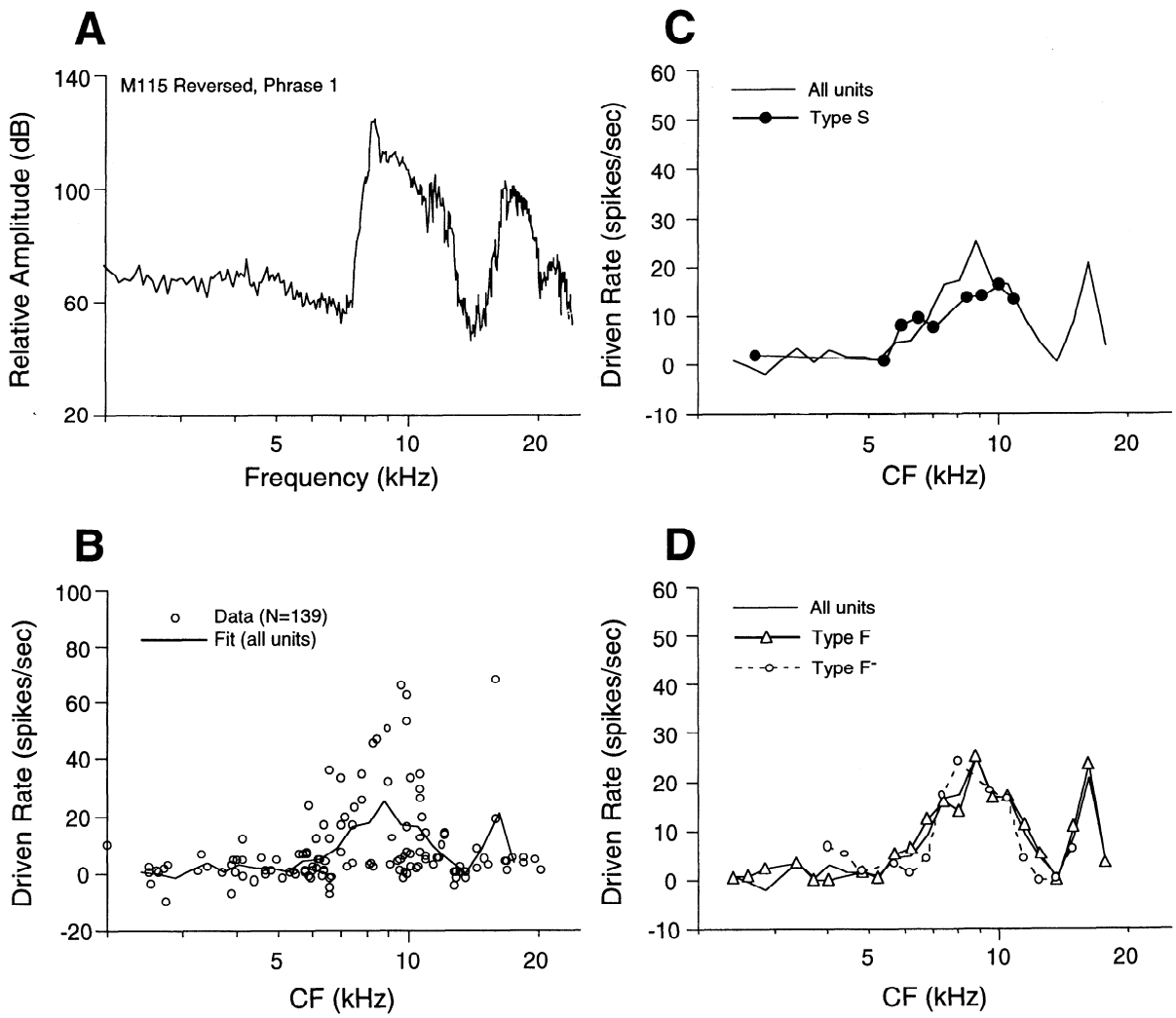

FIG. 15. Spectral representation of the 1st call phrase delivered in the time-reversed call is shown for the experiment described in Fig. 12. The same analytic method and display format were used. Note that the magnitude spectrum of the 1st call phrase in the time-reversed call (*A*) is the same as that in the natural call (Fig. 12*A*); only the phase spectrum is different.

Population versus single-cell representation

Studies of other primate species have shown that A1 is important for the perception of communication sounds (Heffner and Heffner 1986a,b). Assuming that this is also true for the common marmoset, the neuronal representations of its vocalizations in A1 must reflect aspects of the neural processing required for intraspecies communication in this species.

One of the main findings of the present study is a clear demonstration of a close correlation between the spectrotemporal acoustic pattern of a complex vocalization (in the form of spectrograms) and the spectrotemporal discharge patterns within distributed, engaged A1 populations (Figs. 3 and 4). In addition, we found that some A1 neurons responded selectively to more than one of the tested twitter calls, which were highly overlapping in their spectra. These calls were recorded from the monkey under study and from its conspecifics. Our findings demonstrate that their representation is by a large number of neurons. Even a selective subpopulation comprises neurons distributed widely across the engaged A1 sector. Representation is decidedly not by specialized call detector neurons at this cortical level. These results are in disagreement with the specialized neuron hypothesis firmly held by many investigators two decades ago, when researchers extensively studied the cortical representations of species-specific vocalizations in primates (Newman and Wollberg 1973a,b; Winter and Funkenstein 1973; Wollberg and Newman 1972). Results of earlier studies also led others to later suggest population coding as an alternative (Creutzfeldt et al. 1980; Pelleg-Toiba and Wollberg 1991). One advantage of a population coding scheme is that it is capable of encoding stimuli with different features that are highly overlapping in afferent channels on the basis of the same pool of neurons. A second advantage is that the combination of population coding and temporal integration process recorded in this study will be highly tolerant to fluctuations of noises in this stimulus encoding machinery.

Synchronization-based abstraction of complex vocalizations

An important observation of the present study is that the spectrotemporal discharge pattern of A1 neuronal popula-

tions evoked by this studied vocalization is not a replica of a call's acoustic pattern, but rather an abstraction of it (Fig. 3). This is in contradistinction to coding at the auditory periphery. At the level of the auditory nerve, time-varying spectral components like the FM sweeps in marmoset twitter calls produce firing patterns that follow such stimuli very faithfully (Carney and Geisler 1986; Delgutte and Kiang 1984c; Miller and Sachs 1983). That is because auditory nerve fibers are able to encode fine temporal details of complex stimuli with high fidelity, i.e., their responses can be phase-locked to the carrier and envelope of a complex stimulus (e.g., squirrel monkey: Rose et al. 1967; rat: Møller 1976; frog: Feng et al. 1991; cat: Joris and Yin 1992; Wang and Sachs 1993). At the level of A1, this stimulus-following ability is degraded, as also demonstrated by studies using tonal or AM stimuli (see review by Langner 1992). It has been shown that the vast majority of A1 neurons only follow modulation frequencies up to ~30–40 Hz in adult cats (de Ribeaupierre et al. 1972; Schreiner and Urbas 1988) and monkeys (Fastl et al. 1986; Müller-Preuss 1986). When a complex sound like a twitter call is presented, individual A1 neurons only follow the stimulus envelope but not fine temporal details, because they are highly synchronized at a fixed time relative to each call phrase. That is, when the stimulated population of units is considered, A1 neuronal responses are also synchronized to each other (Fig. 3). Thus distributed A1 neurons have captured a stimulus feature, like an FM sweep, by responding synchronously across CF.

The fact that modifications of the time course of the spectral components in a vocalization (e.g., time-compressed or expanded) produced changes in the output of A1 neurons indicates that A1 neurons integrate stimulus energy over significant periods of time before discharging. When the spectrotemporal characteristics of a stimulus varied within this interval, A1 responses varied accordingly. This integration interval, as judged by FM sweeps in the twitter calls, is on the order of 20–40 ms. One can also predict similar time constant from responses to AM sounds. Because thalamic neurons can follow AM and FM stimuli at much faster rates than cortical neurons (Creutzfeldt et al. 1980; Langner 1992; Schreiner and Langner 1988), much of this integration must take place in A1. The consequence of such firing patterns is that the spectrotemporal acoustic pattern of the vocalization is represented by the spectrotemporal discharge pattern of A1 neurons with reduced temporal complexity. Furthermore, data presented in Figs. 12–14 indicate that A1 represents the spectrum of a complex vocalization such as a natural twitter call with reduced spectral complexity. Thus the detailed representations of important features of complex sounds at the periphery appear to be converted into a more abstract, synchronization-based population code in cortical field A1. This population representation is presumably subject to further complex integration and segregation at higher cortical levels, and is likely to progress to still more abstract forms.

Interpretation of selective subpopulations

By using temporally modified vocal stimuli, the existence of subpopulations of A1 neurons that selectively responded to natural calls was documented in the present study. These selective subpopulations were defined solely on the basis of their temporal response properties. The fact that they were found to carry just as much or more information on the stimulus spectrum as that carried by all sampled units suggests that they are of functional importance in coding these complex sounds. We should point out that the definition of a "selective subpopulation" is criteria dependent. Thus the percentages of cortical neurons belonging to a particular selective population (e.g., type S or type F) are meaningful only in the context of how a selective subpopulation was defined. However, the important point is that there were preferential responses to a given natural call in some (a subpopulation) but not all responding A1 neurons. These observations suggest the existence of (putative) stimulus-specific neuronal assemblies that encode individual natural sounds. Such cell assemblies are presumably formed in learning and maintained dynamically throughout the life of a monkey. Furthermore, our observations also showed that a subpopulation of A1 neurons (type S or F) that is selective for one natural call *1*) shares some of its neuronal members with subpopulations selective for other natural calls and *2*) has neuronal members that are not selective for those calls. Thus each A1 neuron may participate in encoding more than one natural call.

It needs to be emphasized that natural vocalizations evoked responses not only in these selective subpopulations, but in other A1 neurons as well. However, these selective neuronal assemblies carry more information about a particular vocalization than do other nonselective A1 neurons. Therefore, when a behaviorally relevant vocalization is presented to a marmoset, it does not produce all-or-none responses in individual A1 neurons, but evokes responses in a subpopulation of A1 cells that are stronger in magnitude and more coherent in temporal relationship than responses to other stimuli with similar spectral structures.

The findings of the present study demonstrate that a natural vocalization is represented differently in A1 than other stimuli of equivalent complexity, as illustrated by responses to natural and time-reversed calls (Fig. 10). The differential responses could be due to two factors. First, the Hebbian-like plasticity mechanisms operating in A1, principally through the strengthening of positive interconnections in the cortical network (Ahissar et al. 1992; Merzenich and Sameshima 1993; Recanzone et al. 1992b), may give rise to stronger and more coherent responses to a natural call that stimulates an animal's auditory system in a behaviorally important way. Second, the difference in acoustic structure between a natural call and the synthetic ones may also contribute to the difference in their A1 responses. However, the fact that the spectral features of a natural call are represented by type F but not by type F^- units (Fig. 12*D*), whereas the spectral features of a time-reversed call are equally represented by both type F and type F^- units (Fig. 15*D*), indicates that their differential representations in A1 do not solely result from their acoustical differences. These two factors can not be completely sorted out on the basis of the results from the present study and need to be explored in future investigations.

Nature of complex sound representation in A1

Findings from the present study demonstrate that representation of a complex sound involves a large population of

A1 neurons. Our analysis of temporal and spectral characteristics of cortical responses to marmoset twitter calls suggests that the representation of a spectrally and temporally complex vocalization at A1 is *1*) temporally integrated, *2*) spectrally distributed, and *3*) spatially dispersed and synchronized. As a result, these complex acoustic signals are mapped into distinct neural patterns that form a base for discrimination and categorization by successive processing stages.

The results presented in this report demonstrate a correlation between A1 discharge patterns and the spectrotemporal features of a complex vocalization. They also raise an important question: how can spectrotemporally highly overlapping twitter calls of different monkeys be discriminated on the basis of spectrotemporal firing patterns that are even more overlapping as a result of the reduction in temporal and spectral complexity? Behavioral studies conducted in the pygmy marmosets showed that they are able to distinguish calls from individual members by their idiosyncratic features (Snowdon and Cleveland 1980). It is important to point out that cortical units that form the spectrotemporal firing pattern bearing important information on the stimulus are in fact spatially distributed across A1 along both tonotopic and isofrequency axes. We suggest that the discrimination of spectrotemporally highly overlapping vocalizations is based on these distributed spatial firing patterns that are manifested by the call-selective cell assemblies. Subpopulations selectively responding to different calls share some but not all of their neuronal members, and thus can form a basis of their discrimination to be read out by higher cortical stages. Therefore the A1 representation of different stimuli does not have to be composed of isolated cortical zones, each of which corresponds to an individual stimulus, but rather it can consist of a collection of distributed cortical patches that may partially overlap each other. This may represent a general strategy of the cortex to convert temporal information (within neuronal integration periods) into spatial information as a requirement for recognizing complex sensory input patterns. Quantitative analyses of these spatial pattern representation characteristics will be considered in subsequent reports.

The dramatic reduction in temporal resolution at A1 as compared with the periphery also suggests that cortical processing of complex sounds operates on a "segment-by-segment" basis rather than on a more or less "point-by-point" basis operating at the auditory nerve. With this slowdown of the incoming signal speed, some buffer time is available for complex integration and comparison at this level of the system.

Implication of A1 representational forms

An important feature of A1 representational forms for complex sounds illustrated by responses to marmoset vocalizations is the reduction in temporal complexity. In other words, the temporal resolution of the discharges from A1 units for a time-varying stimulus stream is limited by an integrative time constant. The representation of more rapidly varying stimulus components thus has to be in a spatial form. On the other hand, Phillips and Hall (1990) have shown that the distribution of the standard deviation of the mean first spike latency of A1 neurons is approximately equivalent to that of the auditory nerve fibers. This, together with results of the present study, suggests that A1 neurons precisely mark the occurrence timing of an individual stimulus event while coding other temporal detail in a spatial form. Such a representational form has implications for processing complex sounds at later stages. For a higher cortical area to optimally read A1 outputs, its integrative time constant should be equal to or greater than that of A1. Findings from studies using AM stimuli on different cortical areas support this hypothesis. Schreiner and Urbas (1988) showed that the mean best modulation frequency is much lower in several secondary cortical fields outside A1, an indication of longer integration time constant. These secondary fields were known to receive moderate to strong inputs from A1 (see review by Winer 1992).

The synchronization of discharges to a wideband stimulus provides a basis for the enhancement of stimulus-related firing patterns at higher processing stages. Synchronized discharges from spatially distributed neurons have a greater probability of passing through the next level of cortical processing when they are integrated. Because cortical fields projecting to and from A1 generally have wider spectral tuning (Aitkin 1990), spatial integration is likely to take place when A1 outputs are read into those fields. By such integration, noisy discharge backgrounds will be filtered and information-bearing discharges further enhanced. Because learning mechanisms in the cortex are coincident input based (Merzenich et al. 1993; Wang et al. 1994), synchronized A1 outputs can profoundly engage downstream cortical learning mechanisms.

Technical considerations

Because our data were obtained from barbiturate-anesthetized animals, one may question the extent to which the spatiotemporal discharge patterns and selective responses we observed reflect neural representations of these vocalizations in the cortex of an awake marmoset. The spectral aspect of the spatiotemporal discharge patterns depends on the underlying tonotopic organization in this cortical field. The evidence so far indicates that the orderly tonotopic organization found in A1 of anesthetized primates (macaque: Merzenich and Brugge 1973; owl monkey: Imig et al. 1977; common marmoset: Aitkin et al. 1986) also exists in A1 of awake mammals or primates (Brugge and Merzenich 1973; Pelleg-Toiba and Wollberg 1989). We believe that the spectral representation of the marmoset vocalization demonstrated in the present study should largely hold in the awake cortex. In the temporal domain, the frequencies at which responses of A1 neurons can follow AM stimuli were found to be similar in anesthetized (Schreiner and Urbas 1988) and unanesthetized (Creutzfeldt et al. 1980) animals, suggesting that the kind of temporal integration described here should take place in unanesthetized animals as well. In awake preparations, the spontaneous firing of A1 units is generally higher than that in anesthetized animals. Higher spontaneous activity should not fundamentally alter the spatiotemporal representation of a complex vocalization, because the stimulus-triggered discharges are highly synchronized across frequency channels (Fig. 4) whereas the spontaneous discharges are not. Therefore a coherent image of a complex sound can still emerge from an elevated "noisy" background. In addition, the R_s measure we used is not affected by adding random spikes. The major aspects of spectral and

temporal representations of marmoset vocalizations described here are not likely to be fundamentally affected by a reduction in response magnitude typically caused by barbiturate anesthesia (Erulkar et al. 1956; Goldstein et al. 1959; Tees and Kiang 1964). However, the selectivity curve for a natural call (Fig. 8*A*) might be sharper in unanesthetized preparations. It remains to be seen in future investigations what roles the behavioral state of an animal, e.g., arousal and attention, will play in defining and shaping A1 representations of species-specific vocalization.

The multiunit recordings used in this study, which enabled us to sample a large number of cortical sites over a limited period of time, should not effect the temporal aspect of the analysis because all spikes were recorded individually unless there was a considerable overlapping of spikes from different neurons. This was not very likely given the low spontaneous and driven discharge rates in these preparations. What could potentially be affected by the multiunit recording procedure is any analysis based on response magnitude such as driven discharge rate. This may explain relatively large scattering in data shown in Figs. 12, 13, and 15. However, because we sampled cortical units in a given experiment more or less in a random order, the recording procedure did not bias toward a particular frequency range and thus fluctuations in response magnitude caused by this method should be distributed fairly evenly among all studied units. We therefore believe that the average rate-CF profiles shown in Figs. 12–15 do reflect the spectral characteristics of the A1 representation.

Relevance to previous studies

A number of studies have investigated the responses of auditory cortical fields to species-specific vocalizations in primates (Glass and Wollberg 1979, 1983a,b; Manley and Müller-Preuss 1978; Newman and Wollberg 1973a,b; Pelleg-Toiba and Wollberg 1991; Steinschneider et al. 1982; Winter and Funkenstein 1973; Wollberg and Newman 1972). Data from previous studies showed that although a large portion of cortical neurons was responsive to species-specific vocalizations, few could be described as call detectors, i. e., neurons responsive only to specific calls in a primate's vocal repertoire (see Aitkin 1990; Clarey et al. 1992 for reviews). Findings of the present study are consistent with this view in that neurons in marmoset A1 did not respond to complex, behaviorally important vocalizations in an all-or-none fashion (Fig. 8). Furthermore, we quantitatively demonstrate that subpopulations of A1 neurons exhibited preferential responses to particular vocalizations (Figs. 8 and 9).

Time-reversed natural vocalizations were also used in previous investigations (Glass and Wollberg 1983a,b; Pelleg-Toiba and Wollberg 1991). In those studies, it was concluded that A1 responses to reversed vocalizations were not significantly different from the response to natural vocalizations, although only a very small percentage of neurons (~2%) was found to exhibit a "mirror image" in the temporal structure of their discharges to the reversed calls (Pelleg-Toiba and Wollberg 1991). This conclusion, however, was largely based on qualitative observations or spike counts over the period of an entire call. Through quantitative analysis, we found in the present study that cortical responses to natural and reversed vocalizations were significantly different if a stimulus-related, temporally based response measure such as R_s was used (Fig. 9).

Creutzfeldt et al. (1980) noticed in their study on A1 of guinea pigs that, in comparison with thalamic responses, cortical responses exhibited loss of information on the fine structure of modulation in the vocalizations tested. The present study more comprehensively documented the reduction in temporal resolution in the spectrotemporal discharge patterns of many cortical neurons across frequency channels evoked by a behaviorally important vocalization (Fig. 3). We interpret this characteristic as being the consequence of temporal integration of afferent inputs by cortical neurons, and having implications for temporal-spatial transformation in stimulus encoding at A1.

The property of A1 neurons of seeming to follow the envelope of a portion of a complex vocalization was also observed by Pelleg-Toiba and Wollberg (1991), who called the phenomenon "peak tracking." In this study we quantitatively analyze the relationship between responses of an A1 unit and the corresponding stimulus envelopes and are thus able to establish the quantitative correlation between the two (Fig. 6).

We thank E. Pirali and the Peabody Computer Music Department of Johns Hopkins University for providing the SoundView software used in stimulus synthesis, A. Wong for assistance in sound analysis, and Dr. R. L. Snyder for helpful comments on the manuscript.

This research was supported by a Kleberg Foundation Postdoctoral Fellowship to X. Wang, National Institute of Neurological Disorders and Stroke Grant NS-10414, The Coleman Fund, Office of Naval Research contract N00014–91-J-1317, and Hearing Research, Inc.

Present address and address for reprint requests: X. Wang, Dept. of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Ave., Ross 424, Baltimore, MD 21205.

Received 30 January 1995; accepted in final form 24 July 1995.

REFERENCES

AHISSAR, E., VAADIA, E., AHISSAR, M., BERGMAN, H., ARIELI, A., AND ABELES, M. Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. *Science Wash. DC* 257: 1412–1415, 1992.

AITKIN, L. *The Auditory Cortex.* London: Chapman & Hall, 1990.

AITKIN, L. M., KUDO, M., AND IRVINE, D. R. Connections of the primary auditory cortex in the common marmoset, *Callithrix jacchus jacchus. J. Comp. Neurol.* 269: 235–248, 1988.

AITKIN, L. M., MERZENICH, M. M., IRVINE, D. R., CLAREY, J. C., AND NELSON, J. E. Frequency representation in auditory cortex of the common marmoset (*Callithrix jacchus jacchus*). *J. Comp. Neurol.* 252: 175–185, 1986.

AITKIN, L. AND PARK, V. Audition and the auditory pathway of a vocal new world primate, the common marmoset. *Prog. Neurobiol.* 41: 345–367, 1993.

ALTMANN, S. A. The structure of primate communication. In: *Social Communication Among Primates,* edited by S. A. Altmann. Chicago, IL: Chicago Univ. Press, 1967, p. 325–362.

ANDREW, R. J. The origins and evolution of the calls and facial expressions of the primates. *Behaviour* 20: 1–109, 1963.

BLACKBURN, C. C. AND SACHS, M. B. The representations of the steady-state vowel sound /ϵ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *J. Neurophysiol.* 63: 1191–1212, 1990.

BRUGGE, J. F. AND MERZENICH, M. M. Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J. Neurophysiol.* 36: 1138–1158, 1973.

CALFORD, M. B. AND TWEEDALE, R. Immediate and chronic changes in responses of somatosensory cortex in adult flying-fox after digit amputation. *Nature Lond.* 332: 446–448, 1988.

CALHOUN, B. M. AND SCHREINER, C. E. Spatial frequency filters in cat auditory cortex. *Soc. Neurosci. Abstr.* 23: 1422, 1993.

CARNEY, L. H. AND GEISLER, C. D. A temporal analysis of auditory-nerve fiber responses to spoken stop consonant-vowel syllables. *J. Acoust. Soc. Am.* 79: 1896–1914, 1986.

CLAREY, J. C., BARONE, P., AND IMIG, T. J. Physiology of thalamus and cortex. In: *The Mammalian Auditory Pathway: Neurophysiology,* edited by A. N. Popper and R. R. Fay. New York: Springer-Verlag, 1992, p. 232–334.

CLARK, S. A., ALLARD, T., JENKINS, W. M., AND MERZENICH, M. M. Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs. *Nature Lond.* 332: 444–445, 1988.

CREUTZFELDT, O., HELLWEG, F.-C., AND SCHREINER, C. Thalamocortical transformation of responses to complex auditory stimuli. *Exp. Brain Res.* 39: 87–104, 1980.

DELGUTTE, B. AND KIANG, N. Y. S. Speech coding in the auditory nerve. I. Vowel-like sounds. *J. Acoust. Soc. Am.* 75: 866–878, 1984a.

DELGUTTE, B. AND KIANG, N. Y. S. Speech encoding in the auditory nerve. III. Voiceless fricative consonants. *J. Acoust. Soc. Am.* 75: 887–896, 1984b.

DELGUTTE, B. AND KIANG, N. Y. S. Speech coding in the auditory nerve. IV. Sounds with consonant-like dynamic characteristics. *J. Acoust. Soc. Am.* 75: 897–907, 1984c.

DI LORENZO, P. M. Across unit patterns in the neural response to taste: vector space analysis. *J. Neurophysiol.* 62: 823–833, 1989.

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

EPPLE, G. Comparative studies on vocalization in marmoset monkeys (*Hapalidae*). *Folia Primatol.* 8: 1–40, 1968.

ERULKAR, S. D., ROSE, J. E., AND DAVIES, P. W. Single unit activity in the auditory cortex of the cat. *Bull. Johns Hopkins Hosp.* 99: 55–86, 1956.

FASTL, H., HESSE, A., SCHORER, E., URBAS, J., AND MÜLLER-PREUSS, P. Search for neural correlates of the hearing sensation fluctuation strength in the auditory cortex of squirrel monkeys. *Hear. Res.* 23: 199–203, 1986.

FENG, A. S., HALL, J. C., AND SIDDIQUE, S. Coding of temporal parameters of complex sounds by frog auditory nerve fibers. *J. Neurophysiol.* 65: 424–445, 1991.

FLANAGAN, J. L. AND GOLDEN, R. M. Phase vocoder. *Bell Syst. Tech. J.* 45: 1493–1509, 1966.

FOSSEY, D. Vocalizations of the mountain gorilla (*Gorilla gorilla beringei*). *Anim. Behav.* 20: 36–53, 1972.

GAUTIER, J.-P. AND GAUTIER, A. Communication in old world monkeys. In: *How Animals Communicate,* edited by T. A. Sebeok. Bloomington, IN: Indiana Univ. Press, 1977, p. 809–964.

GEORGOPOULOS, A., LURITO, J., PETRIDES, M., SCHWARTZ, A., AND MASSEY, J. Mental rotation of the neuronal population vector. *Science Wash. DC* 243: 234–236, 1989.

GILBERT, C. D. AND WIESEL, T. N. Receptive field dynamics in adult primary visual cortex. *Nature Lond.* 356: 150–152, 1992.

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.

GOCHIN, P. M., COLOMBO, M., DORFMAN, G. A., GERSTEIN, G. L., AND GROSS, C. G. Neural ensemble coding in inferior temporal cortex. *J. Neurophysiol.* 71: 2325–2337, 1994.

GOLDBERG, J. M. AND BROWN, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* 32: 613–636, 1969.

GOLDSTEIN, M. H. J., KIANG, N. Y.-S., AND BROWN, R. M. Responses of the auditory cortex to repetitive acoustic stimuli. *J. Acoust. Soc. Am.* 31: 356–364, 1959.

GREEN, S. Variations of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): a field study. In: *Primate Behavior: Developments in Field and Laboratory Research,* edited by L. A. Rosenblum. New York: Academic, 1975, p. 1–102.

HEFFNER, H. E. AND HEFFNER, R. S. Hearing loss in Japanese macaques following bilateral auditory cortex lesions. *J. Neurophysiol.* 55: 256–271, 1986a.

HEFFNER, H. E. AND HEFFNER, R. S. Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *J. Neurophysiol.* 56: 683–701, 1986b.

HEIL, P., RAJAN, R., AND IRVINE, D. R. F. 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, 1992a.

HEIL, P., RAJAN, R., AND IRVINE, D. R. F. Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. II. Organization of response properties along the "isofrequency" dimension. *Hear. Res.* 63: 135–156, 1992b.

IMIG, T. J., RUGGERO, M. A., KITZES, L. M., JAVEL, E., AND BRUGGE, J. F. Organization of auditory cortex in the owl monkey (*Aotus trivirgatus*). *J. Comp. Neurol.* 177: 111–128, 1977.

IRVINE, D. R. F. Physiology of the auditory brainstem. In: *The Mammalian Auditory Pathway: Neurophysiology,* edited by A. N. Popper and R. R. Fay. New York: Springer-Verlag, 1992, p. 153–231.

JENKINS, W. M., MERZENICH, M. M., OCHS, M. T., ALLARD, T., AND GUÍC-ROBLES, E. Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *J. Neurophysiol.* 63: 82–104, 1990.

JORIS, P. X. AND YIN, T. C. T. Responses to amplitude-modulated tones in the auditory nerve of the cat. *J. Acoust. Soc. Am.* 91: 215–232, 1992.

KAAS, J. H., KRUBITZER, L. A., CHINO, Y. M., LANGSTON, A. L., POLLEY, E. H., AND BLAIR, N. Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science Wash. DC* 248: 229–231, 1990.

KIM, D. O. AND MOLNAR, C. E. A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. *J. Neurophysiol.* 42: 16–30, 1979.

LANGNER, G. Periodicity coding in the auditory system. *Hear. Res.* 60: 115–142, 1992.

MANLEY, J. A. AND MÜLLER-PREUSS, P. Response variability of auditory cortex cells in the squirrel monkey to constant acoustic stimuli. *Exp. Brain Res.* 32: 171–180, 1978.

MENDELSON, J., SCHREINER, C., SUTTER, M., AND GRASSE, K. Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. *Exp. Brain Res.* 94: 65–87, 1993.

MERZENICH, M. M., ALLARD, T., AND JENKINS, W. M. Neural ontogeny of higher brain function: implications of some recent neurological findings. In: *Information Processing in the Somatosensory System,* edited by O. Franzén and P. Westman. London: Macmillan, 1990a, p. 293–311.

MERZENICH, M. M. AND BRUGGE, J. F. Representation of cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res.* 50: 275–296, 1973.

MERZENICH, M. M., KAAS, J. H., WALL, J. T., NELSON, R. J., SUR, M., AND FELLEMAN, D. Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience* 8: 33–55, 1983.

MERZENICH, M. M., KNIGHT, P. L., AND ROTH, G. Representation of cochlea within primary auditory cortex in the cat. *J. Neurophysiol.* 38: 231–249, 1975.

MERZENICH, M. M., NELSON, R. J., STRYKER, M. P., CYNADER, M. S., SCHOPPMANN, A., AND ZOOK, J. M. Somatosensory cortical map changes following digital amputation in adult monkeys. *J. Comp. Neurol.* 224: 591–605, 1984.

MERZENICH, M. M., RECANZONE, G. H., AND JENKINS, W. M. How the brain functionally rewires itself. In: *Natural and Artificial Parallel Computations,* edited by M. Arbib and J. A. Robinson. New York: MIT Press, 1990b, p. 177–210.

MERZENICH, M. M. AND SAMESHIMA, K. Cortical plasticity and memory. *Curr. Opin. Neurobiol.* 3: 187–196, 1993.

MERZENICH, M. M., SCHREINER, C., JENKINS, W., AND WANG, X. Neural mechanisms underlying temporal integration, segmentation, and input sequence representation: some implications for the origin of learning disabilities. *Ann. NY Acad. Sci.* 682: 1–22, 1993.

MILLER, M. I. AND SACHS, M. B. Representation of stop consonants in the discharge patterns of auditory-nerve fibers. *J. Acoust. Soc. Am.* 74: 502–517, 1983.

MØLLER, A. A. Dynamic properties of primary auditory fibers compared with cells in the cochlear nucleus. *Acta Physiol. Scand.* 98: 157–167, 1976.

MÜLLER-PREUSS, P. On the mechanisms of call coding through auditory neurons in the squirrel monkeys. *Eur. Arch. Psychiatry Neurol. Sci.* 236: 50–55, 1986.

NELKEN, I., PRUT, Y., VAADIA, E., AND ABELES, M. Population responses to

multifrequency sounds in the cat auditory cortex: one- and two-parameter families of sounds. *Hear. Res.* 72: 206–222, 1994.

NEWMAN, J. D. Primate hearing mechanisms. In: *Comparative Primate Biology,* edited by H. D. Steklis and J. Erwin. New York: Liss, 1988, p. 469–499.

NEWMAN, J. D. 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, J. D. AND WOLLBERG, Z. Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res.* 54: 287–304, 1973b.

OPPENHEIM, A. V. AND SCHAFER, R. W. *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

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.

PETERSEN, M. R. The perception of species-specific vocalizations by primates: a conceptual framework. In: *Primate Communication,* edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen. Cambridge, UK: Cambridge Univ. Press, 1982, p. 171–211.

PHILLIPS, D. P. AND HALL, S. E. Response timing constraints on the cortical representation of sound time structure. *J. Acoust. Soc. Am.* 88: 1403–1411, 1990.

PHILLIPS, D. P., ORMAN, S. S., MUSICANT, A. D., AND WILSON, G. F. Neurons in the cat's primary auditory cortex distinguished by their responses to tones and wide-spectrum noise. *Hear. Res.* 18: 73–86, 1985.

PHILLIPS, D. P., REALE, R. A., AND BRUGGE, J. F. Stimulus processing in the auditory cortex. In: *Neurobiology of Hearing: The Central Auditory System,* edited by R. A. Altschuler, R. P. Bobbin, B. M. Clopton, and D. W. Hoffman. New York: Raven, 1991, p. 335–366.

PONS, T. P., GARRAGHTY, P. E., PMMAYA, A. K., KAAS, J. H., TAUB, E., AND MISHKIN, M. Massive cortical reorganization after sensory deafferentation in adult macaques. *Science Wash. DC* 252: 1857–1860, 1991.

RAJAN, R., IRVINE, D. R. F., WISE, L. Z., AND HEIL, P. Effect of unilateral partial cochlear lesions in adult cats on the representation of lesioned and unlesioned cochleas in primary auditory cortex. *J. Comp. Neurol.* 338: 17–49, 1993.

RECANZONE, G. H., MERZENICH, M. M., JENKINS, W. M., GRAJSKI, K. A., AND DINSE, H. R. Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained in a frequency-discrimination task. *J. Neurophysiol.* 67: 1031–1056, 1992a.

RECANZONE, G. H., SCHREINER, C. E., AND MERZENICH, M. M. Plasticity in the primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13: 87–104, 1992b.

DE RIBEAUPIERRE, F., GOLDSTEIN, M. H., JR., AND YENI-KOMISHIAN, G. H. Cortical coding of repetitive acoustic pulses. *Brain Res.* 205–225, 1972.

ROBERTSON, D. AND IRVINE, D. R. Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *J. Comp. Neurol.* 282: 456–471, 1989.

ROSE, J. E., BRUGGE, J. F., ANDERSON, D. J., AND HIND, J. E. Phase-locked response to low frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* 30: 269–293, 1967.

SACHS, M. B. AND YOUNG, E. D. Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J. Acoust. Soc. Am.* 66: 470–479, 1979.

SCHREINER, C., MENDELSON, J., AND SUTTER, M. Functional topography of cat primary auditory cortex: representation of tone intensity. *Exp. Brain Res.* 92: 105–122, 1992.

SCHREINER, C. E. AND LANGNER, G. Coding of temporal patterns in the central auditory system. In: *Auditory Function,* edited by G. M. Edelmann, W. E. Gall, and W. M. Cowan. New York: Wiley, 1988, p. 337–361.

SCHREINER, C. E. AND MENDELSON, J. R. Functional topography of cat primary auditory cortex: distribution of integrated excitation. *J. Neurophysiol.* 64: 1442–1459, 1990.

SCHREINER, C. E. AND URBAS, J. V. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear. Res.* 32: 49–63, 1988.

SEIDEN, H. R. *Auditory Acuity of the Marmoset Monkeys (Hapale jacchus)* (PhD thesis). Princeton, NJ: Princeton Univ, 1957.

SEYFARTH, R. M., CHENEY, D. L., AND MARLER, P. Monkey response to three different alarm calls: evidence of predator classification and semantic communication. *Science Wash. DC* 210: 801–803, 1980.

SHAMMA, S. A., FLESHMAN, J. W., WISER, P. R., AND VERSNEL, H. Organization of response areas in ferret primary auditory cortex. *J. Neurophysiol.* 69: 367–383, 1993a.

SHAMMA, S. A., VERSNEL, H., AND KOWALSKI, N. A. Organization of primary auditory cortex evident in responses to rippled complex sounds stimuli. *Soc. Neurosci. Abstr.* 23: 1422, 1993b.

SINEX, D. G. AND GEISLER, C. D. Responses of auditory-nerve fibers to consonant-vowel syllables. *J. Acoust. Soc. Am.* 73: 602–615, 1983.

SMITH, H. J., NEWMAN, J. D., AND SYMMES, D. Vocal concomitants of affiliative behavior in squirrel monkeys. In: *Primate Communication,* edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen. Cambridge, UK: Cambridge Univ. Press, 1982, p. 30–49.

SNOWDON, C. T. Linguistic and psycholinguistic approaches to primate communication. In: *Primate Communication,* edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen. Cambridge, UK: Cambridge Univ. Press, 1982, p. 212–238.

SNOWDON, C. T. AND CLEVELAND, J. Individual recognition of contact calls by pygmy marmosets. *Anim. Behav.* 28: 717–727, 1980.

STEINSCHNEIDER, M., AREZZO, J., AND VAUGHAN, H. G. J. Speech evoked activity in the auditory radiations and cortex of the awake monkeys. *Brain Res.* 252: 353–365, 1982.

STRUHSAKER, T. T. Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In: *Social Communication Among Primates,* edited by S. A. Altmann. Chicago, IL: Chicago Univ. Press, 1967, p. 238–324.

SUGA, N. Auditory neuroethology and speech processing: complex sound processing by combination-sensitive neurons. In: *Functions of the Auditory System,* edited by G. M. Edelman, W. E. Gall, and W. M. Cowan. New York: Wiley, 1988, p. 679–720.

SUGA, N. Cortical computational maps for auditory imaging. *Neural Networks* 3: 3–21, 1990.

TEES, D. C. AND KIANG, N. Y.-S. Evoked responses from the auditory cortex. *Exp. Neurol.* 10: 91–119, 1964.

VAN LAWICK-GOODALL, J. A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. In: *Primates: Studies in Adaptation and Variability,* edited by P. C. Jay. New York: Holt, Rinehart, & Winston, 1968, p. 313–374.

WALL, J. T., FELLEMAN, D. J., AND KAAS, J. H. Recovery of normal topography in the somatosensory cortex of monkeys after nerve crush and regeneration. *Science Wash. DC* 221: 771–773, 1983.

WANG, X., MERZENICH, M. M., SAMESIMA, K., AND JENKINS, W. M. Afferent input integration and segregation in learning are input timing dependent. *Soc. Neurosci. Abstr.* 20: 1427, 1994.

WANG, X. AND SACHS, M. B. Neural encoding of single-formant stimuli in the cat. I. Responses of auditory-nerve fibers. *J. Neurophysiol.* 70: 1054–1075, 1993.

WEINBERGER, N. M. AND DIAMOND, D. M. Physiological plasticity in auditory cortex: rapid induction by learning. *Prog. Neurobiol.* 29: 1–55, 1987.

WINTER, P. AND FUNKENSTEIN, H. H. The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkeys (*Saimiri sciureus*). *Exp. Brain Res.* 18: 489–504, 1973.

WINTER, P., PLOOG, D., AND LATTA, J. Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Exp. Brain Res.* 1: 359–384, 1966.

WINER, J. A. The functional architecture of the medial geniculate body and the primary auditory cortex. In: *The Mammalian Auditory Pathway: Neuroanatomy,* edited by A. N. Popper and R. R. Fay. New York: Springer-Verlag, 1992, p. 222–409.

WOLLBERG, Z. AND NEWMAN, J. D. Auditory cortex of squirrel monkey: response patterns of single cells to species specific vocalizations. *Science Wash. DC* 175: 212–214, 1972.

XERRI, C., STERN, J. M., AND MERZENICH, M. M. Alterations of the cortical representation of the rat ventrum induced by nursing behavior. *J. Neurosci.* 14: 1710–1721, 1994.

YOUNG, E. D. AND SACHS, M. B. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J. Acoust. Soc. Am.* 66: 1381–1403, 1979.