

On cortical coding of vocal communication sounds in primates

Xiaoqin Wang*

Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Ross 424, Baltimore, MD 21205

Understanding how the brain processes vocal communication sounds is one of the most challenging problems in neuroscience. Our understanding of how the cortex accomplishes this unique task should greatly facilitate our understanding of cortical mechanisms in general. Perception of species-specific communication sounds is an important aspect of the auditory behavior of many animal species and is crucial for their social interactions, reproductive success, and survival. The principles of neural representations of these behaviorally important sounds in the cerebral cortex have direct implications for the neural mechanisms underlying human speech perception. Our progress in this area has been relatively slow, compared with our understanding of other auditory functions such as echolocation and sound localization. This article discusses previous and current studies in this field, with emphasis on nonhuman primates, and proposes a conceptual platform to further our exploration of this frontier. It is argued that the prerequisite condition for understanding cortical mechanisms underlying communication sound perception and production is an appropriate animal model. Three issues are central to this work: (i) neural encoding of statistical structure of communication sounds, (ii) the role of behavioral relevance in shaping cortical representations, and (iii) sensory–motor interactions between vocal production and perception systems.

Communication sounds are a subset of acoustic signals vocalized by a species and used in intraspecies interactions. Human speech and species-specific vocalizations of nonhuman primates are two examples of communication sounds. Vocal repertoires of many animal species also include sounds that are not communicative in nature but are essential for the behavior of a species. For example, echolocating bats emit sonar signals that are used to determine target properties of prey (e.g., distance, velocity, etc.) but are not used in social interactions between bats. Many avian species such as songbirds have rich vocal repertoires. Communication sounds of nonhuman primates are a class of acoustic signals of special interest to us. Compared with other nonhuman species, primates share the most similarities with humans in the anatomical structures of their central nervous systems, including the cerebral cortex. Neural mechanisms operating in the cortex of primates thus have direct implications for those operating in the human brain.

Although field studies provide full access to the natural behavior of primates, it is difficult to combine them with physiological studies at the single neuron level in the same animals. The challenge is to develop appropriate primate models for laboratory studies where both vocal behavior and underlying physiological structures and mechanisms can be systematically investigated. This is a prerequisite if we ever want to understand how the brain processes vocal communication sounds at the level of single neurons. Most primates have a well-developed and sophisticated vocal repertoire in their natural habitats. However, for many larger primate species like macaque monkeys, their

vocal activities diminish under the captive conditions commonly found in research institutions, in part because of the lack of a behaviorally suitable housing environment. Fortunately, some primate species such as New World monkeys (e.g., marmosets, squirrel monkeys) remain highly vocal in properly configured captive conditions. Fig. 1A shows a vocal exchange between a male and a female marmoset recorded from a captive marmoset colony. These primate species may serve as excellent models for us to study in detail their vocal behavior as well as underlying neural mechanisms in the brain.

The cerebral cortex is known to play an important role in processing species-specific vocalizations. Studies have shown that lesions of auditory cortex cause a deficit in speech comprehension in humans and discrimination of vocalizations in primates (1, 2). Anatomically, humans and primates have similar cytoarchitecture in the superior temporal gyrus where the sensory auditory cortex is located (3). It has been shown in both Old World and New World primates that the sensory auditory cortex consists of a primary auditory field (A1) and surrounding secondary fields (4–6). Afferent information on acoustic signals is processed first in A1 and then in the secondary fields. From there, the information flows to the frontal cortex and to other parts of the cerebral cortex. The issues discussed in this article concern coding of vocalizations in the superior temporal gyrus, the sensory auditory cortex.

Issues in Understanding Cortical Processing of Communication Sounds

Vocal communication sounds are a special class of signals that are characterized by their acoustical complexity, their biological importance, and the fact that they are produced and perceived by the same species. As a result, the representations of these signals in the brain are likely to be different from representations of other types of acoustic signals (e.g., behaviorally irrelevant sounds, sounds from prey and predators). The nature of vocal communication sounds requires that we consider the following issues in our studies.

Neural Encoding of Statistical Structures of Communication Sounds.

There are at least three important behavioral tasks for which primates rely on their vocalizations in a natural environment. The tasks are to (i) identify messages conveyed by members of a social group or family, (ii) identify the caller of a vocalization, and (iii) determine the spatial location of a caller. There has been ample evidence that primates use their vocalizations in these behavioral tasks (7). These tasks require the auditory system of primates to solve “what”, “who,” and “where” problems on the basis of vocalizations. Before one can understand how the

This paper was presented at the National Academy of Sciences colloquium “Auditory Neuroscience: Development, Transduction, and Integration,” held May 19–21, 2000, at the Arnold and Mabel Beckman Center in Irvine, CA.

Abbreviation: CF, characteristic frequency.

*E-mail: xwang@bme.jhu.edu.

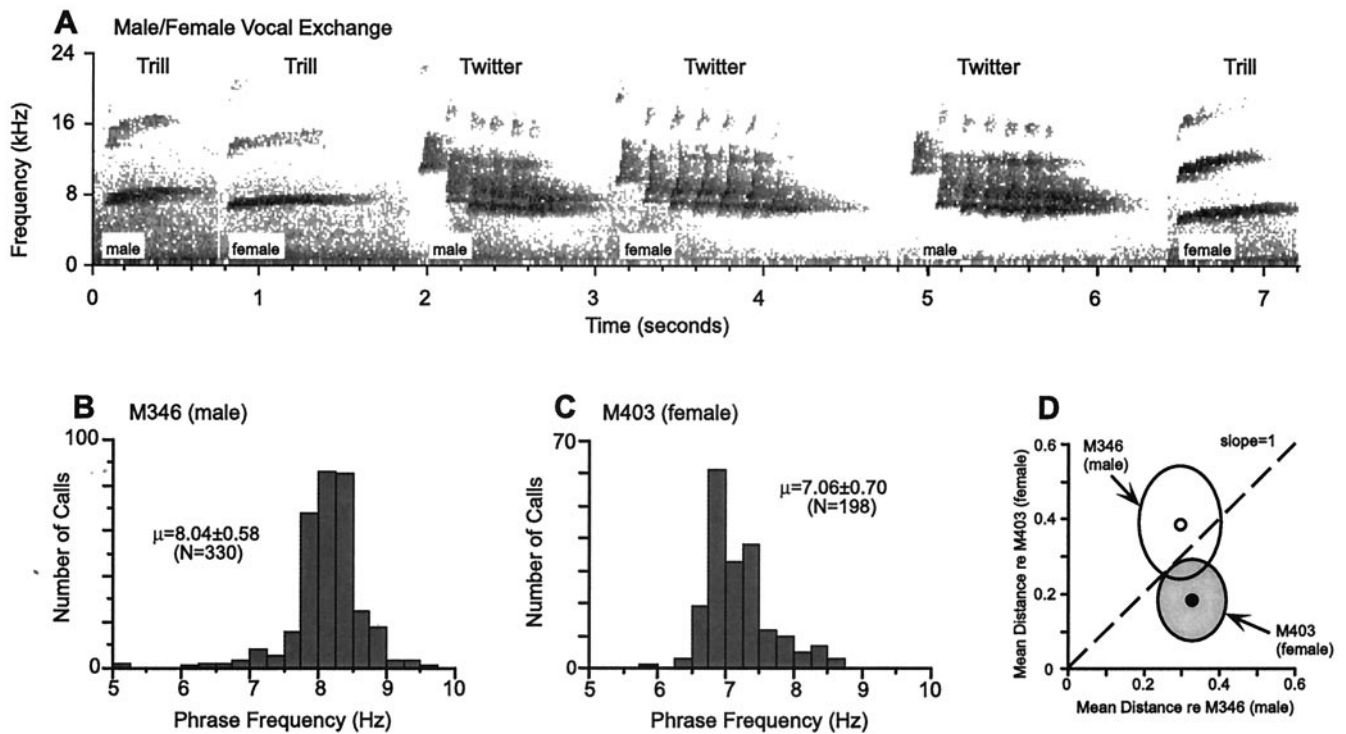


Fig. 1. (A) A real-time recording of vocal exchanges between a pair of marmosets (one male, one female), shown in the form of a spectrogram. This typical vocal exchange contains trill and twitter calls. Call type and caller identity of each call are indicated. (B and C) Distribution of *phrase frequency* of twitter calls from a male (B, M346) and female (C, M403) marmoset. Both marmosets lived in the same colony and engaged in frequent vocal exchanges such as the example shown in A. A twitter call contains several discrete upward FM sweeps, each of which is referred to as a phrase. The intervals between phrases are relatively constant in each twitter call. Fourier analysis of a twitter call's envelope, obtained by using the Hilbert transform, revealed a local maximum reflecting the repetition frequency of the phrases. The frequency at this maximum is defined as the phrase frequency (27). (D) The results of a multidimensional clustering analysis of twitter calls of the pair of marmoset monkeys in B and C. Four parameters were calculated for each twitter call sample and used in the analysis. These parameters included (i) number of phrases, (ii) phrase frequency, (iii) spectral-peak frequency of the first phrase, and (iv) spectral-peak frequency of the second phrase. The spectral-peak frequency is computed from the magnitude spectrum of each phrase (27). Two ellipsoids are drawn by using the mean and standard deviation of a distance measure d_{ij} (see definition below). The open circle marks the mean distances of calls made by the male monkey (M346) with respect to its own group mean (abscissa) and the female monkey's group mean (ordinate); the open ellipsoid (male monkey, M346) outlines the standard deviations along both axes. The filled circle and shaded ellipsoid are calculations for calls from the female monkey M403.

$$d_{ij}(n) = \sqrt{\sum_{k=1}^4 (P_{ik}(n) - \bar{m}_{jk})^2} \quad n = 1, 2, \dots, N_i; \quad \bar{m}_{jk} = \frac{\sum_{n=1}^{N_j} P_{jk}(n)}{N_j}$$

where i, j are animal designation (1: M346, 2: M403); N_i is number of call samples from i th animal ($N_1 = 330, N_2 = 198$); $P_{ik}(n)$ is k th parameter in the n th call of the i th animal; and \bar{m}_{jk} is mean value of the k th parameter of call samples from i th animal.

auditory system solves these problems, however, one must understand how the information is coded in the acoustic structures of vocalizations. After all, in the absence of visual and other kinds of sensory information (a realistic situation encountered by monkeys living in jungles with heavy vegetation), vocalization is the only messenger. Because spatial location of a sound is computed by the auditory system by using binaural cues as well as spectral cues provided by the external ear, the fundamental information carried by vocalizations is for “what” and “who” problems. Primates are known to produce distinct call types that presumably encode “what” information. It is also known that each animal has its own idiosyncratic features in its vocalizations, which are likely cues to represent caller identity.

However, just like human speech sounds, an individual utterance by a primate does not fully represent the underlying structure of a particular call type or caller because of the jittering of the vocal production apparatus and noises in the environment. There is some degree of randomness reflected by the nature of all sound production systems. The underlying statistical properties of these vocalizations, however, ought to be invariant from

call to call. Thus, the corresponding cortical coding should be and can only be fully understood at this level. In other words, it is not enough to know how cortical neurons respond to a particular call sample. One must know how the cortex extracts the invariant statistical structure from which all utterances of a call type or caller are generated.

The Role of Behavioral Relevance in Shaping Cortical Representations.

As biologically important signals, species-specific vocalizations have an ensured representation in the brain. To understand how these types of sound are represented in the cortex and how one can generalize principles from these representational schemes, one needs to understand the biological basis that shapes the cortical responses. There are three factors that may bias cortical responses to communication sounds in a particular species. These factors are (i) evolutionary predisposition, (ii) developmental plasticity, and (iii) experience-dependent plasticity in adulthood, each of which differs in its time scale. In a primate species whose vocal spectrum occupies much higher frequency range than that of humans, its auditory cortex devotes a larger

portion to that frequency range than does human auditory cortex. For example, human speech is concentrated on 0.5–3 kHz (the range of formant frequency), whereas vocalizations of marmoset monkeys are centered at 7–15 kHz (8). As a result, marmosets have an expanded representation of 7–15 kHz in their auditory cortex, whereas representation of frequencies below 3 kHz is much more limited (9). Such species-specific differences are formed through evolution over many millions of years and likely have a genetic underpinning. On a shorter time scale, changes throughout the developmental period, both prenatal and postnatal, may influence how the cortex processes vocalizations, given what is known of developmental plasticity of the cerebral cortex from studies of other sensory cortices. Finally, because the cortex is known to be subject to experience-dependent plasticity in adulthood (10, 11), what an animal hears on a daily basis must shape its cortical representation. The time scale for such changes is the shortest, probably in terms of months, weeks, or even days.

What do these considerations tell us in studying vocal communication sounds? For one, it is clear that one has to study directly mechanisms for encoding vocalizations within a species. However, because of developmental and experience-dependent plasticity, one may not fully reveal cortical coding mechanisms when using calls from unrelated conspecifics to which the experimental subjects have never been exposed. A simple but useful analogy is that one cannot study how the cortex codes Chinese in a native English speaker who never learned Chinese. Nor can one study cortical coding of Chinese in one of Chinese descent who was never exposed to Chinese. Although it is not yet clear to what extent these analogies are true in primates, they are powerful reminders that one must pay close attention to the behavioral meaning of sounds in any model systems of primates.

Sensory–Motor Interactions Between Vocal Production and Perception Systems. One important distinction between vocal communication sounds and other biologically important but nonvocal signals (such as sounds by prey and predators) is that the former are produced by the species perceiving them. It has long been known in studies of human speech that our perception of speech is biased by the nature of our ability to produce speech (12). Recent imaging studies showed that cortical areas outside the sensory auditory cortex on the superior temporal gyrus are activated during passive listening experience (13, 14). These findings are consistent with observations from studies in human epileptics undergoing neurological treatment with subdural or depth electrodes, through which electrical stimulation of the frontal cortex interrupted a patient’s ability in listening comprehension tasks (1, 15). This evidence suggests that processing of vocal communication sounds involves both sensory and motor systems. The frontal cortex likely plays an important role in these sensory–vocal interactions. It has been shown that anatomically the frontal cortex is reciprocally connected with auditory sensory cortex on the superior temporal gyrus in both Old World and New World primates (5, 16, 17). Specifically, the connections originate from the secondary auditory fields where strong neural responses to vocalizations were found in primates [ref. 18; X.W. (1999) *Association of Research in Otolaryngology Abs.* 22, 173]. It is likely that feedback from these higher-order processing centers outside the superior temporal gyrus influences cortical coding in the sensory auditory cortex.

Other important issues in cortical coding include whether vocalizations are represented by neural ensembles or by specialized cells and how representation of vocalizations is transformed from the primary auditory cortex to the secondary areas, which will be discussed below.

Progress and Current Work

Previous Studies. Over the past several decades, a number of experimental attempts have been made to elucidate the mechanisms of cortical coding of species-specific vocalizations in the auditory cortex of primates. The effort to understand cortical representation of vocalizations reached a peak in the 1970s, when a number of reports were published. The results of these studies were mixed, with no clear or consistent picture emerging as to how communication sounds are represented in the auditory cortex of primates (19). This lack of success may be accounted for in part, retrospectively, by expectations of the form of cortical coding of behaviorally important stimuli. For a time, it was thought that primate vocalizations were encoded by highly specialized neurons, the so-called “call detectors” (20, 21). However, individual neurons in the auditory cortex were often found to respond to more than one call or to various features of calls (20–23). The initial estimate of the percentage of the “call detectors” was relatively high. Later, much smaller numbers were reported as more calls were tested and more neurons were studied. At the end of this series of explorations, it appeared, at least in the initial stages of auditory cortical pathway, that the notion of highly specialized neurons is doubtful. Perhaps because of these seemingly disappointing findings, no systematic studies on this subject were reported for more than a decade afterward.

One shortcoming in the earlier studies was that responses to vocalizations were not adequately related to basic functional properties of a neuron such as its receptive field, temporal dynamics, etc. In addition, responses to vocalizations were not interpreted in the context of the overall organization of a cortical field or overall structure of the entire auditory cortex. Such information became available only in later years, and much of the structure of the auditory cortex is still being defined in primates. It is clear now that until a good understanding of the properties of a neuron is achieved, besides its responses to vocalizations, cortical coding of species-specific vocalizations will not be fully understood. Another lesson learned from the earlier studies is that one must consider the underlying statistical structure of a species’ vocalizations. In earlier studies, only vocalization tokens were used. This made it difficult, if not impossible, to accurately interpret cortical responses, as argued earlier in this article. One has to notice that in those earlier days, before powerful computers and digital technology became available, quantifying primate vocalizations would have been a formidable task. Nonetheless, these earlier explorations of the primate auditory cortex by using species-specific vocalization served as an important stepping stone for current and future studies of this important problem in auditory neurophysiology; they pointed to the right direction for seeking the correct answers.

Quantitative Characterizations of Communication Sound Repertoire.

After a long period of silence, studies in this field became active again in recent years. On the forefront of understanding acoustic structure of communication sounds, my laboratory has systematically studied the vocal repertoire of communication sounds in a highly vocal primate species, the common marmoset (*Callithrix jacchus jacchus*). In this study, we quantitatively characterized statistical properties not only of various call types but also of acoustic features related to individual identity. Fig. 1 *B–D* illustrates the ideas of this analysis, which was based on extensive samples of vocalizations from a large colony of marmosets at Johns Hopkins University (Baltimore, MD) [ref. 24; J. A. Agamaite & X.W. (1997) *Association of Research in Otolaryngology Abstr.* 20, 144]. The results of this study showed that (*i*) marmosets have discrete call types with distinct acoustic structures, and (*ii*) idiosyncratic vocal features of individual monkeys are quantifiable on the basis of their acoustics and are clustered in a multidimensional space (Fig. 1*D*). Acoustic structures of

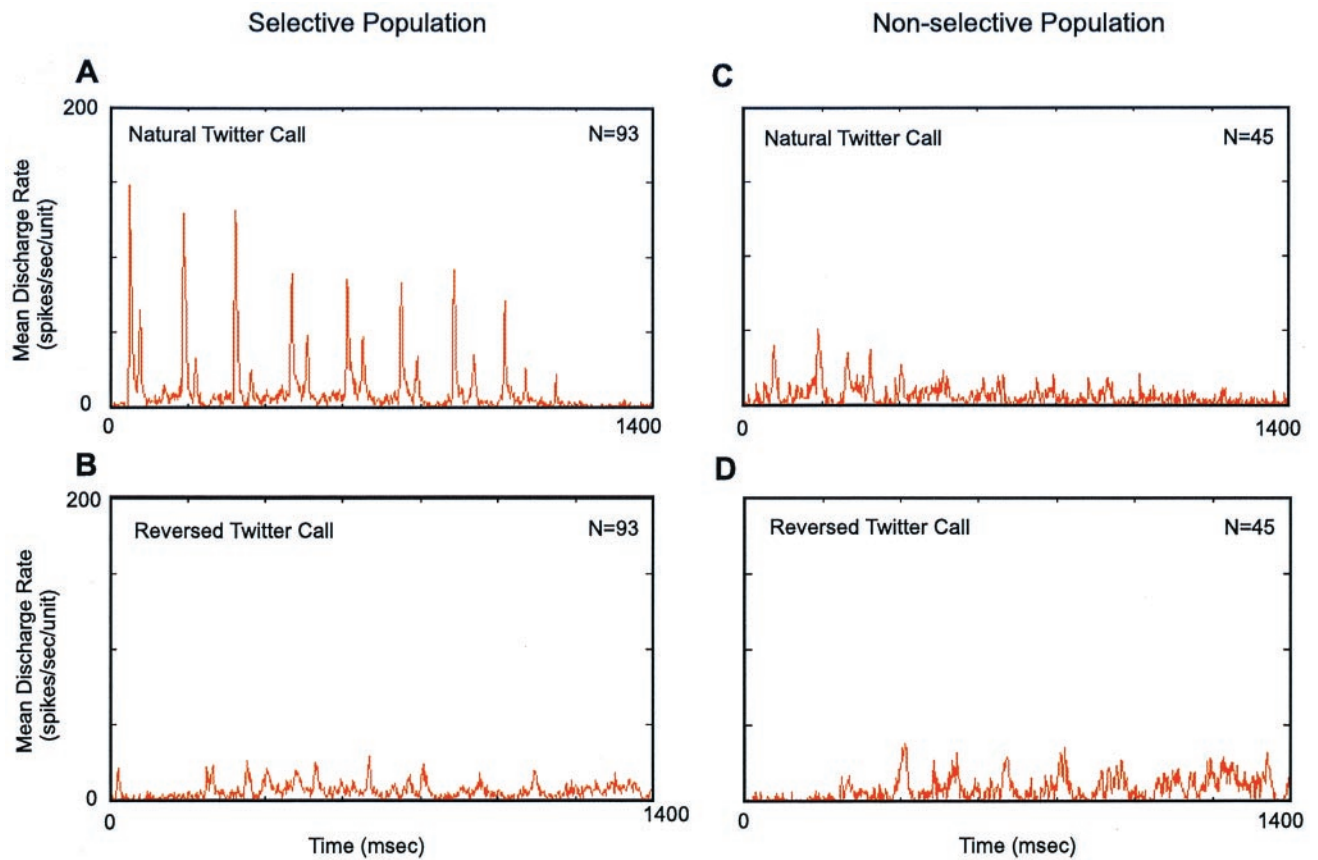


Fig. 2. Averaged temporal discharge patterns of responses to a natural twitter call and its time-reversed version recorded from the primary auditory cortex of a marmoset (27). In each of 138 sampled units, its discharge rate to the natural twitter call was compared with that to the reversed twitter call. The sampled units were divided into two subpopulation based on this analysis. Units included in the *selective population* (A, B) responded more strongly to the natural twitter call than to the reversed twitter call, whereas the units included in the *nonselective population* (C, D) responded more strongly to the reversed twitter call than to the natural twitter call. In A–D, a mean poststimulus histogram (PSTH) is shown for each neuronal population under one of the two stimulus conditions (bin width = 2.0 ms).

marmoset vocalizations were found to contain sufficient information for the discrimination of both caller and gender. In earlier studies of squirrel monkeys, individual call types were identified, but the statistical structure of call types was not fully evaluated, nor were vocal features related to an individual monkey (25). Our studies in marmosets filled this gap and provided a solid basis to further the exploration of cortical coding of communication sounds. Among the published reports, the most comprehensive analysis of the vocal repertoire of communication sounds in a mammalian species was recently conducted in mustached bats in the laboratory of Nobuo Suga (26).

One open issue is how the vocal repertoire recorded in captivity differs from what exists in the wild. It is likely that certain types of calls that are observed in a more behaviorally enriched natural environment are not observable in captivity. However, in the case of common marmoset, there has been no evidence that the basic acoustic structures of calls from captive animals differ fundamentally from those of corresponding calls sampled from animals in the wild. Ideally, quantitative analysis of vocalizations similar to what we have conducted in captive marmosets should be performed in future studies in a wild population of marmosets and other primate species.

Dependence of Cortical Responses on Behavioral Relevance of Vocalizations. In a recent study, it was demonstrated that natural vocalizations of marmoset monkeys produce stronger responses in the primary auditory cortex than do an equally complex but

artificial sound such as a time-reversed call (27), as illustrated in Fig. 2. Moreover, the subpopulation of cortical neurons that were selective to a natural call had a more clear representation of the spectral shape of the call than did the nonselective subpopulation (Fig. 3 A and B). When the reversed call was played, responses from two populations were similar (Fig. 2 B and D and Fig. 3C). These observations suggest that marmoset auditory cortex preferentially responds to vocalizations that are of behavioral significance as compared with behaviorally irrelevant sounds.

To further test this notion, we have directly compared responses to natural and time-reversed calls in the auditory cortex of another mammal, the cat, whose A1 shares similar basic physiological properties (e.g., characteristic frequency, threshold, latency, etc.) to that of the marmoset [ref. 28; X.W., R. Beitel, C. E. Schreiner & M. M. Merzenich (1995) *Soc. Neurosci. Abstr.* 21, 669]. Unlike marmoset, however, cat's auditory cortex does not differentiate natural marmoset vocalizations from their time-reversed version (X.W. & S. Kadia, unpublished observations). One obvious conclusion from both of these lines of evidence is that it is essential to study cortical representation of communication sounds in an appropriate model animal with relevant stimuli. This is not necessarily a trivial issue, because much of our knowledge of subcortical auditory system has been based on studies using artificial stimuli. One likely factor that contributes to the observed response differences is experience-dependent cortical plasticity. As is long known, a behaviorally important stimulus can induce profound changes in neuronal

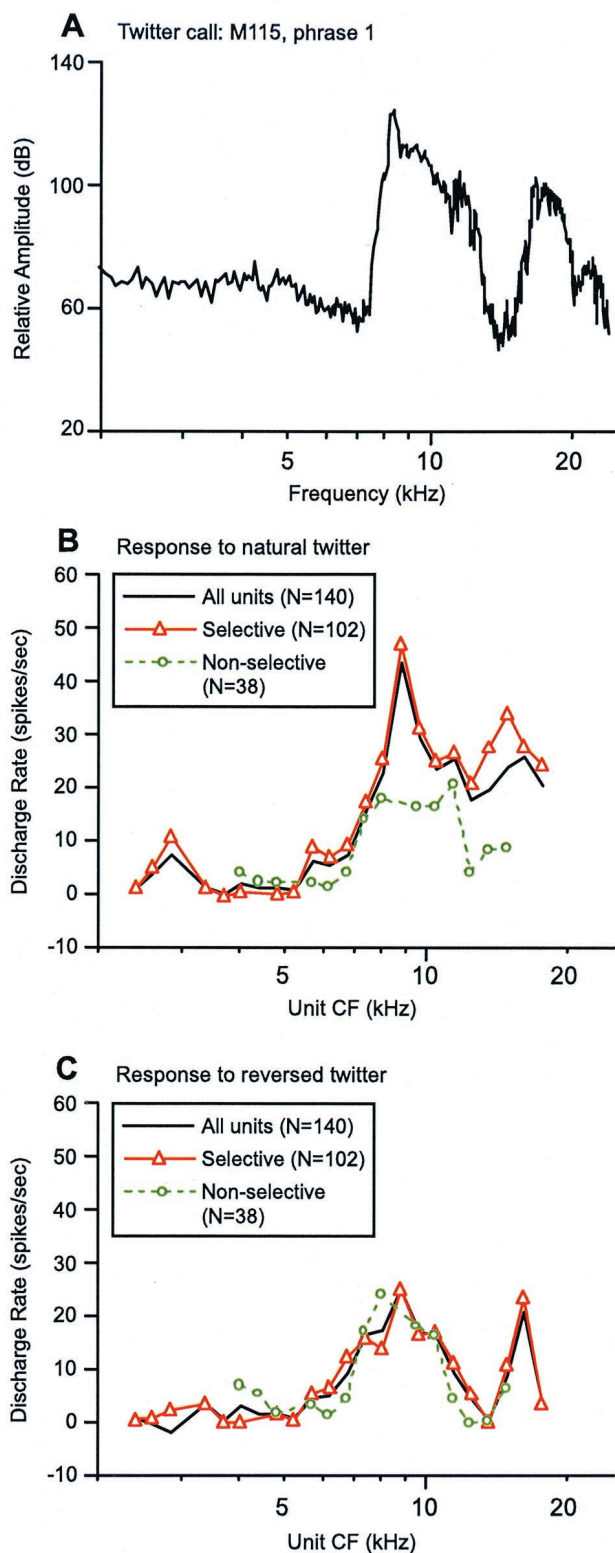


Fig. 3. Population representation of the spectral shape of marmoset vocalizations. Comparison is made between short-term call spectrum of one phrase of the twitter call and rate-CF (discharge rate vs. characteristic frequency) profiles computed over a corresponding time period. Data shown were obtained from the primary auditory cortex of one marmoset (27). (A) Magnitude spectrum of this call phrase in the time-reversed call is the same; only the phase spectrum is different. (B) Rate-CF profiles were constructed based on responses to a natural twitter call from 140 sampled units and were computed by using a

responses of the sensory cortex even within a short period (29, 30). Such contributions would not be observed if vocalizations from unrelated conspecifics were used in obtaining cortical responses. It is possible that the difference between responses from cat and marmoset A1 are due to behavioral modulation on a larger time scale, such as differential experiences through development or even evolution. Further studies are needed to address these issues.

The Nature of Cortical Representation. A fundamental question that has yet to be satisfactorily answered is the nature of cortical representation of species-specific vocalizations. The lack of convincing evidence of call detectors from earlier studies led us to consider other hypotheses regarding cortical coding of complex sounds such as vocalizations. An alternative strategy of encoding complex vocalizations is by the discharge patterns of spatially distributed neuronal populations. Such coding strategies have been demonstrated in the auditory nerve (31, 32) and the cochlear nucleus (33). Recent work (27) in the marmoset has provided evidence that a population coding scheme may also operate at the input stage of the auditory cortex, but in a very different form from that observed at the periphery as illustrated in Fig. 4. Fig. 4A shows population responses to three types of marmoset vocalizations. The difference between cortical and peripheral representations lies largely in the temporal characteristics of neuronal discharges. Compared with the auditory nerve, cortical neurons do not faithfully follow rapidly changing stimulus components (Fig. 4D). We recently examined this issue in the auditory cortex using click train stimuli and found that stimulus-following capacity of cortical neurons is limited to about 20–30 msec (34). Are rapid stimulus components really lost in cortical representation? The answer is no. It turned out that a subpopulation of cortical neurons responded to short interstimulus intervals with changing discharge rate (34). These neurons can potentially signal rapidly changing stimulus components.

It should be pointed out that the notion of population coding does not necessarily exclude the operation of highly specialized neurons. The questions are to what extent such specialized neurons exist and what are their functional roles? In earlier studies, it was expected that a large percentage of these neurons would be found and that the coding of species-specific vocalizations is essentially carried out by such neurons. Various studies since 1970s have cast doubts on this line of thinking. However, a recent study in mustached bats found neurons selective to the bat's communication calls in the FM-FM area of the auditory cortex (35). In the marmoset, a subpopulation of A1 neurons appeared to be more selective to alterations of natural calls than was the rest of the population (27). In our most recent studies in awake marmosets, we identified neurons that were selective to specific call types or even individual callers in both A1 and secondary cortical fields [X.W. (1999) *Association of Research in Otolaryngology Abs.* 22, 173]. Thus, there appear to be two general classes of cortical neurons, one responding selectively to call types or even callers, the other responding to a wider range of sounds including both vocalization and nonvocalization signals. The latter group accounts for a larger proportion of cortical neurons in A1 than in the lateral field (a

triangular weighting window whose base was 0.25 octave wide. The centers of adjacent windows were 0.125 octave apart. Only averages that had at least 3 units in the window were included. Three profiles are shown, all units ($n = 140$, black solid line), selective subpopulation ($n = 102$, red solid line with triangle), and nonselective subpopulation ($n = 38$, green dashed line with circle). The definitions of the two subpopulation of units are given in Fig. 2. (C) Rate-CF profiles are shown for cortical responses to the same call phrase as analyzed in B but delivered in the time-reversed call. The same analytic method and display format are used as in B.

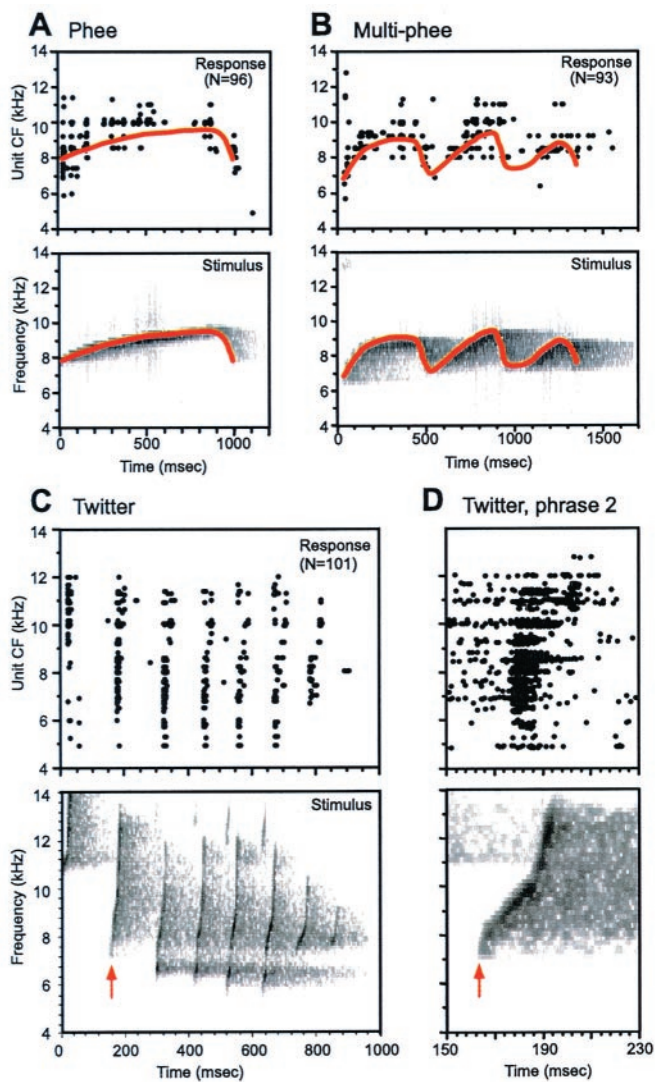


Fig. 4. Comparison between the spectrotemporal acoustic pattern of marmoset vocalizations and the corresponding spectrotemporal discharge patterns recorded in the primary auditory cortex of marmosets. In each plot (A–D), *Upper* shows population responses to a vocalization, and *Lower* shows the corresponding spectrogram of the stimulus. Discharges as they occurred in time (abscissa) from individual cortical units are aligned along the ordinate according to their objectively defined CF. The display of discharges was based on PSTHs computed for each unit (bin width = 2.0 ms). All three vocalizations were delivered at the sound level of 60 dB SPL during the experiments. (A) Population responses to a marmoset *phee* call. An outline of the trajectory (solid line in red) of the call’s time-varying spectral peak is drawn on *Upper* and *Lower* for comparison. (B) Population responses to a marmoset *multiple-phee* call. An outline of the trajectory (solid line in red) of the call’s time-varying spectral peak is drawn on *Upper* and *Lower* for comparison. (C) Population responses to a marmoset *twitter* call. The second call phrase is indicated by a vertical arrowhead (red). (D) An expanded view of cortical responses to the second phrase of the twitter call shown in C (*Upper*) and the corresponding spectrogram of the second call phrase (*Lower*), with a time mark indicated by an arrowhead (red) as in C. Responses of the same group of cortical units shown in C are included but displayed in the form of dot raster. In *Upper*, each recorded spike occurrence within the time period shown is marked as a dot. Spike times from 10 repetitions in each unit are aligned along 10 lines centered at the CF of the unit, shifted by 10 Hz for each repetition for display purpose (i.e., positioned from CF –50 Hz to CF +40 Hz, in 10-Hz step).

secondary cortical area), as observed in our experiments in awake marmosets [X.W. (1999) *Association of Research in Otolaryngology Abs.* 22, 173]. One may speculate that the task of encoding

communication sounds is undertaken by the call-selective neurons alone, however small that population is. On the other hand, one cannot avoid considering the role of the nonselective population that also responds to communication sounds concurrently with the selective population. One hypothesis is that the nonselective population serves as a general sound analyzer that provides the brain with a detailed description of a sound, whereas the selective population serves as a signal classifier whose task is to tell the brain that a particular call type or caller was heard. One can imagine that the former analysis is needed when a new sound is heard or being learned, whereas the latter analysis is used when a familiar sound such as a vocalization is encountered. This and other theories (36, 37) are likely to be tested by new experiments in coming years. Nonetheless, two important lessons have been learned from earlier and recent studies. First, it is not sufficient to study individual call tokens; one must study the representation of statistical structures of communication sounds. Second, cortical representations of complex sounds like vocalizations are no longer isomorphic replicas of the acoustic spectrotemporal pattern of the sound; they have been significantly transformed into a new representational space.

A “Vocal” Pathway. A further question on cortical coding is how the information in communication sounds is routed through cortical systems, including those outside the superior temporal gyrus. It has been suggested that there are “where” and “what” pathways in the auditory cortex (6, 17), a notion largely borrowed from the visual system. However, in our opinion, the auditory cortical system in humans and primates must include another pathway, a “vocal” pathway, in processing vocal communication sounds. This pathway may or may not be involved when nonvocal acoustic signals are processed by the cerebral cortex. There is an important distinction between the auditory and visual system, however. Vocal species such as humans and primates produce their own behaviorally important inputs (i.e., speech and vocalizations) to their auditory system. As has been argued in this article, the “vocal” pathway is likely to be a mutual communication channel between the superior temporal gyrus and the frontal cortex. Anatomically, it has been shown by a number of studies in both humans and primates that the superior temporal gyrus is connected with the frontal lobe, reciprocally, where generation and control of vocal activities take place in humans and possibly in primates as well.

Future Directions. A number of issues remain to be resolved in the overall understanding of cortical coding of communication sounds. First, a careful consideration of the state of an animal must be made when its cortical responses are studied. Much of the understanding of the auditory cortex has been based on studies in anesthetized animals, which obviously carry severe limitations. An important step in moving this field forward is the use of awake and behaving preparations, which have been widely adapted in studies of the visual system. Only then can important issues like attentional modulation of cortical responses be adequately evaluated. Second, two crucial issues that must be answered in this field are vocal production mechanisms and vocal development and learning in primates. The essential question is whether primates possess voluntarily controlled and learned vocalizations. Earlier studies have painted negative pictures of both prospects. In the coming years, these earlier conclusions are likely to be challenged and possibly modified. Finally, to study these and other emerging questions on cortical coding of vocal communication sounds, new techniques are needed. The main limitation of existing neurophysiological methods is that vocal behavior of animals is substantially restricted or eliminated once an animal is restrained. Using implanted electrodes can loosen these restrictions. Such techniques are widely used in the study of rodents with tethered cables to relay signals from the electrodes to a data recorder. However, for highly mobile primates, especially when one is interested in their natural vocal activity, the

ideal way to relay signals from a recording electrode array is by means of a telemetry device.

I thank members of the Laboratory of Auditory Neurophysiology at the Biomedical Engineering Department of the Johns Hopkins University for their contributions toward the research and opinions discussed in this article, in particular James Agamaite, Thomas Lu, Siddhartha Kadia, Ross Snider, Dennis Barbour, Li Liang, Steve Eliades, and Haiyin Chen. I thank Airi Krause and Ashley Pistorio for

assistance in animal training and marmoset colony management. I am especially grateful for the encouragement and support of Dr. Michael Merzenich at the Coleman Laboratory of the University of California at San Francisco, where the work on the marmoset model was initiated. I thank D. Barbour for his excellent help in graphics work and T. Lu and D. Barbour for proofreading the manuscript. This research was supported by National Institutes of Health Grant DC03180 and by a Presidential Early Career Award for Scientists and Engineers (1999–2004).

1. Penfield, W. & Roberts, L. (1959) *Speech and Brain-Mechanisms* (Princeton Univ. Press, Princeton, NJ).
2. Heffner, H. E. & Heffner, R. S. (1986) *J. Neurophysiol.* **56**, 683–701.
3. Brodmann, K. (1909) *Vergleichende Lokalisationslehre der Grobhirnrinde* (Barth, Leipzig).
4. Jones, E. G. & Powell, T. P. S. (1970) *Brain* **93**, 793–820.
5. Pandya, D. N. & Yeterian, E. H. (1985) in *Cerebral Cortex*, eds. Peters, A. & Jones, E. G. (Plenum, New York), Vol. 4, pp. 3–61.
6. Kaas, J. H., Hackett, T. A. & Tramo, M. J. (1999) *Curr. Opin. Neurobiol.* **9**, 164–170.
7. Snowdon, C. T., Brown, C. H. & Petersen, M. R. (1982) *Primate Communication* (Cambridge Univ. Press, Cambridge, U.K.).
8. Epple, G. (1968) *Folia Primatol.* **8**, 1–40.
9. Aitkin, L. M., Merzenich, M. M., Irvine, D. R., Clarey, J. C. & Nelson, J. E. (1986) *J. Comp. Neurol.* **252**, 175–185.
10. Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cynader, M. S., Schoppmann, A. & Zook, J. M. (1984) *J. Comp. Neurol.* **224**, 591–605.
11. Merzenich, M. M., Kaas, J. H., Wall, J. T., Sur, M., Nelson, R. J. & Felleman, D. J. (1983) *Neuroscience* **10**, 639–665.
12. Liberman, A. M. (1996) *Speech: A Special Code* (MIT Press, Cambridge, MA).
13. Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. (1992) *Science* **256**, 846–849.
14. Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M. & Prieto, T. (1997) *J. Neurosci.* **17**, 353–362.
15. Ojemann, G. A. (1991) *J. Neurosci.* **11**, 2281–2287.
16. Morel, A. & Kaas, J. H. (1992) *J. Comp. Neurol.* **318**, 27–63.
17. Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S. & Rauschecker, J. P. (1999) *Nat. Neurosci.* **2**, 1131–1136.
18. Rauschecker, J. P., Tian, B. & Hauser, M. (1995) *Science* **268**, 111–114.
19. Pelleg-Toiba, R. & Wollberg, Z. (1991) *J. Basic Clin. Physiol. Pharmacol.* **2**, 257–272.
20. Winter, P. & Funkenstein, H. H. (1973) *Exp. Brain Res.* **18**, 489–504.
21. Newman, J. D. & Wollberg, Z. (1973a) *Exp. Neurol.* **40**, 821–824.
22. Wollberg, Z. & Newman, J. D. (1972) *Science* **175**, 212–214.
23. Manley, J. A. & Müller-Preuss, P. (1978) *Exp. Brain Res.* **32**, 171–180.
24. Agamaite, J. A. (1997) M.S.E. thesis (Johns Hopkins University, Baltimore, MD).
25. Winter, P., Ploog, D. & Latta, J. (1966) *Exp Brain Res.* **1**, 359–384.
26. Kanwal, J. S., Matsumura, S., Ohlemiller, K. & Suga, N. (1994) *J. Acoust. Soc. Am.* **96**, 1229–1254.
27. Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. (1995) *J. Neurophysiol.* **74**, 2685–2706.
28. Schreiner, C. E., Read, H. L. & Sutter, M. L. (2000) *Annu. Rev. Neurosci.* **23**, 501–529.
29. Recanzone, G. H., Schreiner, C. E. & Merzenich, M. M. (1992) *J. Neurosci.* **13**, 87–104.
30. Wang, X., Merzenich, M. M., Sameshima, K. & Jenkins, W. M. (1995) *Nature (London)* **378**, 71–75.
31. Sachs, M. B. & Young, E. D. (1979) *J. Acoust. Soc. Am.* **66**, 470–479.
32. Young, E. D. & Sachs, M. B. (1979) *J. Acoust. Soc. Am.* **66**, 1381–1403.
33. Blackburn, C. C. & Sachs, M. B. (1990) *J. Neurophysiol.* **63**, 1191–1212.
34. Lu, T. & Wang, X. (2000) *J. Neurophysiol.* **84**, 236–246.
35. Esser, K. H., Condon, C. J., Suga, N. & Kanwal, J. S. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14019–24.
36. Suga, N. (1994) *J. Comp. Physiol. A* **175**, 135–144.
37. Suga, N. (1994) in *The Cognitive Neuroscience*, ed. Gazzanica, M. S. (MIT Press, Cambridge, MA.), pp. 295–313.