

## **What is the Neural Code of Species-specific Communication Sounds in the Auditory Cortex of Primates?**

**Xiaoqin Wang**

*Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, U.S.A.*

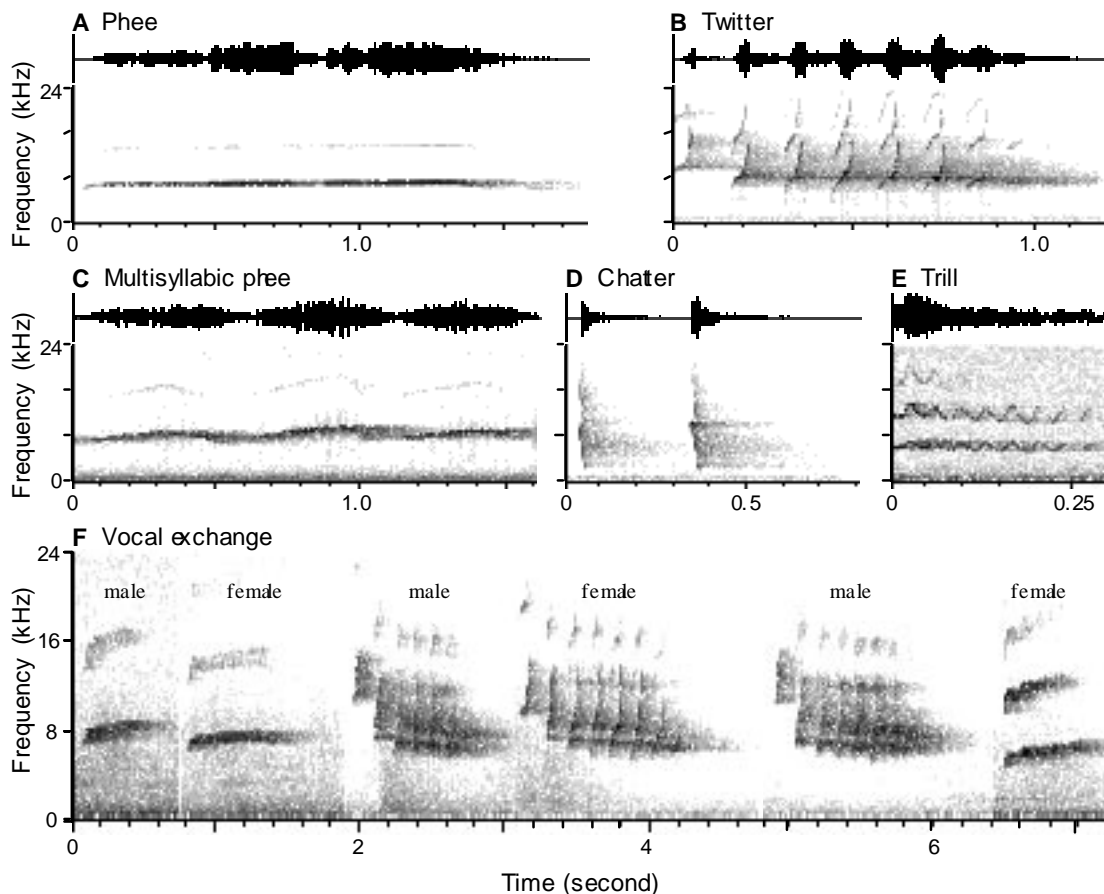
### **Introduction**

The reception of species-specific communication sounds is an important aspect of the auditory behavior of many primate 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 limited, compared with our understanding of other auditory functions such as sound localization and echolocation. This article intends to give a brief review of our recent work in a vocal primate, the common marmoset (*Callithrix jacchus jacchus*), the conceptual basis of this line of work and the hypothesis being tested.

### **Why primate communication sounds are important to study?**

Communication sounds are a subset of vocalizations produced by a species that are used for intra-species communication. Species-specific vocalizations of non-human primates and human speech are examples of communication sounds. Vocal repertoires of species such as echolocation bats also include sonar signals that are used to determine target properties (distance, velocity, etc.). Many avian species also have rich vocal repertoires. Some of the best work in our understanding of brain processing of animal vocalizations has been conducted in bats and song birds (e.g., Suga 1988). Communication sounds of primates are a class of acoustic signals of great interests to us because they are the closest analogy of human speech despite many differences that lie between the two. Comparing with other non-human 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 have direct implications for those operating in the human brain.

Almost all primate species have sophisticated vocal repertoire in their natural behavior. However, many primate species become silent in captive conditions, in part due to the lack of a behaviorally suitable housing environment. 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. Fortunately, some primate species such as marmosets preserve their vocal activity well in properly configured captive conditions. These species provide excellent models for us to study in details their vocal behavior as well as underlying neural mechanisms in the brain. Figure 1 shows examples of marmoset vocalizations, which are only a subset of more than 20 call types that have been identified in this species (Agamaite and Wang 1997).



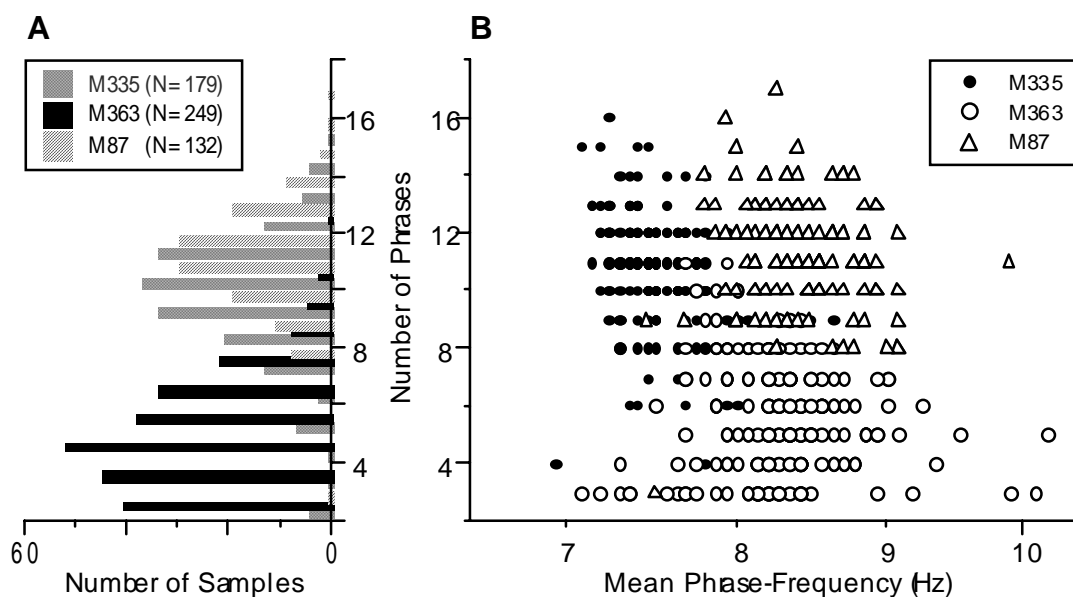
**Figure 1.** A-E: Examples of marmoset vocalizations recorded in captivity. Both the waveform (upper panel) and the spectrogram (lower panel) of each call are shown. F: A vocal exchange between a pair of marmosets is shown in the form of spectrogram.

It should be pointed out that the cortex of primates is capable of processing not only species-specific vocalizations but other acoustic signals as well (e.g., sounds made by predators and preys). The advantage of using communication sounds as a model to unlock cortical mechanisms underlying auditory processing is that the solution to the problem is guaranteed by the behavioral relevance of these sounds. Our understanding of how the cortex solves this unique problem should greatly facilitate our understanding of cortical mechanisms *in general*, so much like finding a special solution to a complex differential equation often produces invaluable insights leading to the general solution of the equation.

### What kind of information do primate communication sounds carry?

In order to understand how the cortex processes communication sounds, it is necessary to know what kind of information these sounds carry and by what acoustic features? There are three important behavioral tasks for which primates rely on their vocalizations, that is, to identify *location*, *meaning* and *caller* of a vocalization. Until recently, studies on informational contents of primate vocalizations remained largely qualitative. Using digital signal processing techniques, more recent studies, including those in our laboratory, have provided clear evidence that acoustic features of primate communication sounds contain precise information to distinguish call types and callers

(Agamaite and Wang 1997). For example, a marmoset phee call is characterized by a slow-varying, continuous frequency sweep (Fig.1A), whereas a twitter call is composed of several segments of rapid, ascending FM sweeps (Fig.1B). These two types of calls can be reliably identified based on their FM characteristics. The problem of identifying signatures of individual callers is a more difficult one. All marmosets vocalize every types of calls, but with their idiosyncratic features. Figure 2A illustrates that some marmosets produce twitter calls with fewer phrases than do others, a property that can be measured by a call feature referred to as *number of phrases*. However, using this feature alone does not adequately distinguish calls made by three animals shown in the example given in Fig.2. If another call feature, *mean phrase-frequency*, is added to the analysis, the three animals can then be reliably distinguished (Fig.2B). Data in Fig.2 illustrate an important observation we have learnt from analyzing marmoset vocalizations, i.e., marmosets can potentially obtain essential information on call types and individual identities using multiple *call features* of a complex vocalization, similar to the way humans analyze speech sounds using particular features such as formants of vowels. Our analysis also indicates that the dimension of this presumably multi-dimensional analysis is relatively small, which suggests that it is feasible to selectively manipulate specific acoustic features in the investigation of the perception and neural encoding of complex vocalizations.



**Figure 2.** Quantitative analysis of individual differences. **A:** Histograms of the number of phrases measured from twitter calls vocalized by three marmosets **B:** Comparison the same group of calls in a two dimensional space, number of phrases versus mean phrase-frequency (equal to the inverse of mean inter-phrase interval measured from the amplitude envelop).

### Earlier studies on cortical encoding of primate communication sounds

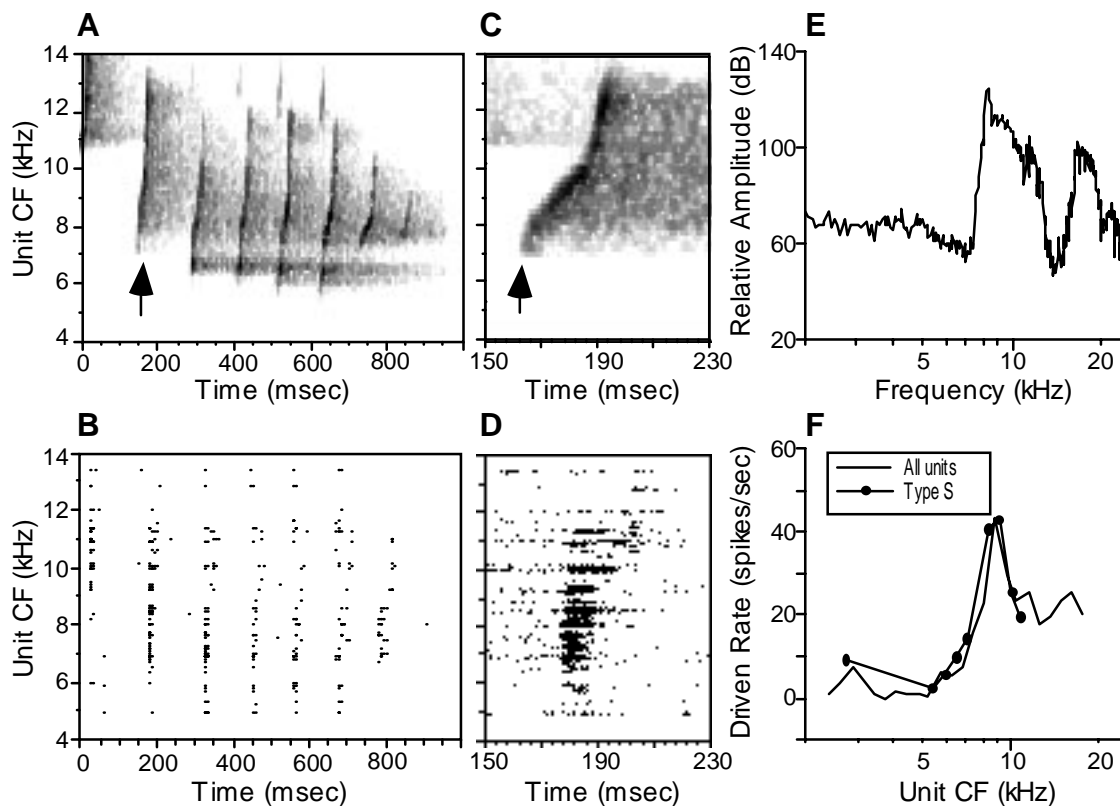
The cortex is known to play an important role in processing species-specific vocalizations. Over the past several decades, a number of experimental attempts have been made to elucidate the forms of cortical representations of species-specific vocalizations in the auditory cortex of primates. The results of these studies have been mixed, with no clear or consistent picture emerging as to how behaviorally relevant complex sounds are "coded" in the auditory cortex (see reviews by Pelleg-Toiba and Wollberg 1991). This lack of

success 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" (Winter and Funkenstein 1973, Newman and Wollberg 1973). However, individual neurons in the auditory cortex were often found to respond to more than one call or to various "features" of calls (Winter and Funkenstein 1973, Newman and Wollberg 1973, Wollberg and Newman 1972, Manley and Muller 1978). The initial estimate of the percentage of "call detectors" was 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 was clear that, at least in the initial stages of cortical pathway, there did not appear to exist highly specialized neurons in primates. Perhaps it was due to these seemingly frustrated findings, no systematic studies on this subject were reported for more than 10 years after then.

### **Recent work in the auditory cortex of marmoset**

An alternative strategy of encoding complex vocalizations is by the discharge patterns of spatially distributed neuronal populations. At the peripheral level, such coding strategy has been demonstrated in a number of studies (e.g., Sachs and Young 1979). Recent work in the primary auditory cortex (A1) of marmoset has provided evidence that a population coding scheme also operates at the level of cortex, but taking a very different form from that seen at the periphery (Wang et al. 1995a). To overcome some of the limitations in the earlier studies, our work on the marmoset auditory cortex began with three objectives: 1) to define population representation on the basis of a large number of neurons recorded from each hemisphere, 2) to relate responses to complex vocalizations to responses to simpler stimuli, and 3) to compare cortical responses to natural vocalizations with those to synthetic vocalizations with altered acoustic structures but similar complexity.

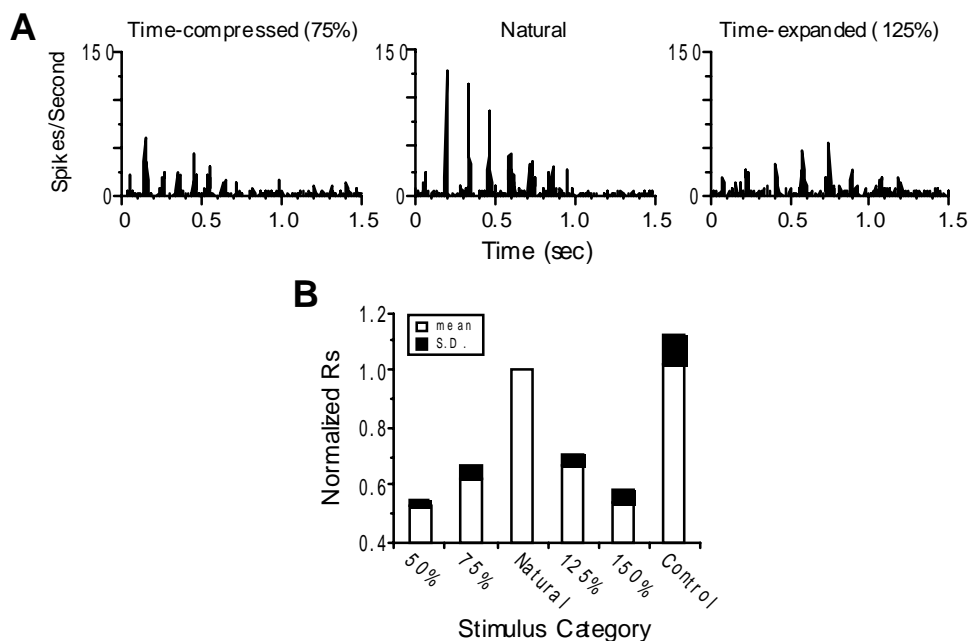
The first important observation was that spectrotemporal discharge pattern of spatially distributed neuronal populations in A1 was correlated with the spectrotemporal acoustic pattern of a complex natural vocalization as illustrated by the example shown in Fig.3A-B. Major acoustic features of this complex vocalization are captured in the population responses of A1 neurons. In the frequency domain, spectral peaks and troughs were represented by their counterparts in rate-CF profiles (Fig.3E-F). In contrast, temporal features are not faithfully represented in A1 response patterns. A time-varying segment of a marmoset twitter call is reduced to firings of many excited neurons randomly distributed within a narrow time window (Fig.3C-D). This finding is consistent with the basic response properties we know of A1 neurons, namely, they are not able to follow stimulus dynamics that vary at a rate faster than 30-50 Hz, even in unanesthetized cortex (Creutzfeldt et al. 1980). The reduction in temporal dynamics in cortical firings suggests that A1 has to encode such information in other forms. This loss of temporal dynamics also has a direct consequence, it increases *temporal synchrony* among many evoked neurons that are distributed far away from each other across cortical surface. This is actually a desirable property because it leads to an enhancement of stimulus-evoked discharges when the outputs of A1 neurons are integrated in the next processing stage. Neurons in the secondary cortical fields generally have wider receptive fields and thus integrate A1 outputs across sections of the tonotopic axis. In summary, the A1 representation of a complex vocalization was not a faithful replication of its acoustic pattern, but has been transformed into a more abstract form than that in the auditory periphery.



**Figure 3.** Comparison between acoustic pattern of a twitter call and the corresponding A1 discharge pattern. **A, B:** Spectrogram and population responses, respectively, of a marmoset twitter call. Discharges as they occurred in time from individual cortical units ( $N=100$ ) are aligned according to their CF (binwidth 2.0 msec). **C, D:** an expanded view of the spectrogram and population responses of the second phrase of the twitter call shown in **A** and **B**. Arrowheads in **A** and **C** point to the same position in time. Responses shown in **D** were from the same group of cortical units shown in **B**, but now displayed in the form of dot roster. **E, F:** Comparison of short-term spectrum (**E**) and rate-CF profiles (**F**) computed over a period of time that includes the first call phrase. Rate-CF profiles shown were based on 140 A1 units (including 37 type-S units, see the definition in Fig.4 legend) recorded in one marmoset, and were computed using a triangular weighting window (base: 0.25 octave, centers of adjacent windows: 0.125 octave apart). Only averages that had at least 3 units in the window were included.

The second major observation in our marmoset study is that A1 neurons did not respond in the same way to natural and acoustically altered vocalizations, both of which have comparable complexity. The acoustic alterations changed temporal characteristics of a natural vocalization (compressed or expanded) while keeping the spectral characteristics unchanged. These manipulations produced significant alterations in call features that are known to be important in call identification, such as the *phrase-frequency* of a twitter call (see Fig.2B). For a subset of A1 neurons (referred to as *type-S* units), cortical responses to natural vocalizations were stronger and much more temporally coherent than responses to temporally altered vocalizations (Fig.4A, also see Wang et al. 1995a). Such response differences may be the results of 1) behavioral relevance of the sounds as altered vocalizations never exist in marmoset's behavior and 2) neuronal sensitivity to specific acoustic features such as amplitude envelope and frequency sweep directions. At this point, we can not conclusively determine specific contributions of these two factors. Experiments are underway to directly address this issue. However, given the evidence from the studies of cortical plasticity in development and in adulthood, it is likely that the behavioral relevance

of natural vocalizations plays an important role in shaping cortical responses such as those shown in Fig.4, by means of experience-induced plasticity. Interestingly, the rate-CF profile for type-S units gives just as clear an indication of the main spectral peak in a marmoset twitter call as does the rate-CF profile for entire population (Fig.3F), an evidence that further supports the suggestion that such a selective subpopulation of neurons play an important role in representing a specific vocalization.



**Figure 4.** Response properties of Type-S units, a subpopulation of A1 neurons that preferentially responded to a natural twitter call. These neurons accounted for about 30% of all A1 neurons studied in the frequency range of marmoset twitter calls. An A1 unit was defined as a type-S unit if it responded better to a natural call than to its temporally-altered variations. **A:** Examples of compound PSTHs generated by summing individual PSTHs derived from 14 type-S units recorded in one marmoset in response to a natural twitter call and to several temporally altered variations of this call. Stimulus onset was at  $t=0$ . **B:** Relative strength of Synchronized Discharge Rate ( $R_s$ ) in different stimulus conditions for type-S units averaged from all experiments (a total of 728 units), normalized by  $R_s$  for the natural call.  $R_s$  measures a portion of the total discharges that are synchronized to the repetitive events in a stimulus. The control stimulus (100%) was a synthetic "natural call", produced using a temporal-alteration ratio of 1.0.

### Neural code of communication sounds in the auditory cortex: a hypothesis

In the primary auditory cortex, as we have seen from the study in marmoset, highly specialized "call detector" neurons do not appear to exist in primates. The neural representation of communication sounds seems to be based on widely distributed neuron populations. Does this mean that we should view A1 as a purely analytic machine that processes complex signals entirely on the basis of their acoustic structures? Observations from our marmoset studies indicate that A1 is more than a static analyzer. Supporting evidence also comes from a long line of cortical plasticity studies (see review by Merzenich and Jenkins 1993) that argue for the formation of stimulus-specific neural assemblies in the sensory cortex through learning and experience (Wang et al. 1995b). Our hypothesis is that a behaviorally important stimulus such as a species-specific vocalization is effectively represented by a stimulus-specific subpopulation of neurons distributed across a cortical field. Neuronal members of a specific subset result from acoustic properties (spectral and

temporal characteristics) of a stimulus as well as the competition induced by experience-dependent neural plasticity in development and adulthood. Member neurons differ from non-member neurons in that they respond to the stimulus to which the subset belongs more strongly and with greater temporal coherence. Because of overlaps between complex stimuli in their acoustic structures and their usage, each A1 neuron may participate in more than one subpopulations that are partially overlapping. In comparison, cortical responses of stimuli that are not involved in an animal's behavioral experience would be more evenly distributed among all excitable neurons in a field. The consequence of having a specific subpopulation of A1 neurons devoted to a behaviorally important stimulus is that its representation in the next cortical processing stage, say a secondary field, will be even more specific, through synaptic integration and plasticity. This should result in more efficient processing of the learned or experienced stimuli, such as sounds in one's native or well-learned second language. Whether this type of chain-reaction effects will progress to the extent that it eventually leads to highly specialized "call detectors" somewhere in the cerebral cortex remains to be seen. On the other hand, we don't necessarily need so specialized detectors or a "pilot" to tell us what we hear, just as one does not have to be in a helicopter to know that he or she is on the top of Mount Everest.

## Acknowledgments

This research is supported, in part, by NIDCD Grant DC03180 and The Deafness Research Foundation.

## References

- Agamaite, J.A. and Wang, X. (1997) Quantitative classification of the vocal repertoire of the common marmoset (*Callithrix Jacchus Jacchus*). Association of Research in Otolaryngology Abs 20, 573.
- Creutzfeldt, O., Hellweg, F.-C., and Schreiner, C. (1980) Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res* 39, 87-104.
- Manley, J.A. and Müller-Preuss, P. (1978) Response variability of auditory cortex cells in the squirrel monkey to constant acoustic stimuli. *Exp Brain Res* 32, 171-80.
- Merzenich, M.M. and Jenkins, W.M. (1993) Reorganization of cortical representations of the hand following alterations of skin inputs induced by nerve injury, skin island transfers, and experience. *J Hand Ther* 6, 89-104.
- Newman, J.D. and Wollberg, Z. (1973) Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res* 54, 287-304.
- Pelleg-Toiba, R. and Wollberg, Z. (1991) Discrimination of communication calls in the squirrel monkey: "call detectors" or "cell ensembles"? *J Basic Clin Physiol Pharmacol.* 2:257-72.
- Sachs, M.B. and Young, E.D. (1979) Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *J. Acoust. Soc. Am.* 66, 470-479.
- Suga, N. (1988) 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: John Wiley & Sons., pp. 679-720.
- Wang, X., Merzenich, M.M., Beitel, R., and Schreiner, C.E. (1995a) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiology* 74, 2685-2706.
- Wang, X., Merzenich, M.M., Sameshima, K., and Jenkins, W.M. (1995b) Remodeling of Hand Representation in Adult Cortex Determined by Timing of Tactile Stimulation. *Nature* 378, 71-75.
- Winter, P. and Funkenstein, H.H. (1973) 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.
- Wollberg, Z. and Newman, J.D. (1972) Auditory cortex of squirrel monkey: response patterns of single cells to species specific vocalizations. *Science* 175, 212-214.