A Neuroethological Perspective on the Perception of Vocal Communication Signals

Timothy Q. Gentner\textsuperscript{1} & Gregory F. Ball\textsuperscript{2}

\textsuperscript{1}Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL USA

\textsuperscript{2}Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218 USA

Please address all correspondence to:

Gregory F. Ball

Department of Psychological and Brain Sciences

Johns Hopkins University

3400 N. Charles Street

Baltimore, MD 21218

Email: gball@jhu.edu

Phone: 410-516-7910

Fax: 410-516-6008
The Relationship between Human and Non-Human Animal Studies of Vocal Behavior

One path breaking observation made about speech in the latter half of the 20th century is that it is “special” (Liberman, 1982). An important step taken by the cognitive revolution initiated in the 1950s was to take speech and language out of the realm of general psychological processes such as learning and perception (e.g., Skinner, 1957) by claiming that unique neural and behavioral mechanisms needed to be postulated to explain the learning, production and perception of human speech (Chomsky, 1959). When the plurality of human language behaviors are conceptualized as species-typical biological adaptations (e.g., Lenneberg, 1967; Pinker, 1994), however, the strong claim of uniqueness appears less universally tenable. To say that human language behaviors have evolved, is to say that the neural and behavioral specializations that mediate speech and language abilities are result of natural (and/or sexual) selection pressures in humans. Given that no extant species of primates, including the closely related great apes, share with Homo sapiens comparable faculties of language and speech, it would surprising if humans did not show anatomical and physiological specializations for language production and processing. At the same time, comparative studies (Kleunder et al., 1987; Kuhl and Miller 1975, 1978) also demonstrate that at least some of the behaviors associated with speech perception (e.g. categorical perception, perceptual constancy) are not restricted to humans. As such, it would be equally surprising to find that some
anatomical structures and physiological mechanisms for vocalization and vocal signal perception are not conserved across taxa.

There are at least two views concerning the value of comparative studies in non-human animals to our understanding of human speech. One perspective in the cognitive science community is that studies of non-human animals are simply irrelevant to the study of human speech (e.g., Chomsky, 2000). Chomsky has championed the notion that the “language faculty” of humans is a natural biological phenomenon and should be studied as such. In its strongest form, this argument claims that very few of the behavioral or cognitive phenomena associated with communication in non-human animals have attributes that even approach human speech. Accordingly, studies of non-human animals can only claim relevance to human language by asserting that something that is not language (i.e., non-human animal communication systems) is language. In contrast, others in the human language community (e.g. Liberman, 1996), while completely embracing the notion that language is a specialized human trait, have argued that comparative studies of non-human animals have much to teach students of human language about how a species-typical trait like speech might develop, be produced, and perceived in adulthood. In a trivial sense, animal studies must, by necessity, inform the cellular basis of human language, because the central nervous systems of humans and many non-human animals share certain cell types such as neurons and glia. Although no one is likely to object to such a claim, commonalities of this sort hardly make a good case for considering non-human animal studies. Yet, as we have already noted, commonalities between human and non-humans go far beyond the cellular level.
Recently, Hauser et al. (2002) have suggested that language abilities can be divided into faculties in the “broad sense” and faculties in the “narrow sense.” Broad faculties include sensory-motor processes underlying language production, and the conceptual-intentional abilities to formulate communication content. For example, some of the perceptual attributes that facilitate speech comprehension in humans and are observed in non-human animals can be considered examples of broad language faculties. Narrow faculties include such skills as the recursive abilities exhibited in grammar and syntax. For example, humans can easily take a finite number of forms and recombine them infinitely. This combinatorial skill is obvious in language but can also be observed in non-linguistic domains such as numerical competence. Hauser et al. (2002) argue that comparative studies can illuminate broad language faculties by identifying developmental and neural mechanisms of processes common to human and non-human animals, and narrow faculties by identifying constraints that limit the occurrence of these human-specific traits in non-human animals. In other words, comparative studies can inform our understanding of human language by pointing out important similarities and differences in underlying processes and mechanisms.

Perhaps the strongest case for considering comparative data from non-human animals in understanding human language learning and production is illustrated empirically by neuroethological studies of birdsong. Given the differences in semantics and syntax between the two communication systems, no one would claim that birdsong is a language. However, to learn and to produce songs birds must solve the problem of how the auditory system regulates the development and production of a vocal behavior. More so than any other known animal behavior, vocal learning in songbirds
presents the most direct analogy to the problems solved by the human central nervous system in learning to produce speech and in maintaining vocal production over time (Marler, 1970; Ball and Hulse, 1998; Doupe and Kuhl, 1999). Therefore, although birds do not have language, components of their vocal behaviors show a striking resemblance to important components of human speech.

The evolved sensory-motor processes that mediate vocal learning in songbirds are almost certainly not homologous in all aspects to the evolved processes that mediate language in humans. Yet, our knowledge of the neural and behavioral mechanisms of birdsong has already yielded insight to analogous processes in the human central nervous system (Margoliash, 2003). The goal of the current essay is to extend this analogy to include neurobiological mechanisms that underlie the perceptual processing of human speech and birdsong. We point to commonalities in the pre-semantic processing of species-typical vocal signals, and explore how studies of one can inform our understanding of the other.

**Neuroethology: Neuroscience of Natural Behavior.**

Until very recently, the study of behavior and the study of the brain proceeded on independent tracks. In contrast, it is now understandable that molecular biologists studying the function of brain-specific gene expression collaborate with neuroscientists who study the behavioral expression of the action of neural circuits. Cognitive psychologists and systems-level neuroscientists are working together on problems such as the neural basis of memory, perception and language in a way that would have been unthinkable even 15 years ago (e.g. Gazzaniga, 1995). Emerging disciplines that involve
cross-fertilization among the fields of behavior and neurobiology are given new names such as “Cognitive Neuroscience”. Another hybrid discipline that straddles the behavioral and neural sciences is “neuroethology”. Investigators in this field are interested in the neural and physiological mechanisms of behavior and perception, as are many systems-level neuroscientists. The focus of neuroethology, however, is on natural (i.e., ethological) behaviors.

The Ethological Perspective

Ethology can be broadly defined as the biological study of animal behavior (Immelmann and Beer, 1989). It generally involves the study of naturally occurring behaviors (i.e. species-typical behaviors) in the context of the animal’s ecology. Ethological investigations can be conducted in a field setting or in the laboratory but the importance of relating findings to naturally occurring behaviors is paramount (Tinbergen, 1963).

In the 1960s, Brown and Hunsperger (1963) coined the term “neuroethology” to describe their studies of the neural basis of species-typical behaviors. Early neuroethology was dominated by studies of invertebrates (see Camhi, 1984). Contemporary work on vertebrate neuroethology is best known to most neuroscientists because of its success in elucidating the neural circuits mediating such behaviors as echolocation in bats, sound localization in owls, electroreception and perception in weakly electric fish, and the learning, production and perception of birdsong.

Modern ethology emerged from a debate between ethologists and comparative psychologists that was carried out in the 1950s (see Grasse, 1956). From this debate a
consensus of sorts emerged that was articulated by Tinbergen (1963), and later given
more detail by Hinde (1970). Tinbergen pointed out that at least some of the
disagreements between the ethologists and the comparative psychologists were due to the
fact that they were unknowingly asking different questions about behavior. He argued
that there are four basic causal questions one can ask about a particular behavior. These
are: 1. Immediate causation: What stimulus and/or physiological factors cause this
behavior to be produced at this time rather than another time; 2. Behavioral development:
What ontogenetic processes mediate the emergence of a given behavior in the adult
animal? ; 3. Adaptive significance: What is the functional significance of engaging in this
behavior in terms of reproductive success or survival? ; 4. Evolutionary history: How did
this behavior evolve over taxonomic history? Questions 1 and 2 are often considered
together under the rubric of “proximate” cause, and questions 3 and 4 are referred to
collectively as “ultimate” causes. These so-called “levels of analysis” (Sherman, 1988)
provide a very useful heuristic in organizing multiple explanations of the causes and
significance of observed behaviors.

**Neuroethology**

Following the scheme developed by Tinbergen, *neuro*ethologists ask questions
related to the physiological mechanisms mediating immediate (i.e. proximate) causation
and development. They focus on how the nervous system and related physiological
systems such as the endocrine system and the immune system function to regulate the
occurrence of species-typical behavior. Although they usually work in laboratory
settings, they often investigate wild species and pose questions in terms of the organism’s natural history.

One of the earliest intersection points between classical ethology and behavioral neuroscience concerns the selectivity of perception. Although it is obvious that perception requires an active selection process of some sort, ethologists discovered that an extraordinarily limited range of stimuli often lead an animal to exhibit a behavioral response. Indeed, based on experiments in a variety of taxa, it became clear that animals often respond to only one special part of the array of complex stimuli presented to them. Stimuli that are especially potent in eliciting behavioral responses are referred to as "sign stimuli" by ethologists (Tinbergen, 1951). These sorts of relationships between a stimulus and response underscore the notion, first articulated by von Uexküll, that one must consider an animal’s functional environment based on its perceptual abilities, to understand the organization of the mechanisms controlling its behavior (von Uexküll, 1934). The concept of a sign-stimulus has since guided neuroethologists in research on the physiological basis of perception in different species.

**Bird Song: Behavior and Perception**

Birdsong has long been viewed is an excellent example of the value in applying all four of Tinbergen’s questions to the study of behavior (e.g., Hinde, 1982). Although we will not try to review the both proximate and ultimate aspects of birdsong in depth here, a good summary of the adaptive significance and evolution of birdsong can by found in Catchpole and Slater (1995).
**Definition and Function of Song**

Vocalizations in birds and other animals are often divided into two general categories: calls and songs. Calls usually refer to simpler vocalizations, produced by both sexes, that are used in contexts such as signaling alarm, maintaining flock cohesion, facilitating parent–young contact, and providing information about food sources. The term “song” refers to vocalizations that are generally more elaborate and used in the context of courtship and mating. The term “bird song” is limited to such vocalizations produced by species in the songbird order (e.g., Thorpe, 1961; Catchpole and Slater, 1995). For most of these species, “song” is usually the most complex vocalization produced; it is usually sung loudly from a prominent perch, and often associated with stereotyped courtship behaviors.

The decision to classify a vocalization as a song or a call is also related to the perceived function of the vocalization. The main functions ascribed to song behavior are territory defense (or spacing behavior) and mate attraction (Kroodsma and Byers, 1991). In many songbird species, especially among species that live in the temperate zone, there are marked sex differences in song behavior, with male song being more complex and more frequently sung than female song (Nottebohm, 1975). In evolutionary terms, the greater use of song by males is likely related to the effects of sexual selection, with both intrasexual (i.e., song is used to repel competing males from the territory) and intersexual selection (i.e., females choose males based on their songs) mechanisms operating differentially on males and females (e.g Searcy and Andersson, 1986). In the tropics, avian social systems differ to some extent from those in the temperate zone, and female
territorial defense along with female song is much more common than among temperate zone species (e.g. Levin, 1996).

**The Study of Song Perception**

One often stated similarity between bird song and human speech is the shared dependence of these two vocal communication signals on developmental learning. To produce functional species-typical vocalizations, both birds and humans must be exposed to ‘model’ vocalizations early in life. The so-called ‘critical period’ for song learning has been the topic of much research. Like any communication signal, however, the function of birdsong (in the adaptive sense) relies on accurate production and perception. Here we focus on the songbirds’ capacity to perceive and make sense of incoming natural acoustic information, i.e. song. What are the basic sensory capacities of the avian hearing system? What higher-level capacities exist in songbirds to process sounds as complex signals? Do songbirds have the capacity to form abstract relations among auditory stimuli? Do they form perceptual categories for their own species song as compared with the songs of other species? How is song processed as a functional signal? Given the range of adaptive functions served by song production, how are these different ‘meanings’ coded in the acoustics of song? And what are the neural mechanisms that underlie song processing?

**Song Perception and the Problem of Vocal Recognition**

In general, songbirds respond to sounds according to psychophysical principles familiar to humans. Virtually all species tested have an audiogram with a lower limit of
frequency sensitivity in the region of 100-200 Hz and a higher limit in the region from 8-10 kHz, with maximum sensitivity in the range of 1 to 3.5 kHz, depending on the species (Fay, 1988). Likewise, the temporal resolution of hearing is similar to that observed in humans, with gap detection (between sine tones) at around 2.5 ms (Klump and Maier, 1989; Okanoya and Dooling, 1990b). Interestingly, while little is known about the information coded in birdsongs at the millisecond timescale, recent data indicate that for birds the sensitivity to fine temporal structure in complex waveforms may exceed that for humans (Dooling et al., 2002). Like speech for humans, conspecific song for birds holds functional (i.e. adaptive) significance. Because selection pressures are likely to shape the physiological mechanisms that regulate song perception, the manner in which these acoustically complex signals are processed deserved special attention.

Variation in communication signals (i.e. songs) can occur in the spectro-temporal properties of the signal itself and also in the spatial-temporal distribution of signal sources. Together, this variation leads to at least two general classes of receiver behavior. First, because not all acoustic events are of equal interest, animals must be able to dissociate appropriate target signals from irrelevant/background noise, including non-target conspecific vocalizations. This problem has been studied in the context of the so-called cocktail party effect (Cherry and Taylor, 1954) and more generally in terms of auditory stream segregation (Bregman, 1990). Despite its obvious importance, relatively few studies have addressed this phenomenon in nonhumans (e.g. Fay, 1988; MacDougall-Shackleton et al., 1998). It may be that the acoustic parameters governing stream segregation of acoustic communication signals vary dramatically from those involved in the segregation of pure tone sequences (c.f. Hulse et al., 1997). In any case, the basic
ability is likely to be widespread. Recent reviews cover both stream segregation and the closely related topic of auditory spatial localization in nonhuman animals (Feng and Ratnam, 2000; Klump, 2000). Most research using conspecific communication signals assumes that the test subject has successfully extracted the target signal by presenting stimuli in isolation, or over a coincidently recorded low-noise background. This assumption implies an independence of localization/segregation processes and the subsequent classification behaviors that may not prove biologically realistic.

Once an auditory object is formed, a second general class of behavior emerges as these objects or events are organized into behaviorally relevant classes. For example, females might use male vocalizations to help choose a mate and therefore are likely, under appropriate conditions, to discriminate between heterospecific and conspecific songs or even the songs of different conspecific males. The presence of such distinctions, or class boundaries, implies the detection of and discrimination among multiple auditory objects along with an association between the object representations and a behavioral response. At the behavioral level, these processes are collectively referred to as perceptual recognition.

According to this definition, recognition can take many forms, depending on the specific boundaries between classes of vocalizations. Often these acoustic boundaries correspond to other behaviorally relevant distinctions (e.g., species, sex, kin, and individual). That is, they are not arbitrary but rather reflect the ecology of the particular animal under consideration. Although not all forms of recognition behavior are likely to be mediated by the same neural mechanisms, there are likely to be shared features across species, particularly when relevant classification requires discrimination among subsets
of conspecific vocalizations. The recognition of communication signals based on intraspecific acoustic variation is widespread and taxonomically diverse. For example, penguin chicks can recognize the calls of their parents (Jouventin et al., 1999; Aubin et al., 2000). Bottlenose dolphins have individually distinctive whistles that mothers and independent offspring can use to recognize one another (Sayigh et al., 1999). Female African elephants appear able to recognize the infrasonic calls of female family, bond group, and even more distant kin (McComb et al., 2000), and female spotted hyenas can recognize specific vocalizations of their own pups (Holekamp et al., 1999).

**Categorization.** From a cognitive standpoint, the common ability to associate sets of vocalizations with external referents raises the possibility that such animals are able to categorize the acoustic information in vocal signals in a more formal sense. For our purposes, categorization can take two forms.

*Phonetic Categorization.* First, we can ask if songbirds are capable of perceiving any features of their song categorically in the sense that humans perceive certain speech sounds categorically (Liberman et al., 1967). For example, in human speech perception there is a sudden switch from perceiving the phoneme "ba" to perceiving it as "pa" as voice onset time (VOT) increases gradually. On either side of the critical phoneme boundary, changes in VOT have little or no effect on perception of the phoneme (“ba” remains “ba”), but across the boundary, the effect of a small change in VOT is sudden and profound as “ba” suddenly switches perceptually to “pa” (Abramson and Lisker, 1967). Phonetic categorization as described has been found in a variety of mammals (Harnad, 1987), and is strongly suggested in at least one species of songbird. Nelson &
Marler (1989) found in a playback study that territorial swamp sparrows were relatively indifferent to initial swamp sparrow song syllables differing in duration until a certain critical duration was reached. At that critical duration, the song duration switched to that characteristic of a syllable with a different natural function, and the birds responded to it accordingly.

*Categorization as Concept Formation.* Apart from categorical perception in the unit-boundary sense just discussed, researchers have asked whether songbirds are capable of assigning broad classes of natural bird song stimuli to categories, such as some aspect of same-species versus other-species song, for example (Emlen, 1972; Falls, 1982; Marler, 1982). In this case, we are using the term category to define a case in which animals must place stimuli into different discriminable groups based on presumed common features, prototypes, or functions among exemplars within a group. The ability to form such categories is important for many reasons (Estes book; Murphy book). For example, if information can be sorted and stored in distinctive categories, there is a gain both in the efficiency with which old information can be retrieved for later use, and in the facility with which new information can be stored. In fact, Dooling et al. (1992) found that budgerigars, canaries, starlings, and zebra finches classified together song exemplars according to species, but distinguished song exemplars of their own species much better than exemplars of other species. The results provide evidence for special processing of ones’ own species song, a process akin to the special processing that is involved in human speech perception as compared with processing arbitrary acoustic stimuli (e.g., (Liberman, 1982; Kuhl, 1989; Mullennix and Pisoni, 1990). Similarly, budgerigars distinguish and classify human vowel sounds much as humans do (Dooling et al., 1987;
Brown et al., 1988). Taken together, these experiments suggest that certain aspects of human speech perception are not unique to humans and may reflect more general principles of perception and categorization.

Whether or not such principles are hardwired remains somewhat of an open question. Several subtypes of the Japanese macaque “coo” vocalizations can be defined on the basis of acoustic variation (Green 1975). Much of this work has focused on two particular subtypes of coo vocalizations, the smooth-early-high (SEH) and smooth-late-high (SLH), so-called because of the relative position of the peak of one frequency component that sweeps up and then down over the course of the call. Although Japanese macaque mothers can discriminate the coos of their young from others (Pereira 1986), the role of individual recognition cues in the coo vocalizations has not been well studied. Nevertheless, the SEH and SLH call types function in different behavioral contexts (Green 1975), and Japanese macaques possess a species-specific bias for discriminations involving these coos when the relevant variation is in the relative timing of the FM peak (Zoloth et al. 1979).

Although initial data suggested that the SEH and SLH calls can be perceived categorically (May et al. 1989), more recent data from the field indicate that many adult female coo vocalizations have FM peaks within the ambiguous zone between “early” and “late” prototypes (Owren and Casale 1994). The categorical boundaries determined in laboratory tests do not coincide with natural variation in the distribution of calls in the field. Nonetheless, all data to date show clear evidence that the coo calls are perceived as perceptually distinct classes (if not categorically), and several studies now substantiate the notion that the relative position of the FM peak within the coo is the most salient cue
to discrimination among different coos (May et al. 1988; Le Prell and Moody 2000). Amplitude cues also appear to function in call discrimination (Le Prell and Moody 1997). Thus, the stimulus dimensions involved in “real-world” classification and/or categorization of coo calls may involve a more dynamic stimulus space than that suggested by the original studies.

**The Songbird Model for Individual Vocal Recognition**

Whether or not songbirds categorize conspecific vocalizations in the strict sense, the occurrence of basic vocal recognition is common, and its evolutionary functions are well described, having been examined more extensively here than in any other group of animals. Various forms of intraspecific vocal recognition have been observed in nearly every species of songbird studied to date (see Falls 1982; Stoddard, 1996). In general, vocal recognition in songbirds provides for the association of specific songs with specific singers or locations, serving as a basis for decisions in more elaborate social behaviors such as female choice (Wiley et al., 1991; Lind et al., 1997), female preference (O'Loghlen and Beecher, 1997), and communal kin recognition (Beecher, 1991). Another complex social behavior in which individual vocal recognition plays an important role is territoriality, where it functions in both the manipulation and maintenance of territorial boundaries (e.g. Godard, 1991) and thus may have indirect effects on reproductive success (Hiebert et al., 1989).

**Song Recognition in Males**
**Field Studies.** The function of male song in maintaining and establishing songbird territories is well established. For example, abolishing a male songbird’s ability to sing has dramatic effects on his success at holding a territory (e.g. Peek, 1972), and simply broadcasting a conspecific song from an unoccupied territory leads to significantly lower rates of settlement in that territory compared to controls (Krebs et al., 1978). Moreover, territory residents often respond weakly, or not at all, to a neighbor singing from a familiar location but more strongly to a stranger singing from that same location. Using these facts along with a variety of clever song-playback techniques in the field, a large number of studies have demonstrated that males in many (at least 23) songbird species are capable of discriminating among neighbors and strangers on the basis of song alone. In several species, listeners are capable of recognizing individual singers on the basis of their songs (see Stoddard, 1996).

**Signal Variation.** There are several ways that singer identity could be represented in the acoustic variation of male birdsong. In the simplest case, individual males might sing a unique song or sets of songs (i.e., repertoires), and recognition would follow by the association of specific songs with specific singers. This strategy appears to be used by song sparrows (Beecher et al., 1994), and European starlings (Gentner and Hulse, 2000). One feature of this strategy is that recognition is constrained by memory capacity. Although recognition memory capacity deserves further study, initial results suggest that the capacity of these systems is, in fact, quite high (see Gentner et al., 2000) and, at least for song sparrows, exceeds the number of exemplars that an individual is likely to face at a single time in the wild (Stoddard et al., 1992).
A second recognition strategy relies on morphological differences in the acoustics of shared song types. In both white-throated sparrows and field sparrows, the songs of neighboring territorial males share several acoustic features but vary slightly in frequency. Neighbors rely on these subtle frequency differences to recognize one another (Brooks and Falls, 1975; Nelson, 1989). Related to this is a third possible strategy for recognition. If the morphology of the vocal-production apparatus varies slightly between individuals, then this variation might impart unique spectral features, or so-called “voice characteristics,” to all of an individual’s vocalizations. A potential role for voice characteristics has been suggested for great tits (Weary and Krebs, 1992), but vocal recognition in song sparrows (Beecher et al., 1994) and starlings (Gentner and Hulse, 2000) is not affected by these putative cues. Finally, vocal recognition might also rely on the sequence in which multiple song types are sung. That is, different males may share song types but sing them in individually distinctive temporal patterns. The role of this final cue has not been extensively studied in songbirds, but there is some evidence to suggest that European starlings are sensitive to the sequence of motifs within familiar song bouts (Gentner and Hulse, 1998).

For species in which males sing multiple songs, the four mechanisms outlined above may not be mutually exclusive. There is no a priori reason to believe that individual vocal recognition in a single species relies on acoustic variation coded in only a single dimension, nor is there any reason to suspect that all species of songbirds use the same recognition strategies. Given the approximately 4,500 different species of songbirds – each singing acoustically distinct songs and the occurrence of vocal recognition in a
wide range of behavioral contexts, it is likely that vocal-recognition information is coded at multiple levels throughout a songbird’s repertoire.

**Laboratory Studies.** Given the likely diversity of vocal-recognition behaviors across songbird species, it is reasonable to consider whether there are corresponding peripheral perceptual specializations among songbirds that in theory might provide an “open channel” of communication within a species while limiting confusion across species. For instance, different species might concentrate the acoustic energy of their songs in defined spectral bands. This hypothesis is supported by several observations of species-specific advantages during operant discriminations of multiple conspecific and heterospecific songs in several different species (Sinnott, 1980; Okanoya and Dooling, 1990a; Cynx and Nottebohm, 1992; Dooling et al., 1992). As mentioned earlier, however, the overwhelming data from psychophysical studies of hearing in birds indicate that most basic sensory processing capabilities (e.g., frequency sensitivity) are conserved across songbird species (Dooling et al., 2000). Thus, it appears that biases for the discrimination of species-specific vocalizations, and hence mechanisms for vocal recognition, result from evolutionary or ontogenic changes in the (presumably) central processing mechanisms that underlie pattern perception. This is consistent with the more general assumption that the perceptual (and cognitive) processes underlying individual vocal recognition take the neural representation of acoustically complex signals (i.e., song) as their input. Recent laboratory studies of European starlings have addressed these questions directly by determining more precisely the form of the acoustic signal controlling recognition in this species.
**Song Recognition in Starlings.** Male starlings present their songs in long episodes of continuous singing referred to as bouts (See Figure 1 for an excerpt of male starling song). Song bouts, in turn, are composed of much smaller acoustic units referred to as motifs (Adret-Hausberger and Jenkins, 1988; Eens et al., 1991), which in turn are composed of still smaller units called notes (Figure 1). Notes can be broadly classified by the presence of continuous energy in their spectrographic representations, and although several notes may occur in a given motif, their pattern is usually highly stereotyped between successive renditions of the same motif. One can thus consider starling song as a sequence of motifs, where each motif is an acoustically complex event. The number of unique motifs that a male starling can sing (i.e., his repertoire size) can be quite large, and consequently different song bouts from the same male are not necessarily composed of the same set of motifs. This broad acoustical variation in their song provides several potential cues that starlings might use when learning to recognize the songs of an individual conspecific and while maintaining that recognition over time. One straightforward recognition mechanism is the association of specific motifs with specific singers. Although some sharing of motifs does occur among captive males (Hausberger and Cousillas, 1995; Hausberger, 1997), the motif repertoires of different males living in the wild are generally unique (e.g. Eens et al., 1989; Chaiken et al., 1993). Thus, learning which males sing which motifs can provide discriminative cues for song classification.

As shown in Figure 2, data from operant studies in starlings indicate that song recognition is based at the level of the motif. Starlings trained operantly to recognize individual conspecifics by one set of songs can readily generalize correct recognition to novel songs from the same singers (Gentner and Hulse, 1998). However, recognition falls...
to chance when these novel song bouts have no motifs in common with the training songs (Gentner et al., 2000). Likewise, starlings trained to discriminate among pairs of motifs will reverse the discrimination when transferred to the same motif sung by the opposite individual and perform at chance when transferred to novel motifs sung by the training singers. This failure to generalize correct recognition to songs composed of novel motifs, or to single novel motifs, is inconsistent with the use of individually invariant source and/or filter properties (voice characteristics) for vocal recognition. Instead, the data suggest that starlings learn to recognize the songs of individual conspecifics by attending to information contained at (or below) the level of the motif. They appear to associate distinct sets of motifs (or variant motif features) with individual singers.

If starlings learn to recognize individuals by the sets of unique motifs that they sing, then once learned, it should be possible to control recognition systematically by varying the proportions of motifs in a given bout that come from two “vocally familiar” males. That is, recognition behavior ought to follow the proportional distribution of motifs from two vocally familiar males rather than the presence or absence of single diagnostic motifs from either male. The behavioral data confirm this prediction by showing that when starlings are compelled to classify conspecific songs, they do so by memorizing large numbers of unique song components (i.e., motifs) and then organizing subsets of these motifs into separate classes (Gentner and Hulse, 2000). As a cognitive recognition strategy, classifying songs according to their component (motif) structure represents a straightforward method of dealing with these complex acoustic signals. Because individual starlings tend to possess unique motif repertoires, disjoint sets of motifs will generally correspond to individual identity. Therefore, attending to the motif
structure captures a significant portion of the individual variation in the signal, albeit at the expense of a large memory capacity.

The behavioral data yield several hypotheses regarding the neural mechanisms underlying the recognition of natural (i.e. high-dimensional) acoustic events. First, the functionality of motifs as auditory objects in recognition behavior implies their explicit representation in the central nervous system. That is, the response functions of single neurons or of populations of neurons in appropriate forebrain auditory regions should reflect the segmentation of song at the level of the motif. Second, because recognition behavior requires the learned association between sets of motifs and singers, motif representations (or the representations of sub-motif features that correspond to unique motifs) should reflect the behavioral relevance of specific motifs. That is, there should be a bias for representations of familiar motifs. Third, the representational mechanisms and capacity (i.e. memory) of the system should permit the acquisition of very large numbers of acoustically complex, natural objects (motifs). The nature of object representation, representational plasticity, and memory capacity are central questions for any researcher interested in the neural coding of natural stimuli, including human speech and language. Below, we consider these questions in the context of the neural representation of bird song.

**Neural Representations of Bird Song**

Bird’s Own Song

Research on the neural mechanisms of bird song has followed the ethological studies in emphasizing strongly male song learning and production. All songbirds possess
a network of cytoarchitectonically distinct brain regions, referred to collectively as the
“song system” (see Figure 3). Together, the various regions of the song system are
thought to function primarily in the production and juvenile acquisition of song. The song
system itself is composed of two sub-systems, the vocal-motor pathway (VMP), which
plays a direct role in the adult song production; and the anterior-forebrain pathway
(AFP), which is thought to function primarily in vocal acquisition among juveniles, and
in subsequent maintenance of song in adults. A number of comprehensive reviews on
song-system neurobiology are available (e.g., Nottebohm, 1996; Brenowitz et al., 1997;
Brainard and Doupe, 2002). One hallmark of cells within many of the song system nuclei
is their selective response to a “bird’s own song”, or BOS (see Margoliash 1987). That is,
one readily finds neurons throughout the song system whose firing rates and/or temporal
response properties are ‘tuned’ to the acoustics of the song that the bird sings. In the
nucleus HVc (see Fig.3), the stimulus specificity observed for BOS has both spectral and
temporal components, with responses contingent upon the presence (and absence) of
acoustic energy in specific frequency bands or upon specific temporal combinations of
sounds (Margoliash, 1983; Margoliash and Fortune, 1992). In addition, these so-called
‘BOS responses’ are strongly modulated by behavioral state (Dave et al 1998; Schmidt
and Konishi, 1999), and are, in at least some cases, observed in cells that also show
strongly coupled pre-motor activity during song production (Yu and Margoliash 1996;
Dave and Margoliash 2000). The sensory-motor integration at both the cellular- and
system-level which gives rise to the BOS response is an active area of research among
birds: song neuroethologists. Understanding these physiological mechanisms will likely
impact the broader study of sensorimotor learning in other systems, and may be of value
in understanding the perceptual role of self-generated sounds in human speech processing (see Margoliash 2003).

Despite early suggestions that BOS selective responses in song system nuclei, specifically the VMP, might reflect a ‘motor theory’ of song perception (*sensu* Liberman et al. 1967), recent data suggest a somewhat different interpretation. The state dependent nature of these responses, and that absence of auditory responses altogether in many VMP neurons in the awake animal (Dave et al. 1998), argue against the notion that the pathways controlling vocal output also contribute to sensory representations of song. Instead, BOS selectivity more likely reflects the involvement of acoustic feedback in ongoing regulation of song production mechanisms. It now appears that only very specific sorts of auditory information, namely BOS, are admitted to the song system so that the bird can detect, and thus correct, any deviations between the intended song and that actually produced. Nonetheless, there are data that suggest a production-independent role for both HVC and the AFP in adult song perception. Lesions to lMAN in canaries affect auditory, but not visual discrimination (Burt et al., 2000); lesions to HVC in female canaries abolish female behavioral preferences for conspecific over heterospecific song (Brenowitz, 1991) and for sexually attractive song phrases over other phrases of conspecific song (Del Negro et al., 1998; but see MacDougall-Shackleton et al., 1998). In both male and female starlings, HVC lesions affect the ability to form new associations with familiar songs while leaving retention of learned conspecific song discriminations intact (Gentner et al., 2000). Clearly the role of these structures, and by extension the general role of vocal-motor systems in sensory perception and cognition requires further study- ideally through electrophysiology in awake animals.
Auditory System

Within the auditory system proper (See Fig. 3) the links between physiological responses and vocal perception are more direct. Auditory signals impinging on the bird's ear (following the general vertebrate pathway) are coded tonotopically in the cochlea, continue through the thalamus, and into Field L2 - the avian analog to mammalian primary auditory cortex. As one moves ‘higher’ into the central structures, away from the primary cortex, increasing selectivity for complex stimuli and species-specific vocalizations emerges (e.g. Leppelsack and Vogt, 1976; Theunissen and Doupe, 1998). The pattern of increasing response selectivity from Field L2 to the higher-order areas continues into NCM and cHV (Müller and Leppelsack, 1985; Sen et al., 2001) suggesting that these regions are involved in the extraction of complex features common to song.

Neurons in NCM (see Fig. 3) are broadly tuned to conspecific songs (Grace et al., 2003) and respond to repeated presentations of a single conspecific song in a stimulus-specific manner. The repeated presentation of a single conspecific song elicits a rapid modulation in the initial firing rate of NCM neurons (Chew et al., 1995; Stripling et al., 1997). If the same song is repeated on the order of 200 times, this initial modulation of the firing rate is no longer observed when that same song is presented on subsequent trials, even though the initial response modulation can still be observed for other conspecific songs. These stimulus-specific changes in the response properties of NCM neurons have led to the hypothesis that NCM may play an important role in individual vocal recognition. The putative role in song memory processes is supported by the fact that many neurons in NCM (and cHV) show a rapid and selective up-regulation of the
immediate early gene (IEG) *zenk* in response to the presentation of conspecific songs (Mello et al., 1992), in manner sensitive to the acoustics of particular song syllables (Ribeiro et al., 1998). Interestingly, these IEG responses also habituate to the repeated presentation of the same conspecific song (Mello et al., 1995) and are elevated during specific components of vocal-recognition in starlings (Gentner et al., 2001). In mammals, the homolog to *zenk* is required for expression of certain forms of LTP and the consolidation of long-term memories in mice (Jones et al., 2001). This suggests that *zenk* expression in NCM and cHV may be related to learning about conspecific songs.

Recent data demonstrates directly the role of cmHV in learned recognition of song. After training starlings to recognize two sets of conspecific songs, Gentner and Margoliash (2003) observed that single neurons and populations of neurons in the medial cHV respond selectively to acoustic features contained in those songs that the birds had learned to recognize. In contrast, no neurons were selective for similar features in songs that were novel to the birds. This argues very strongly that the response functions of cmHV, at both the single unit and the population level, are a direct product of each bird’s unique sensory experience. Mechanisms of experience-dependent plasticity act to modify the responses of cmHV neurons based on the functional demands of song recognition. Several additional results from this study are consistent with this notion. First, the spectro-temporal tuning properties of cmHV cells correspond closely to song features correlated with individual motifs. That is, the same auditory objects that control recognition behavior, also predict the responses of selective cells in cmHV. Second, the variation in neuronal response strength among the familiar songs was dependent upon the reinforcement contingencies used for recognition training. For animals trained with a
go/nogo procedure to discriminate between two sets of songs, the S+ songs elicited the significantly stronger responses than S- songs, which in turn elicited significantly stronger responses than novel songs. When positive reinforcement is available for both sets of songs, the response strengths associated with each set of familiar songs are similar, but still greater than those associated with novel songs. Thus, the response profiles of neurons in cmHV are shaped not only by task relevant acoustic features of conspecific songs in a ‘bottom-up’ fashion, but also by so-called ‘top-down’ mechanisms presumably through reward systems (Gentner and Margoliash, 2003).

Conclusions

In the absence of an efficient coding scheme, the high-dimensionality and range of natural stimuli could easily overwhelm the capacity and fidelity of any representational system. In songbirds coding efficiencies are improved by two mechanisms. First, the song is parsed into perceptual objects (i.e. motifs) that capture the behaviorally relevant variation among very complex signals (songs). Second, the representation of these complex auditory objects is dynamic, and is dependent upon the functional demands of the task at hand. At the present time, it unclear whether or not these dynamic, functional object representations arise from the convergence of more static representations of embedded, simpler song features, or more general representations of the input signal. Regardless, the songbird data raise the interesting hypothesis that similarly well-defined and dynamic representations of functional speech units may exist in humans,
corresponding to perhaps single phonemes or high-probability phonemic patterns relevant in specific contexts.

To be sure, human language processing, and thus by necessity the processing of speech, contains innumerable aspects that far exceed the perceptual and cognitive abilities of any other species. Human language is inarguably special. Still, like any behavior it is constrained, and many of the constraints on speech are shared with other organisms that rely on vocal communication. We have tried to point out some of those constraints. While the mechanisms that regulate any one biological system should never be expected to map one-to-one onto another system, we gain knowledge by understanding both the differences and the similarities.
REFERENCES


Cherry E, Taylor W (1954) Some further experiments upon the recognition of speech, with one and with two ears. The Journal of the Acoustical Society of America 26:554-559.


Kuhl PK, Miller JD (1978). Speech perception by the chinchilla: identification function


Figure 1. Sonogram of starling song segment. Power across the frequency spectrum is shown as a function of time. Darker regions show higher power. Starling song bouts are organized hierarchically. Normal bouts of song can last over a minute, and are comprised of series of repeated motifs. (A) Shows a short sequence of motifs as they might appear in a much longer song bout. A single motif is outlined in (B). Complete song bouts contain many different motifs. Motifs are comprised of stereotyped note patterns. An example of one note is shown in (C).

Figure 2. Vocal recognition behavior in European starlings. (A) Mean (±SEM) proportion of correct responses given during asymptotic performance on an operant recognition task (“Baseline”), and during initial transfer to novel songs containing familiar motifs (“Novel bout”). (B and C) Mean (±SEM) proportion of correct responses during transfer from the baseline training to novel songs from the same singers composed of “Novel motifs”. Data in “B” show the transfer when the subjects were exposed to the training and test songs outside of the operant apparatus, whereas the data in “C” show the results for the same transfer after controlling for this experience. Note that in “C” recognition of the novel motifs falls below chance. (D) Data showing the close (and approximately linear) relationship between the statistical distribution of familiar motifs from two different singers and song recognition.

Figure 3. Schematic of the songbird primary auditory pathways (light gray), vocal motor pathway (VMP, gray), and anterior forebrain pathway (AFP, dark gray). Ov: nucleus ovoidalis; L1-L3: field L complex; NCM: caudomedial neostriatum; cHV: caudoventral
hyperstriatum; NIIf: nucleus interfacialis; Uva: nucleus uvaeformis; RA: robust nucleus of
the archistriatum; lMAN: lateral magnocellular nucleus of the anterior neostriatum;
DLM: dorsolateral region of the medial thalamus; X: area X; HVc is used as the proper
name. Ov comprises a core and a surround. HVc and RA have subjacent regions, the
“shelf” and “cup”, respectively.