2 • Birdsong and human speech: common themes and mechanisms

Allison J. Doupe and Patricia K. Kuhl

INTRODUCTION

Experts in the fields of human speech and birdsong have often commented on the parallels between the two in terms of communication and its development (Marler, 1970a; Kuhl, 1989). Does the acquisition of song in birds provide insights regarding learning of speech in humans? This review provides a critical assessment of the hypothesis, examining whether the similarities between the two fields go beyond superficial analogy. The often cited commonalities provide the topics of comparison that structure this review.

First, learning is critical to both birdsong and speech. Birds do not learn to sing normally, nor infants to speak, if they are not exposed to the communicative signals of adults of the species. This is an exception among species: most animals do not have to be exposed to the communicative signals of their species to be able to reproduce them. The fact that babies and songbirds share this requirement has intrigued scientists.

Second, vocal learning requires both perception of sound and the capacity to produce sound. At birth, both human infants and songbirds have been hypothesized to have innate perceptual predispositions for the vocal behavior of their own species. We review the nature of the predispositions in the two cases and the issue of whether they are similar. Given that innate predispositions exist, another important question is how subsequent experience alters perception and production in each case. Moreover, vocal perception and production are tightly interwoven in the vocal learning process. We examine what is known about the relationship between perception and production an whether in these different vocal learners it is similar.

In addition, neural substrates of vocal communication in humans and birds have often been compared Human brains are asymmetric and language tend to be organized in the left hemisphere as opposed t the right. Birds are also often assumed to show simila hemispheric specialization for song. What are the rea parallels between the neural substrates in the two cases

Finally, critical (sensitive) periods are evidenced i both species. Neither birds nor babies appear to lear their communicative signals equally well at all phases o the life cycle. This raises the questions of what cause the change in the ability to learn over time and wi experience, and whether the causes are the same i human infants and songbirds. And if the plasticity c the brain is altered over the life cycle, what neur mechanisms control this changing ability to learn?

The research reviewed here relates to ongoin work in developmental biology, ethology, linguistic cognitive psychology, and computer science, as we as in neuroscience, and should be of interest to indivduals in many of these fields. What our review reveal is that although the comparisons between birdson and speech are not simple, there is a surprisingl large number of areas where it is fruitful to compar the two. Going beyond the superficial analogy, however, requires some caveats about what may be compaable and what clearly is not. In the end, understandin both the similarities and differences will provide broader spectrum in which to view the acquisition communication in animals and humans.

Editors' note: Adapted and reprinted, with permission from the authors and from the *Annual Review of Neuroscience*, Volume 22, © 1999 by Annual Reviews www.annualreviews.org. The sections on "Specialized neural substrates for song and speech learning (pp. 596–604) on "Lateralization" (pp. 606–609) and on "possible neural mechanisms underlying the sensitive period and its clsure" (pp. 618–619) have been omitted. The reader is referred relevant chapters in this volume for reviews of more recent wo on neural mechanisms.

Neuroscience of Birdsong, ed. H. Philip Zeigler and Peter Marler. Published by Cambridge University Press. © Cambridge University Press 2008

SPEECH AND BIRDSONG: DEFINITIONS

Speech and song production

Both birdsong and human speech are complex acoustic signals. Figure 2.1 shows a spectrographic (frequency versus time) display of a spoken human phrase ("Did you hit it to Tom?") and Figure 2.2 a similar display of songs of two different songbird species. In both songbirds and humans, these sounds are produced by the flow of air during expiration through a vocal system. In humans, the process is relatively well understood: air from expiration generates a complex waveform at the vocal folds, and the components of this waveform are subsequently modified by the rest of the vocal tract (including the mouth, tongue, teeth, and lips) (Stevens, 1994).

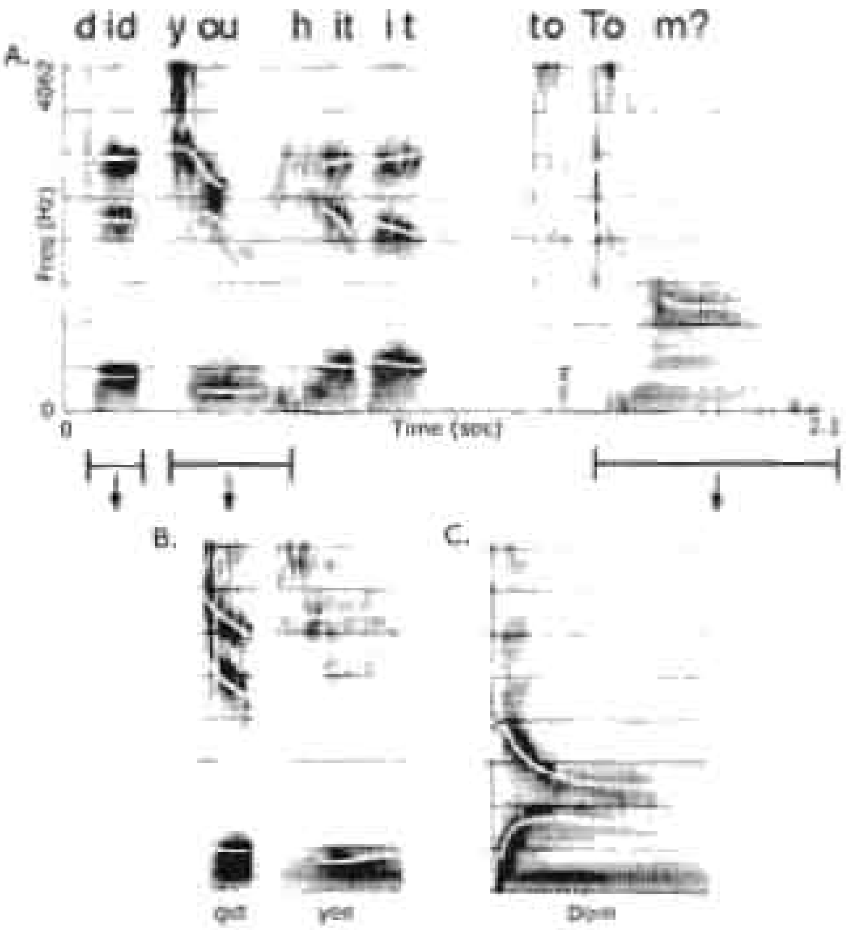

Figure 2.1 Human speech. Three dimensions of speech are shown in a spectrogram: time or duration along the horizontal axis; frequency along the vertical axis; and intensity, which is correlated with loudness, by the relative darkness of each frequency. This spectrogram shows the phrase "Did you hit it to Tom?" spoken by a female (A). White lines are the formants that characterize each individual phoneme. (B–C) Variations on words from the full sentence. (B) A place of articulation contrast using a spectrogram of the nonsense word "gid," which differs from its rhyme "did" (in A) in that it has a decreasing frequency sweep in the second and third formants (between 2000 and 3000 Hz). This decreasing formant pattern defines the sound "g" and a pattern of flat formants defines the sound "d." (C) The words "Tom" and "Dom" contrast in voice onset time (VOT). Notice the long, noisy gap in "Tom" (A), which has a long VOT, compared with the short gap in "Dom."

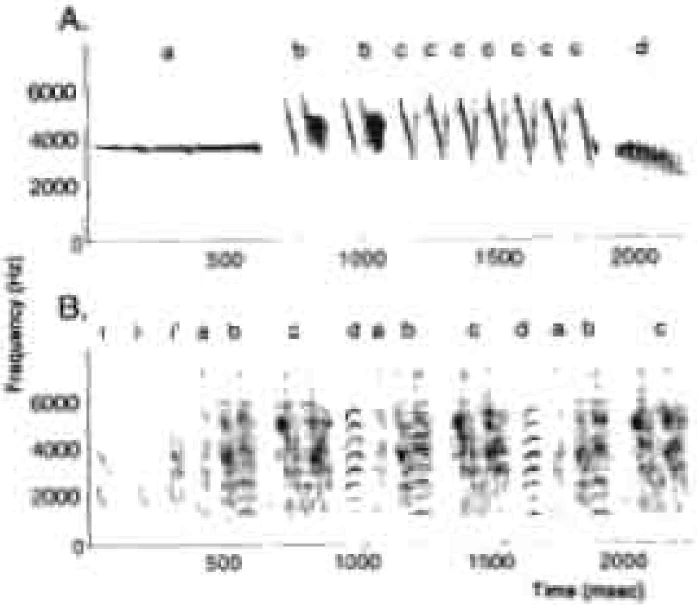

Figure 2.2 Examples of birdsongs from two species. (A) A typical song of a white-crowned sparrow. The smallest elements, the notes, are combined to form syllables (lower case letters), and these are repeated to form phrases. White-crowned sparrow songs typically begin with (a) a long whistle followed by (b, c) trills and (d) buzzes. (B) A typical song of a zebra finch. Note the noisy spectral quality (more like humans) that distinguishes it from more tonal species like the sparrows. Zebra finch songs start with a number of introductory syllables (marked with i), followed by a sequence of syllables (lower case letters), that can be either simple or more complex, with multiple notes (e.g. b, c). Particular sequences of syllables are organized into phrases called motifs, which are repeated.

The vocal tract acts as a filter, creating concentrations of energy at particular frequencies, called formant frequencies (Figure 2.1). Vowels are characterized by relatively constant formant frequencies over time (Figure 2.1A, C), whereas during consonant production the formant frequencies change rapidly (20–100 ms), resulting in formant transitions (Figure 2.1A, B, D).

In songbirds, sounds are produced by the flow of air during expiration through an organ called the syrinx, a bilateral structure surrounded by specialized muscles, which sits at the junction of the bronchi with the trachea. A number of aspects of syringeal function are understood, although the exact mechanism of sound generation is controversial and is under active investigation (Gaunt, 1987; Goller and Larsen, 1997a; Suthers, 1997; Fee *et al.*, 1998). Also, there are indications that the upper vocal tract in bird structures sound in a manner like the upper vocal tract in humans. Recent research suggests that the width of beak opening (known as beak gape) affects sound frequency (Westneat *et al.*, 1993; Suthers, 1997), and there may be some degree of coupling between the syrinx and the vocal tract (Nowicki, 1987). Regardless of differences in component structures, for both birdsong and speech the production of highly structured and rapidly changing vocalizations requires elaborate neural control and coordination of respiration with a variety of vocal motor structures.

The structure of speech and song

It is useful to define the basic terms used in each field, and the various ways in which vocal behavior is described, in order to assess what aspects of each of the signals are comparable. Human speech can be described at many different levels. It can be written, spoken, or signed (using a manual language such as American Sign Language). In all these forms, language consists of a string of words ordered by the rules of grammar to convey meaning. Structurally, language can be analyzed from the standpoint of semantics (conceptual representation), syntax (word order), prosody (the pitch, rhythm, and tempo of an utterance), the lexicon (words), or phonology (the elementary building blocks, phonemes, that are combined to make up words).

Speech, and especially its development, has been intensively studied at the phonological level. Phonetic units are the smallest elements that can alter the meaning of a word in any language, for example the difference between /r/ and /l/ in the words "rid" and "lid" in American English. Phonemes refer to the phonetic units critical for meaning in a particular language. The phonetic difference between /r/ and /l/ is phonemic in English, for example, but not in Japanese. Each phonetic unit can be described as a bundle of phonetic features that indicate the manner in which the sound was produced and the place in the mouth where the articulators (tongue, lips, teeth) were placed to create the sound (Jakobson *et al.*, 1969). The acoustic cues that signal phonetic units have been well documented and include both spectral and temporal features of sound (Figure 2.1) (Stevens, 1994). For instance, the distinction between /d/ and /g/ depends primarily on the frequency content of the initial burst in energy at the beginning of the sound and the direction of formant transition change (Figure 2.1A, B). An example of a temporal acoustic dimension of speech is voice-onset time (VOT), which refers to the timing of periodic laryngeal vibration (voicing) in relation to the beginning of the syllable (Figure 2.1A, D). This timing difference provides the critical cue used to identify whether a speech sound is voiced or voiceless (e.g. /b/ versus /p/, /do/ versus /to/) and is a classic distinction used in many speech studies.

Which aspects of birdsong can be usefully compared with speech? Birdsongs are distinct from bird calls (which are brief and generally not learned), last from a few seconds to many tens of seconds, and, like speech, consist of ordered strings of sounds separated by brief silent intervals (Figure 2.2). The smallest level of song usually identified is the note or element, defined as a continuous marking on a sound spectrogram; these may be analogous to the smallest units of speech, or phonetic units. Notes can be grouped together to form syllables, which are units of sound separated by silent intervals. When singing birds are interrupted by an abrupt light flash or sound, they complete the syllable before stopping (Cynx, 1990); thus, syllables may represent a basic processing unit in birdsong, as posited for speech.

Another feature that birdsong and language share is the conspicuous timing and ordering of components on a timescale longer than that of the syllable. Song syllables are usually grouped together to form phrases or motifs (Figure 2.2), which can be a series of identical or different syllables. Many songbirds sing several phrases in a fixed order as a unit, which constitutes the song, whereas other species such as mockingbirds and warblers produce groups of syllables in fixed or variable sequences. The timing and sequencing of syllables and phrases are rarely random but instead follow a set of rules particular to a species. In the songbird literature, the ordering of syllables and phrases in song is often called song syntax. The same word applied to human speech, however, implies grammar, i.e. rules of ordering words from various grammatical classes to convey meaning. Therefore, in this review, we avoid using the word syntax for song and simply use "order." Thus, language and song share a dependence on timing on several timescales: a shorter timescale (on the order of tens of milliseconds), as in phonemes and syllables, and a longer one, up to many hundreds of milliseconds (as in syllable, phrase, and word ordering).

Language is also characterized by a boundless and flexible capacity to convey meaning, but this property is not shared with birdsong. The whole set of different songs of a bird is known as its song repertoire and can vary from one (in species such as the zebra finch or white-crowned sparrow) to several hundreds (for review see Konishi, 1985). Numerous behavioral studies, usually using the receiver's response, suggest that songs communicate species and individual identity (including "neighbor" and "stranger"), an advertisement for mating, ownership of territory, and fitness. Some birds with multiple song types use different songs for territorial advertisement and for mate attraction (Catchpole, 1983; Searcy and Nowicki, 1998). Nonetheless, large song repertoires do not seem to convey many different meanings, nor does song have the complex semantics of human speech. The definitions above suggest that the phonology (sound structure), the rules for ordering sounds, and perhaps the prosody (in the sense that it involves control of frequency, timing, and amplitude) are the levels at which birdsong can be most usefully compared with language, and more specifically with spoken speech, and are thus the focus of this review.

VOCAL LEARNING IN HUMANS AND SONGBIRDS

Which animals are vocal learners?

Many animals produce complex communication sounds but few of them can and must learn these vocal signals. Humans are consummate vocal learners. Although there is emerging evidence that social factors can influence acoustic variability among nonhuman primates (Sugiura, 1998), no other primates have yet been shown to learn their vocalizations. Among the mammals, cetaceans are well known to acquire their vocal repertoire and to show vocal mimicry (McCowan and Reiss, 1997); there are also some bats whose vocalizations may be learned (Boughman, 1998). Among avian species, songbirds, the parrot family, and some hummingbirds meet the criteria for vocal learning, but the term birdsong is usually reserved for the vocalizations of passerine (perching) songbirds and that is the focus of this review. The many thousands of songbird species, as well as the parrots and hummingbirds, stand in striking contrast to the paucity of mammalian vocal learners.

Nonhuman primates can, however, make meaningful use of vocalizations: for instance, vervets use different calls to indicate different categories of predators. Production of these calls is relatively normal even in young vervets and does not appear to go through a period of gradual vocal development, but these animals must develop the correct associations of calls to predators during early ontogeny (Seyfarth and Cheney, 1997). What songbirds and humans share is not this development of associations of vocalizations with objects or actions, but the basic experience-dependent memorization of sensory inputs and the shaping of vocal outputs.

Evidence for vocal learning

The basic phenomenology of learning of song or speech is strikingly similar in songbirds and humans. Initial vocalizations are immature and unlike those of adults: babies babble, producing consonant–vowel syllables that are strung together (e.g. bababa or mamama), and young songbirds produce subsong, soft and rambling strings of sound. Early sounds are then gradually molded to resemble adult vocalizations. The result of this vocal development is that adults produce a stereotyped repertoire of acoustic elements: these are relatively fixed for a given individual, but they vary between individuals and groups (as in languages and dialects, and the individually distinct songs and dialects of songbirds within a particular species). This variability is a reflection of the fact that vocal production by individuals is limited to a subset of all sounds that can be produced by that species. Layered on top of the developing capacity to produce particular acoustic elements is the development of sequencing of these elements: for humans this means ordering sounds to create words and, at a higher level, sentences and grammar; in birds this means sequencing of elements and phrases of song in the appropriate order. An important difference to remember when making comparisons is that the numerous languages of humans are not equivalent to the songs of different species, but rather to the individual and geographical variations of songs within a species.

LEARNED DIFFERENCES IN VOCAL BEHAVIOR

That the development of a mature vocal repertoire reflects learning rather than simply the expression of innate programs is apparent from a number of observations. Most important, for both birds and humans, there exist group differences in vocal production that clearly depend on experience. Obviously, people learn the language to which they are exposed. Moreover, even within a specific language, dialects can identify the specific region of the country in which a person was raised. Likewise, songbirds learn the songs sung by adults to which they are exposed during development: this can be clearly demonstrated by showing that birds taken from the wild as eggs or nestlings and exposed to unrelated conspecific adults, or even simply to tape recordings of the song of these adults, ultimately produce normal songs that match those that were heard (Marler, 1970b; Thorpe, 1958, 1961). Even more compelling are cross-fostering experiments, in which birds of one species being raised by another will learn the song, or aspects thereof, of the fostering species (Immelmann, 1969). In addition, many songbirds have song "dialects," particular constellations of acoustic features that are well defined and restricted to local geographic areas. Just as with human dialects, these song dialects are culturally transmitted (Marler and Tamura, 1962).

VOCALIZATIONS IN THE ABSENCE OF EXPOSURE TO OTHERS

Another line of evidence supporting vocal learning is the development of abnormal vocalizations when humans or birds with normal hearing are socially isolated and therefore not exposed to the vocalizations of others. The need for auditory experience of others in humans is evident in the (fortunately rare) studies of children raised either in abnormal social settings, as in the case of the California girl, Genie, who was raised with almost no social contact (Fromkin *et al.*, 1974), or in cases in which abandoned children were raised quite literally in the wild (Lane, 1976). These and other documented instances in which infants with normal hearing were not exposed to human speech provide dramatic evidence that in the absence of hearing speech from others, speech does not develop normally. Similarly, songbirds collected as nestlings and raised in isolation from adult song produce very abnormal songs (called "isolate" songs) (Marler, 1970b; Thorpe, 1958). This need for early auditory tutoring has been demonstrated in a wide variety of songbirds (for reviews see Catchpole and Slater, 1995; Kroodsma and Miller, 1996). Strikingly, although isolate songs are simplified compared with normal, learned song, they still show some features of species-specific song (Marler and Sherman, 1985).

One caveat about studies of isolated songbirds or humans is that many aspects of development are altered or delayed in such abnormal rearing conditions. Nonetheless, the results of isolation in humans and songbirds are in striking contrast to those seen with members of closely related species, such as nonhuman primates and nonsongbirds such as chickens, in whom vocalizations develop relatively normally even when animals are raised in complete acoustic isolation (Konishi, 1963; Kroodsma, 1985; Seyfarth and Cheney, 1997). In combination with the potent effects of particular acoustic inputs on the type of vocal output produced, these results demonstrate how critically both birdsong and speech learning depend on the auditory experience provided by hearing others vocalize.

THE IMPORTANCE OF AUDITION IN SPEECH AND SONG

The importance of hearing one's own vocalizations

Vocal learning, shared with few other animals, is also evident in the fact that both humans and songbirds are acutely dependent on the ability to hear themselves in order to develop normal vocalizations. Human infants born congenitally deaf do not acquire spoken language, although they will, of course, learn a natural sign language if exposed to it (Petitto, 1993). Deaf infants show abnormalities very early in babbling, which is an important milestone of early language acquisition. At about 7 months of age, typically developing infants across all cultures will produce this form of speech. The babbling of deaf infants, however, is maturationally delayed and lacks the temporal structure and the full range of consonant sounds of normal-hearing infants (Oller and Eilers, 1988; Stoel-Gammon and Otomo, 1986). The strong dependence of speech on hearing early in life contrasts with that of humans who become deaf as adults: their speech shows gradual deterioration but is well preserved relative to that of deaf children (Cowie and Douglas-Cowie, 1992; Waldstein, 1990).

Songbirds are also critically dependent on hearing early in life for successful vocal learning. Although birds other than songbirds, e.g. chickens, produce normal vocalizations even when deafened as juveniles, songbirds must be able to hear themselves in order to develop normal song (Konishi, 1963, 1965b; Nottebohm, 1968). Songbirds still sing when deafened young, but they produce a very abnormal, indistinct series of sounds that are much less songlike than are isolate songs; although it varies from species to species, often only a few features of normal songs are maintained, primarily their approximate duration (Marler and Sherman, 1983). As with humans, once adult vocalizations have stabilized, most songbird species show decreased dependence on hearing (Konishi, 1965b; but see below).

The effects of deafness in early life do not differentiate between the need for hearing others and a requirement for hearing oneself while learning to vocalize. In birds, however, there is often a separation between the period of hearing adult song and the onset of vocalizations, and this provided the opportunity to demonstrate that song is abnormal in birds even

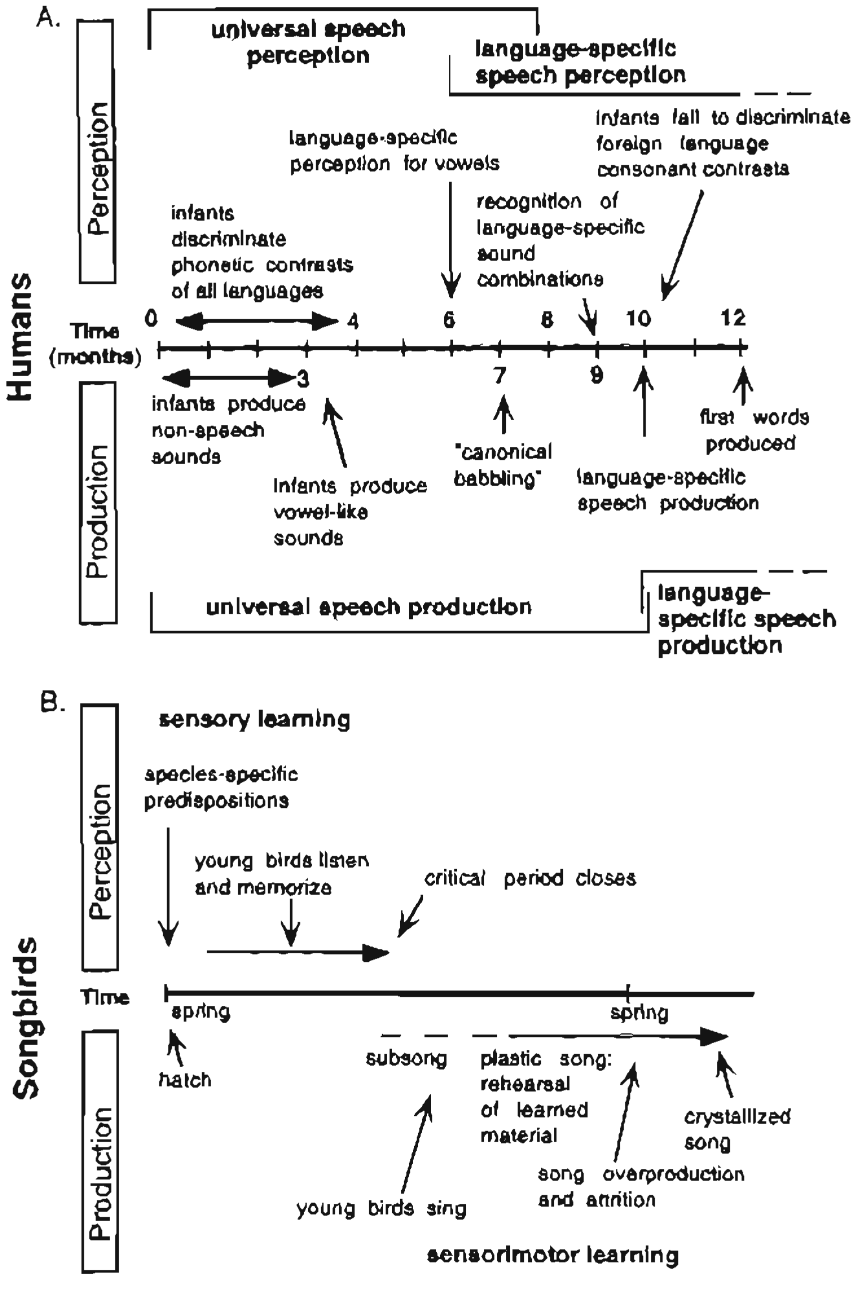
A.
Humans
Perception
universal speech perception
language-specific speech perception
language-specific perception for vowels
infants fail to discriminate foreign language consonant contrasts
infants discriminate phonetic contrasts of all languages
recognition of language-specific sound combinations
Time (months)
0
4
6
8
10
12
3
7
9
Production
infants produce non-speech sounds
infants produce vowel-like sounds
"canonical babbling"
language-specific speech production
first words produced
universal speech production
language-specific speech production
B.
Songbirds
Perception
sensory learning
species-specific predispositions
young birds listen and memorize
critical period closes
Time
spring
spring
Production
hatch
subsong
plastic song: rehearsal of learned material
crystallized song
song overproduction and attrition
young birds sing
sensorimotor learning

Figure 2.3 Timelines of speech and song learning. (A) During the first year of life, infant perception and production of speech sounds go through marked changes. (A, top) The developmental milestones associated with listening to speech; (A, bottom) the type of sounds produced throughout an infant's first year, leading up to the meaningful production of words. In both aspects of development, infants change from being language-general in the earliest months to language-specific toward the end of the first year. (B) Similar timelines show the early perceptual learning of seasonal songbirds (approximately 2–3 months), followed by sensorimotor learning in the fall and especially the next spring. In zebra finches this entire learning takes place over 3–4 months, with the critical period ending around 60 days of age, and much more overlap between sensory and sensorimotor phases (with singing beginning around 30 days of age).

communication). Thus, although auditory feedback is not as essential for ongoing vocal production in adult birds and humans as in their young, it clearly has access to the adult vocal system, and can have dramatic effects on vocal behavior if it is not well matched with vocal output.

INNATE PREDISPOSITIONS AND PERCEPTUAL LEARNING

Key features of vocal learning are the perception of sounds, the production of sounds, and the (crucial) ability to relate the two. In the next section, two questions, which roughly parallel the course of vocal development and have preoccupied both speech and song scientists, are addressed. What are the perceptual capabilities and innate predispositions of vocal learners at the start of learning? And what does subsequent experience do to perception?

Speech and song perception and production: innate predispositions

Experience clearly affects vocal production in humans and songbirds, but there is compelling evidence that learning in both species does not occur on a *tabula rasa*. Rather, there is evidence of constraints and predispositions that bias the organism in ways that assist vocal learning.

At the most fundamental level, the physical apparatus for vocalization constrains the range of vocalizations that can be produced (Podos, 1996). What is surprising, however, is that motor constraints do not provide the strongest limitations on learning. Both bird and human vocal organs are versatile, and although some sounds are not possible to produce, the repertoire of human and songbird sounds is large.

Looking beyond these peripheral motor constraints, there are centrally controlled perceptual abilities that propel babies and birds toward their eventual goal, the production of species-typical sound. In humans, perceptual studies have been extensively used to examine the initial capacities and biases of infants regarding speech, and they have provided a wealth of data on the innate preparation of infants for language. At the phonetic level, classic experiments show that early in postnatal life, infants respond to the differences between phonetic units used in all of the world's languages, even those of languages they have never heard (Eimas, 1975a, b; Streeter, 1976; for review see Kuhl, 1987). In these studies, infants are tested using procedures that indicate their ability to discriminate one sound from another. These include the high-amplitude sucking paradigm (in which changes in sucking rate indicate novelty), as well as tests in which a conditioned head turn is used to signal infant discrimination. These tests demonstrate the exquisite sensitivity of infants to the acoustic cues that signal a change in the phonetic units of speech, such as the VOT differences that distinguish /b/ from /p/ or the formant differences that separate /b/ from /g/ or /r/ from /l/.

Moreover, as with adults, infants show categorical perception of sounds, a phenomenon initially demonstrated in adults during the 1950s (Liberman *et al.*, 1967). Tests of categorical perception use a computer-generated series of sounds that continuously vary in small steps, ranging from one syllable (e.g. /ba/) to another (/pa/), along a particular acoustic dimension (in the case of /ba/ and /pa/, the VOT). Adult listeners tend not to respond to the acoustic differences between adjacent stimuli in the series but perceive an abrupt change in the category – the change from /ba/ to /pa/ – at a particular VOT (hence the name categorical perception). In adults, categorical perception generally occurs only for sounds in the adult's native language (Miyawaki *et al.*, 1975). Very young infants not only perceive sounds categorically (Eimas *et al.*, 1971, Eimas, 1975a), but also demonstrate the phenomenon for sounds from languages they have never heard as well as for sounds from their native language (Streeter, 1976; Lasky *et al.*, 1975). These studies provided the first evidence that infants, at birth, have the capacity to discriminate any and all of the phonetic contrasts used in the languages of the world, a feature of auditory perception that greatly enhances their readiness for language learning.

Later studies revealed that nonhuman mammals (chinchillas and monkeys) respond to the same discontinuities in speech that human infants do (Kuhl and Miller, 1975; Kuhl and Padden, 1983), which suggested that human speech evolved to take advantage of the existing auditory capacities of nonhuman primates (Kuhl, 1986). Data also showed that human infant sensitivities extended to nonspeech sounds that contained acoustic dimensions critical to speech but that were not identifiable as speech (Jusczyk *et al.*, 1977). These data caused a shift in what was theorized to be innate (Kuhl, 1986, 1994; Jusczyk, 1981). Initial theories had argued that humans were endowed at birth with "phonetic feature detectors" that defined all possible phonetic units across languages (Eimas, 1975b). These detectors were thought to specify the universal set of phonetic units. When data revealed that the categorical perception of speech was not restricted to humans nor to speech, theories were revised to suggest that what was innate in humans was an initial discriminative capacity for speech sounds, rather than a specification of speech sounds themselves. Infants' discriminative capacities are currently viewed as "basic cuts" in auditory perception. Though not precise, they allow infants to discriminate the sounds of all languages (Kuhl, 1994). Evidence supporting this comes from studies showing that, with exposure to language, the accuracy of discrimination increases substantially for native-language sounds (Kuhl *et al.*, 1997b; Burnham *et al.*, 1987). Theorists noted that these innate perceptual abilities, although not unique to humans, provided infants with a capacity to respond to and acquire the phonology of any language.

As with humans, young songbirds begin life endowed with the capacity for responding to the sounds of their own species, before they have done any singing themselves. Studies of changes in heart rate in young birds in response to song playback initially demonstrated that both male and female sparrows innately discriminate conspecific from heterospecific song (Dooling and Searcy, 1980). Measurement of white-crowned sparrow nestling begging calls in response to tape-recorded song also revealed the much greater vocal behavior of young birds in response to their own species' song than to alien song, providing further evidence of inborn sensory recognition of conspecific song (Nelson and Marler, 1993). This assay also used simplified versions of these songs containing single phrases or modified songs with altered order, to begin to define the minimal acoustical cues critical for this recognition (Whaling *et al.*, 1997).

There is a subtle but important difference between most studies of innate predispositions in songbirds and in humans, however. In birds, what has been examined is not discrimination of sounds within a set of possible songs from a particular species, which would be analogous to studies of phonemes from different human languages. Rather, most studies have looked at learning and listening preferences between songs of different songbird species. This is not possible in humans because one cannot isolate humans in order to expose them to the sounds of other species (to macaque monkey calls, for example) to determine whether they would learn such calls. In birds with whom these experiments have been done, both innate conspecific song recognition and preference are evident in the choice of models for learning song. A variety of experiments, using tape playback of tutor songs, showed that songbirds prefer their own species' song over alien songs as tutor models (Marler and Peters, 1977, 1982a). Songbirds are capable of imitating alien songs, or at least producing modified versions of them, especially in situations in which these are the only songs they hear. When given a choice of conspecific and heterospecific song, however, they preferentially copy the song of their own species. They also usually make much more complete and accurate copies of the conspecific model than of the alien song and may take longer to learn heterospecific song (Konishi, 1985; Marler, 1997; Marler and Peters, 1977). The ability to compare different species has provided evidence that there exists some rudimentary model of species-typical song even in the absence of experience. In humans, there is no convincing experimental evidence that infants have an innate description of speech, but only a few preference tests analogous to those in birds have examined the issue (e.g. Hutt *et al.*, 1968), and the results are not conclusive. Moreover, because infants hear their mothers' voices both through the abdominal wall and through bone conduction and have been shown to learn aspects of speech (prosodic cues) while still in the womb (e.g. DeCasper and Spence, 1986; Moon *et al.*, 1993) (see below), it will be difficult to determine whether infants are endowed with an innate description of speech prior to experience.

In birds, where there is an experimentally verified innate song preference, one can then ask what aspect of the song is required for recognition. Marler and Peters (1989) created synthetic tutor songs with syllables from two different species (the closely related swamp sparrows and song sparrows), arranged in temporal patterns characteristic of one or the other species. Using these songs to tutor the two types of sparrows, they demonstrated that predispositions vary across species. For instance, swamp sparrows copied syllables from their own species' song, regardless of the temporal arrangement of syllables in the synthetic tutor song. In contrast, song sparrows could copy swamp sparrow notes, but only when these were ordered in the usual multipart pattern of song sparrow song. Thus, for the swamp sparrow a critical cue (presumably innately specified) appears to be syllable structure, whereas for song sparrows it is syllable ordering as well as syllable structure. Certain acoustic cues may also serve as attentional flags that permit the acquisition of heterospecific notes. For instance, when the calls of ground squirrels were incorporated into tutor songs that began with the long whistle universally found in white-crowned sparrow song, these sparrows could be shown to learn these squirrel sounds, which they would normally never acquire (Soha, 1995).

In addition to the fact that most studies in birds compare species, another difference between the studies of innate predispositions for song and those for language learning is that in many cases the assay in birds is the song that the bird eventually produces. Any deduction of initial perceptual capacities from the final vocal output confounds initial capacities with subsequent sensory learning and motor production. Nonetheless, the studies of sensory capacities in birds with heart rate or begging call measures provide direct support for the idea that birds innately recognize their own species' song. This recognition is presumed to underlie much of the innate predisposition to learn conspecific song evident in the tutoring experiments. Thus, both humans and birds start out perceptually prepared for specific vocal learning. It may be that songbirds also have more complex innate specifications than do humans, or simply that the analogous experiments (pitting speech against nonspeech sounds) have not been or cannot be done with humans.

Another way of examining innate neural biases is to look at vocal production that emerges prior to, or in the absence of, external acoustic influences. For obvious reasons, relatively few data are available from humans. Deaf babies do babble, but their productions rapidly become unlike those of hearing infants. At a higher level of language analysis, there is some evidence that children exposed only to simple "pidgin" languages, and deaf children exposed to no acoustic or sign language, develop some elements (words or gestures, respectively) and order them in a way that is consistent with a rudimentary grammar (Petitto, 1993; Bickerton, 1990; Goldin-Meadow and Mylander, 1998). It remains disputed, however, whether this reflects an innate model specific to language (Chomsky, 1980; Fodor, 1983) or a more general innate human capacity to learn to segment and group complex sensory inputs (Elman *et al.*, 1996; Bates, 1992).

Songbirds again provide an opportunity to study this issue because analysis of the songs of birds reared in a variety of conditions can provide extensive data relevant to the issue of what may be innate in a vocal learner. In normally reared songbirds, the song of every individual bird within a species differs, but there are enough shared characteristics within a species that songs can also be used for species identification. The songs of birds raised in complete isolation vary between individuals but always contain some of the species-specific structure, although these songs are much less complex than those of tutored birds: the songs of white-crowned sparrow isolates tend to contain one or more sustained whistles, swamp sparrow isolates sing a trilled series of downsweeping frequencies, and song sparrow isolates produce a series of notes ordered in several separate sections. Even when white-crowned sparrows have copied alien song phrases, they often add an "innate" whistle ahead of these (Konishi, 1985; Marler, 1997, 1998). Thus, there is innate information that provides rough constraints on the song even in the absence of tutoring experience. Strikingly, almost all these features require auditory feedback to be produced. Because these features must be translated into vocal output via sensorimotor learning, they cannot be completely prespecified motor programs; they must involve some sensory recognition and feedback. Thus, the innate mechanisms that direct isolate song might bear some relationship to the neural mechanisms that allow innate sensory recognition of song. Recent behavioral evidence, however, suggests that there is not complete overlap between isolate song and the features found to be critical for innate conspecific recognition (Whaling *et al.*, 1997).

Innate sensory recognition and learning preferences in both humans and songbirds suggest that there must be underlying genetic mechanisms, perhaps specifying auditory circuitry specialized for processing complex sounds in special ways. An advantage of songbirds is that, unlike humans, there are many different, but closely related, species and even subspecies of vocal learners that show variation in their capacity to learn

(Kroodsma and Canady, 1985; Nelson *et al.*, 1996). An intriguing example is the recent result of Mundinger (1995), who showed that the roller and border strains of canaries, which differ in note types, simply do not learn or retain in their songs the note types most specific of the other strain. However, hydrid offspring of the two breeds readily learn both types, and analysis of the patterns of inheritance of this capacity in these birds and in back-crosses has even begun to point to chromosome linkage (Mundinger, 1998). Comparisons of perceptual and motor learning and their neural substrates in birds like these may facilitate eventual understanding of the neural mechanisms contributing to innate biases for vocal learning.

Perceptual learning and the effects of experience

Although neither human nor songbird brain starts out perceptually naïve, abundant evidence in both fields suggests that innate predispositions are subsequently modified by experience. In addition, both speech and song scientists are grappling with the question of how experience alters the brain. In purely selective models of learning, sensory experience simply selects the sounds to be used to guide vocal learning from an extensive set of pre-encoded possibilities. In purely instructive models, there is no innate information about what is to be learned, and experience simply instructs a wide open brain about what to memorize. In fact, studies of both song and speech are converging on the idea that the mechanisms underlying learning are not described by either of these extreme models but combine aspects of each.

PERCEPTUAL LEARNING IN HUMANS MODIFIES INNATE PREDISPOSITIONS

As described, at the phonetic level of language, infants initially discriminate phonetic units from all languages tested, showing that they perceive and attend to the relevant acoustic features that distinguish speech sounds. By 6 months of age, however, infants have been affected by linguistic experience and show recognition of the specific phonetic units used in their native language. At this age, they respond differently to phonetic prototypes (best instances of phonetic categories) from the native as opposed to a foreign language (Kuhl, 1991; Kuhl *et al.*, 1992). By 9 months, they have learned the stress patterns of native-language words, and the rules for combining phonetic units (Jusczyk *et al.*, 1993), phrasal units (Jusczyk *et al.*, 1992), and the statistical probabilities of potential word candidates (Saffran *et al.*, 1996). Finally, by 12 months of age, native-language learning is evident in the dramatic changes seen in perceptual speech abilities (Werker and Tees, 1992) (Figure 2.3A). Infants no longer respond to speech contrasts that are not used in their native language, even the ones that they did discriminate at earlier ages (Werker and Tees, 1984; Kuhl *et al.*, 1997b). Instead, one-year-old infants show the pattern typical of adult native-language listeners wherein discrimination of foreign-language contrasts has been shown to be difficult: adult English speakers fail to discriminate Hindi consonant-vowel combinations (Werker and Tees, 1984, 1992), American speakers fail on Spanish /b/ and /p/ (Abramson and Lisker, 1970), and speakers of Japanese fail to discriminate American English /r/ and /l/ (Miyawaki *et al.*, 1975). The decline in the language-universal perception of infants has been directly demonstrated for Canadian infants tested sequentially over time with Hindi contrasts (Werker and Tees, 1984) and, most recently, for Japanese infants listening to American English /r/ and /l/ (Kuhl *et al.*, 1997b).

In humans, there is evidence that perceptual learning of the more global, prosodic aspects of language actually commences prior to birth. Studies using the sucking and heart rate paradigms show that exposure to sound in utero has resulted in a preference of newborn infants for native-language over foreign-language utterances (Moon *et al.*, 1993), for the mother's voice over another female's voice (DeCasper and Fifer, 1980), and for simple stories the mother read during the last trimester over unfamilar stories (DeCasper and Spence 1986). This indicates that the prosodic aspects of human speech, including voice pitch and the stress and intonation characteristics of a particular language and speaker, are transmitted to the fetus and are learnable.

All these studies on learning in the first year of life indicate that prior to the time that infants learn the meanings of individual words or phrases, they learn to recognize general perceptual characteristics that describe phonemes, words, and phrases that typify their native language. Thus, as a first step toward vocal learning, infants avidly acquire information about the perceptual regularities that describe their native language and commit them to memory in some form. Understanding the nature of this early phonetic learning and the mechanisms underlying it is one of the key issues in human language development.

PERCEPTUAL LEARNING IN SONGBIRDS

A variety of experiments provide evidence that what occurs in the first, or sensory, phase of song learning is the memorization of the sensory template, which is a subset of all possible vocalizations of the species (Marler, 1970b). This phase is thus in many ways analogous to the early perceptual learning of human infants. The study of perceptual learning in songbirds that is most similar to studies of humans measures vocal behavior of 10- to 40-day-old white-crowned sparrows in response to playback of tutored and novel songs (Nelson *et al.*, 1997). After 10-day periods of tape tutoring with pairs of songs, male white-crowned sparrows not only gave significantly more calls to tutor songs than to novel songs, they also called significantly more to the song of the pair they would subsequently produce than to the nonimitated song of that pair. This suggests that the vocal assay reflected sensory learning that would ultimately be used for vocal production.

Most studies of the sensory learning period in songbirds, however, have assessed what is learned by using adult song production as an assay, after tutoring birds either for short blocks of time beginning at different ages or with changing sets of songs for a long period of time (Marler, 1970b; Nelson, 1997). Measuring learning using song production may underestimate what is perceptually learned. In many of these tutoring experiments, however, the song ultimately produced reflected experiences that had occurred long before the birds had begun to produce vocalizations; these studies, therefore, provide strong evidence that the first phase of learning involves the memorization of song.

In contrast to the emerging data on in utero learning in humans, prehatch or even immediately posthatch experience has not yet been shown to have much influence on song learning. Rather, in the well-studied white-crowned sparrows, the sensory period begins around day 20 and peaks in the next 30 days, with some acquisition possible up to 100 or 150 days (Baptista and Petrinovich, 1986; Marler, 1970b) (Figure 2.3). The timing of sensory learning may be similar for many other seasonal species (Kroodsma and Miller, 1996; Catchpole and Slater, 1995). Studies of zebra finches in which birds were separated from their tutors at different ages suggest that different aspects of the tutor song are memorized in sequence, with the individual component sounds being learned first and the overall order and temporal pattern acquired later (Immelmann, 1969). Careful comparisons of related white-crowned sparrow subspecies under identical learning conditions show that genetics also plays a role in the exact timing of learning, because subspecies of sparrows from harsh climates with short breeding seasons learn earlier and more than their coastal cousins (Nelson *et al.*, 1995). Such differences between birds provide an opportunity to identify the factors governing sensory learning.

HOW DOES EXPERIENCE ALTER PERCEPTUAL ABILITIES IN HUMANS?

The initial studies demonstrating categorical perception of speech sounds in infants and its narrowing with language exposure led many speech theorists to take a strongly nativist or selective view of speech learning. By this hypothesis, infants were thought to be biologically endowed with either phonetic feature detectors that specified all the phonetic units used across languages (e.g. Eimas, 1975b), or with knowledge of all linguistically significant speech gestures (Liberman and Mattingly, 1985). The subsequent decline in speech discrimination was seen as a process of atrophy of the prespecified phonetic representations in the absence of experience. Recent studies of languages and of experience-dependent perceptual maps are changing theories of language learning and the role of innate and learned factors in the acquisition process. Rather than experience only selecting from prespecified categories, experience is thought to establish memory representations for speech that specify the phonetic units used in that language and that alter the perceptual system of the infant (Kuhl, 1994). On this view, experience is instructive as well as selective.

Several lines of evidence support this changing view. For one, cross-linguistic studies show that across languages, even ostensibly similar vowels (such as the vowel /ʌ/) show a great deal of variation (Ladefoged, 1994). This suggests that prestoring all possible phonetic units of the world's languages would not be an efficient process. A second line of evidence against a simple atrophy of phonetic representations from lack of exposure is that, often, listeners are exposed to the categorical variations that they eventually fail to perceive. For instance, approximations of both English /r/ and /l/ are produced interchangeably by Japanese adults, although they do not change the meanings of words (Yamada and Tohkura, 1992). Japanese infants are therefore exposed (albeit randomly) to variants of both /r/ and /l/; similarly, American infants are exposed to variants of Spanish /b/ and /p/. Yet, both groups will eventually fail to respond to those distinctions. Finally, more detailed studies on the changes in infant phonetic perceptions brought about by experience suggest that perceptual learning is not in fact a simple sensory memory of the sound patterns of language. Instead, it seems to be a complex mapping in which perception of the underlying acoustic dimensions of speech is warped to create a recognition network that emphasizes the appropriate phonetic differences and minimizes those that are not used in the language (Kuhl, 1994, 1998, Kuhl and Meltzoff, 1997). This warping of the underlying dimensions is language specific such that no adult speakers of any language perceive speech sounds veridically. Rather, in each language group, perception is distorted to enhance perception of that language: this has been called the perceptual magnet effect (PME).

This last line of evidence results from studying perception of sounds in more detail than simply identifying category boundaries. Kuhl (1998) used large grids of systematically varying consonant–vowel syllables spanning the phonetic boundary between American English /r/ and /l/ to test American and Japanese adults. They asked listeners to rate the perceptual similarity of all possible pairs of stimuli and used multidimensional scaling techniques to create a map of the perceived physical distances between stimuli. The maps for American and Japanese speakers indicated that although the real physical distances between each stimulus in the grid were equal, American and Japanese adults perceived the sounds, and the distances between them, very differently. Americans identified the sounds as belonging to two clearly different categories, /r/ and /l/, whereas Japanese identified all stimuli but one as Japanese /r/ (the only phoneme of this type normally used in Japanese). Moreover, American listeners perceived many sounds as if they were closer to the best, most prototypical examples of /r/ and /l/ (called prototypes) than they really were. This is the origin of the term perceptual magnet effect, meant to describe how prototypes seem to act as magnets for surrounding sounds. Americans also perceived a larger than actual separation between the two categories. Japanese listeners showed no magnet effects, and no separation between the two categories. Thus, neither of the two groups perceive the real physical differences between the sounds. Instead, language experience has warped the underlying physical space so that if certain categories of sounds are used in a language, differences within a category are perceptually shrunk, whereas differences between categories are perceptually stretched. The PME may aid in perception by reducing the effects of the variability that exists in physical speech stimuli.

Critically for theories of speech learning, further studies suggest that these mental maps for speech are being formed or altered early in life as a function of linguistic experience. At 6 months of age, infants being raised in different cultures listening to different languages show the PME only for the sounds of their own native language (Kuhl *et al.*, 1992). Moreover, when American and Japanese infants were tested at 6–8 months of age, both groups showed the ability to discriminate American English /r/ and /l/, as expected from previous studies. By 10–12 months, however, not only did Japanese infants show a dramatic decline in performance, but American infants had also increased their accuracy of discrimination. This suggests that experience is not simply preventing atrophy (Kuhl *et al.*, 1997b). Finally, monkeys do not show the PME, indicating that, unlike categorical perception, it is not an effect that is inherent in the auditory processing of speech stimuli in many animals (Kuhl, 1991). The implication is that magnet effects explain the eventual failure of infants to discriminate foreign-language contrasts. Japanese infants, for example, would form a phonetic prototype for Japanese /r/ that is located between American /r/ and /l/. The magnet effect formed by experience with Japanese would eventually cause a failure to discriminate the American sounds. Although the studies show that magnet effects are altered by experience, it is not yet known whether magnet effects initially exist for all sounds of all languages and are then modified by experience, or whether they do not exist initially and are formed as a function of experience (Kuhl, 1994).

The special kind of speech that adults use when they speak to infants ("parentese") may play a role in the normal infant development of these phonemic maps. It has long been known that adults speak to infants using a unique tone of voice, and that when given a choice, infants prefer this kind of speech (Fernald, 1985, Fernald and Kuhl, 1987; Grieser and Kuhl, 1988). Early work on parentese emphasized the prosodic differences (the increased fundamental frequency or pitch of the voice, its animated intonation contours, and its slower rate). Recent data show, however, that infant-directed speech also provides infants with greatly exaggerated instances (hyperarticulated prototypes) of the phonetic units of language (Kuhl *et al.*, 1997b). When speaking to infants, humans may intuitively produce a signal that emphasizes the relevant distinctions and increases the contrast between phonetic instances.

The studies described above all lend support to the newly emerging view that the initial abilities of infants to discriminate the auditory dimensions employed in speech contrasts are dramatically altered simply by listening to ambient language, resulting in a new and more complex map of the relevant linguistic space. The perception of speech in infants is thus both highly structured at birth, promoting attention to the relevant acoustic distinctions signaling phonetic differences, and highly malleable, allowing the brain to lay down new information, instructed by experience.

HOW DOES EXPERIENCE ACT ON THE SONGBIRD BRAIN?

Studies of perceptual learning in humans suggest that initial basic divisions of sound space are gradually altered by experience with the native language. The same questions about how this occurs that have been raised in humans can be asked about the effects of sensory experience in birds. The two extreme models (instructive and selective) discussed in the case of human speech have also been raised in the case of birdsong (Marler, 1997).

A purely instructive model would suggest that birds have little foreknowledge about the song of their species and are equally ready and able to learn virtually any song to which they are exposed. This is not consistent with innate preferences for learning conspecific song (Marler, 1997; Marler and Peters, 1982a). It also cannot explain isolate songs. These songs vary a great deal between individuals, however, which suggests that the innate template only coarsely defines the species song. The instructive model does account for the fact that prior to the production of the final learned songs, many birds produce copies of syllable types that are not used later in their final songs (Marler and Peters, 1982a). Even the syllables of alien species to which a bird was exposed can be reproduced in this way (Thorpe, 1961; Konishi, 1985). This phemenonon of overproduction of syllables suggests that birds are instructed by experience to memorize and even produce multiple songs, including songs of other species. The instructive model has difficulty, however, explaining the usual attrition later in song learning of syllables not appropriate for the species. A more realistic view of the instructive model would posit that during the impressionable phase, birds memorize a variety of songs, perhaps memorizing more easily or more completely songs that match their prespecified preferences. Later, during sensorimotor learning, birds listen to their vocalizations and use the memorized songs as templates to assess how well their vocal output matches them. They then ultimately elect to produce as adults a subset of those songs; the selection of this subset may be guided by a combination of genetic biases and experience (Nelson and Marler, 1994). Thus, even the simplest instructive model contains some elements of selection, both at the early (sensory) and at the late (sensorimotor) learning stages.

Alternatively, a strictly selective model of song learning can be proposed, in which the songbird brain has extensive innate knowledge about its species song, and this knowledge is then simply activated by experience. Evidence in favor of this includes innate song learning preferences and the surprising lack of variability seen in nature when the song patterns of an entire species are analyzed (Marler and Nelson, 1992; Marler, 1997). In contrast to the drift that might be expected in a culturally transmitted behavior operating by instruction alone, there are a number of features of song that are always shared, so-called species universals. None of these universals develop fully in birds raised in isolation, however. According to the pure selection model, therefore, all possible universals are pre-encoded in the brain, but most of them must be activated by the sensory experience of matched sounds in order to be available for later guidance of motor development, whereas the species universals that are not heard atrophy. Consistent with this idea, although not conclusive, is the surprisingly small number of sensory exposures necessary for learning. For example, white-crowned sparrows can learn from as few as 30 repetitions of a song, and nightingales have been shown to learn songs presented only twice a day for 5 days (Peters *et al.*, 1992; Hultsch and Todt, 1989a).

As with the strict instructive model, however, even highly selective models seem likely to have some elements of instruction, for instance to allow the significant culturally transmitted variation seen within each category of universals (much like the variations in the vowel /i/ in human languages), and the copying of complex sequences, without requiring a multitude of templates. Moreover, because some features of song are produced in isolated birds, there must be two sorts of pre-encoded templates, ones that require no auditory experience of others to be active and a much larger set that do require auditory experience (Marler, 1997). In addition, and perhaps most important, a purely selective and species-based mechanism does not explain why birds can learn songs of heterospecifics, when birds are raised with those songs alone or even sometimes in the presence of conspecific songs as well (Baptista and Morton, 1981; Immelmann, 1969). One must therefore postulate two different learning mechanisms, one for conspecific song and a different one (perhaps a more general sensory learning) when other songs are learned. Although this is possibly consistent with data suggesting that birds take more time to learn alien song, it also necessitates a multiplication of learning substrates and makes it harder to explain why birds may incorporate both conspecific and heterospecific syllables into a single song. Finally, some of the lack of variability in the final crystallized song of many birds could be due not to selection at the early memorization stage, but rather in part to the highly socially controlled selection process active during late plastic song and crystallization, in which birds choose to crystallize the songs most similar to their neighbors (Nelson and Marler, 1994). Clearly, more studies are necessary to resolve the question of how sensory experience acts on the brain. Already it seems likely, however, that some combination of selection and instruction act both in series and in parallel in song learning. In many ways this is strikingly similar to the issues in the speech field, where purely innate and selection-based models are now making way for the idea that initial capacities are revised by instructive effects of experience.

Better understanding of the neural mechanisms underlying learning might also help resolve this issue. For instance, pre-existing circuitry and innate auditory predispositions might be revealed at the neural level, both in humans (using imaging) and in songbirds. The brain of songbirds contains a system of areas devoted to song learning and production and in adult birds these contain numerous neurons that respond selectively to the sound of the bird's own song and poorly to the songs of other conspecifics or to temporal alterations of the bird's own song (Margoliash, 1983, 1986; Margoliash and Fortune, 1992). In young birds just in the process of learning to sing, however, these same neurons are broadly selective for any conspecific songs, and they only gradually develop selectivity for their own song during learning (Doupe, 1997; Solis and Doupe, 1997; Volman, 1993). This suggests that at least this part of the song system contains neurons that are initially nonselective, i.e. without specific foreknowledge of the song the bird will sing, and that are subsequently instructed by experience.

SOCIAL EFFECTS ON SENSORY LEARNING

Both songbirds and humans demonstrate that learning is not solely dependent on innate predispositions and acoustic cues. Social factors can dramatically alter learning. Songbirds have been shown to learn alien songs from live tutors when they would reject the same songs presented by tape playback (Baptista and Petrinovich, 1986), and zebra finches will override their innate preference for conspecific song and learn from the Bengalese finch foster father feeding them, even when adult zebra finch males are heard nearby (Immelmann, 1969). Zebra finches, a more highly social and less territorial species than many songbirds, are particularly dependent on social factors even for selection of a particular conspecific tutor, as demonstrated in a series of experiments from the laboratory of Slater and colleagues. These experiments showed that zebra finches, which do not learn well from tapes, required visual interaction with the tutor in a neighboring cage in order to copy it, even if they could hear it (Eales, 1989; Slater *et al.*, 1988). Zebra finch fledglings prevented by eye patches from seeing, however, would still learn from a tutor if it was in the same cage, allowing the usual local social interactions (pecking, grooming, etc.) seen between zebra finch tutors and young. Finally, Adret (1993) showed that replacing the social interaction with a taped recording that the young zebra finch had to activate with a key press resulted in the zebra finch actively key pressing and then learning from that tape. Thus, the social factors required by zebra finches can come in a variety of modalities, all of which may serve to open some attentional or arousal gate, which then permits sensory learning. Such attentional mechanisms may also explain birds' preferential selection of a conspecific tutor during sensory learning and their choice of a particular song for crystallization.

Social interaction has been suggested to play a critical role in language learning as well (Locke and Snow, 1997; Kuhl and Meltzoff, 1996, 1997), although clearly studies of humans cannot withdraw social interaction to study the effects on vocal learning. Consistent with the importance of social cues are the speech patterns of adults in language addressed to infants. These patterns are greatly modified in ways that may aid language learning. In addition, neglected infants are developmentally delayed in language (Benoit *et al.*, 1996), and much of early word learning is deeply embedded in shared social activities. It is not clear whether a tape recorded or televised speaker would permit language learning in infants, although this could be addressed in studies of second language learning. Infants engaged in social interaction appear to be highly aroused and attentive, which may play a role in their ability to react to and learn socially significant stimuli. As in birds, such arousal mechanisms might help to store and remember stimuli and to change their perceptual mapping (Kilgard and Merzenich, 1998).

VOCAL PRODUCTION AND ITS INTERACTION WITH PERCEPTION

In vocal learning by humans and songbirds, both perception and production of sound are crucial. One must perceive both the vocal models of others and one's own sounds, and one must learn the mapping from one's own motor commands to the appropriate acoustic production. It has been clear for a long time that these two aspects of vocalization interact strongly, and in fact early speech theorists suggested that sound decoding requires creation of a model of the motor commands necessary to generate those sounds (Liberman *et al.*, 1967). In songbirds, however, memorization of sounds clearly precedes their generation. Recently, studies showing that human perception of speech is highly sophisticated at birth and then rapidly sculpted by experience, prior to the emergence of a sophisticated capacity for sound production, have led to a new view in studies of speech that is strikingly similar to that in birdsong. By this hypothesis, acoustic targets that are a subset of all possible species' vocalizations are perceptually learned by the young individual (bird or human) by listening to others. This perceptual learning then powerfully constrains and guides what is (and can be) produced. Subsequent production then aids in creating auditory-articulatory maps; the relationship between production and perception continues to be highly interactive but is derived, at least initially, from perceptual maps.

Production and perception in humans

In humans, the interaction between perception and production has been studied in two ways, by examining the infant's own production of sound and by examining the infant's reactions to the sight of others producing sound. Both assess what infants know about speech production and its relation to perception.

One strategy is to describe the progression of sounds produced by infants across cultures as they mature, examining how exposure to language alters speech production patterns. Characteristic changes in speech production occur as a child learns to talk, regardless of culture (for review see Stoel-Gammon, 1992). All infants progress through a set of universal stages of speech production during their first year: early in life, infants produce nonspeech gurgles and cries; at 3 months, infants coo, producing simple vowel-like sounds; by 7 months infants begin to babble; and by 1 year first words appear (Figure 2.3A). The cross-cultural studies also reveal, however, that by 10–12 months of age, the spontaneous vocalizations of infants from different language environments begin to differ, reflecting the influence of ambient language (de Boysson-Bardies, 1993). Thus, by the end of the first year of life, infants diverge from the culturally universal speech pattern they initially exhibit to one that is specific to their culture, indicating that vocal learning has taken place.

It is not the case, however, that the remarkable ability of infants to imitate the speech patterns they hear others produce begins only toward the end of their first year. Recent laboratory studies indicate that infants have the capacity to imitate speech at a much earlier age. Infants listening to simple vowels in the laboratory alter their vocalizations in an attempt to approximate the sounds they hear, and this ability emerges around 20 weeks of age (Kuhl and Meltzoff, 1996). The capability for vocal motor learning is thus available very early in life. In adults, the information specifying auditory–articulatory relations is exquisitely detailed and allows almost instantaneous reaction to changes in load or position of the articulators in order still to produce the appropriate sound (Perkell *et al.*, 1997). Although speech production skills improve throughout childhood, showing that auditory–articulatory maps continue to evolve over a long period, the early vocal imitation capacities of infants indicate that these maps must also be sufficiently formed by 20 weeks of age to allow infants to approximate sounds produced by others.

A comparison of the developmental timelines relating speech perception and speech production suggests that early perceptual mapping precedes and guides speech production development (Kuhl and Meltzoff, 1996, 1997). Support for this idea comes from a comparison of changing perceptual abilities and production in the infant studies just described: a language-specific pattern emerges in speech perception prior to its emergence in speech production. For instance, although infant vocalizations produced spontaneously in natural settings do not become language-specific until 10–12 months of age, the perceptual system shows specificity much earlier (Mehler *et al.*, 1988; Kuhl *et al.*, 1992). In addition, at an age when they are not yet producing /r/- or /l/-like sounds, infants in America and Japan already show language-specific patterns of perception of these sounds. These data suggest that stored representations of speech in infants alter perception first and then later alter production as well, serving as auditory patterns that guide motor production. This pattern of learning and self-organization, in which perceptual patterns stored in memory serve as guides for production, is strikingly similar to that seen in birdsong, as well as in visual-motor learning, such as gestural imitation (Meltzoff and Moore, 1977, 1997).

A second experimental strategy reveals the link between perception and production for speech. In this case, studies demonstrate that watching another talker's mouth movements influences what subjects think they hear, indicating that representational maps for speech contain not only auditory but visual information as well. Some of the most compelling examples of the polymodal nature of speech are the auditory–visual illusions that result when discrepant information is sent to two separate modalities. One such illusion occurs when auditory information for /b/ is combined with visual information for /g/ (McGurk and MacDonald, 1976; Massaro, 1987; Kuhl *et al.*, 1994). Perceivers report the strong impression of an intermediate articulation (/da/ or /tha/), despite the fact that this information was not delivered to either sense modality. This tendency of human perceptual systems to combine the multimodal information (auditory and visual) to give a unified percept is a robust phenomenon.

Infants 18–20 weeks old also recognize auditory–visual correspondences for speech, akin to what adults do when they lip-read. In these studies, infants looked longer at a face pronouncing a vowel that matched the vowel sound they heard than at a mismatched face (Kuhl and Meltzoff, 1982). Young infants therefore demonstrate knowledge about both the auditory and the visual information contained in speech. This supports the notion that the stored speech representations of infants contain information of both kinds.

Thus, early perceptual learning – primarily auditory but perhaps also visual – may underpin and guide speech production development and account for infants' development of language-specific patterns by the end of the first year. Linguistic exposure is presumably the common cause of changes in both systems: memory representations that form initially in response to perception of the ambient language input then act as guides for motor output (Kuhl and Meltzoff, 1997).

Production and perception in birdsong

The observation that perceptual learning of speech may precede and guide production in humans makes it strikingly similar to birdsong, which clearly does not require immediate motor imitation while the young bird is still in the presence of the tutor. Many seasonal species of birds begin the sensorimotor learning phase, in which they vocally rehearse, only many months after the tutor song has been heard and stored (Figure 2.3B). Thus, birds can remember complex acoustical patterns (that they heard at a young age) for a long time and use them much later to guide their vocal output.

The lack of overlap between the sensory and sensorimotor phases of song learning is not as complete as often supposed, however, and in this sense some songbirds are also more like humans than previously thought. This is most obvious in the zebra finch (Immelmann, 1969; Arnold, 1975a), which is not a seasonal breeder and develops song rapidly over a period of 3–4 months, and in which sensory and sensorimotor learning phases overlap for at least a month. Thus, as in humans, these finches continue to copy new sounds after sensorimotor learning has started. Even in the classical seasonal species, birds often produce the amorphous vocalizations known as subsong as early as 25 days of age, well within the 100-day sensitive phase (Nelson *et al.*, 1995). These early vocalizations of songbirds may allow calibration of the vocal apparatus and an initial mapping between motor commands and sound production, a function similar to that proposed for human babbling (Marler and Peters, 1982a; Kuhl and Meltzoff, 1996, 1997). Moreover, in more complex social settings, the schedule for the onset of singing and sensorimotor learning can be dramatically accelerated (Marler, 1970b; Baptista and Petrinovich, 1986).

Nonetheless, in many species, perceptual learning of the tutor is complete before the so-called sensorimotor stage of learning begins in earnest, usually toward the end of a seasonal bird's first year of life (Figure 2.3B). This stage begins with a great increase in the amount of singing, and soon thereafter, vocalizations show clear evidence of vocal rehearsal of learned material, at which point they are termed plastic song. These are gradually refined until they resemble the tutor song. Along the way, however, birds produce a wide variety of copied syllables and songs, only to drop them before crystallization (Marler and Peters, 1982a; Nelson and Marler, 1994). During the plastic song phase, birds also often incorporate inventions and improvisations that make their song individual. At the end of the sensorimotor phase, birds produce a stable, or "crystallized", adult song, which in most species remains unchanged throughout life (except for open learners; see below). Because tutor learning occurs largely before production, it cannot depend on motor learning. During sensorimotor learning, however, sensory processing of sounds might conceivably change or become more dependent on knowledge of motor gestures. This question could be studied in songbirds raised with normal sensory exposure to others but experimentally prevented from producing sounds.

Just as in humans, not all sensory effects on song learning are mediated solely by auditory feedback. Not only do zebra finches require some sort of visual or social interaction to memorize a tutor, but male cowbirds will choose to crystallize the particular one of their several plastic songs that elicits a positive visual signal, a wingflap, from a female cowbird (West and King, 1988). Thus, visual cues can also affect song learning by acting on the selection of songs during motor learning. Along the same lines, Nelson and Marler (1993) demonstrated that late juvenile sparrows just arriving at the territory where they will settle will choose to crystallize the plastic song in their repertoire that is the most similar to the songs sung in that territory. This result was replicated in the laboratory by playing back to a sparrow just one of four plastic songs that it was singing, which invariably resulted in that song being the one crystallized (Nelson and Marler, 1994). These social effects on crystallization may allow the matched countersinging frequently observed in territorial birds. By allowing visual and auditory cues to influence song selection, birds incorporate the likelihood of successful social interaction into their final choice of vocal repertoire.

PERCEPTION OF SELF AND ITS INTERACTION WITH PRODUCTION

Although auditory processing of the sounds of others is important in speech and song learning, the interaction between perception of one's own sounds and vocal production is also crucial, because vocal learning depends on the ability to modify motor output using auditory feedback as a guide. In both birdsong and speech, the sensory and motor processes are virtually inseparable. One striking demonstration of this is that in frontal and temporoparietal lobes of humans, stimulation at single sites disrupts both the sequential orofacial movements used in speech production and the ability to identify and discriminate between phonemes in perception tasks (Ojemann and Mateer, 1979). This provides more evidence that the traditional description of Broca's and Wernicke's aphasias as expressive and receptive is oversimplified. Likewise, the song premotor nucleus HVC also contains numerous song-responsive neurons (Margoliash, 1983, 1986; McCasland and Konishi, 1981).

The important question of how and where the auditory feedback from self-produced vocalizations acts and how it relates to vocal motor processes remains unclear for both humans and songbirds. In humans, the majority of individual speech-related neurons studied thus far have been active only during either speech production or speech perception (even with identical words presented and then spoken). Thus, the vocal control system seems in some way to inhibit the response of these neurons to the sound of self-vocalized words. More striking, this link between auditory and vocal systems already exists in nonhuman primates: more than half of the auditory cortex neurons responsive to the presentation of calls in squirrel monkeys did not respond to these calls when they were produced by the monkeys (Müller-Preuss and Ploog, 1981). Similarly, in songbirds, despite the strong responses of HVC song-selective neurons to presentation of the bird's own song, these neurons are not obviously activated by the sound of the bird's own song during singing, and in many cases they are clearly inhibited during and just after singing in adult birds (McCasland and Konishi, 1981). Thus, information that there is vocal activity is provided to auditory and even vocal control areas in both primates and songbirds, but it is not clear how the sounds made by this activity are used. This puzzle is evident in some but not all PET studies as well: even though Wernicke's area is strongly activated during auditory presentation of words, a number of such studies have shown surprisingly little activation of the same area from reading or speaking aloud (Ingvar and Schwartz, 1974; Petersen *et al.*, 1989; Hirano *et al.*, 1996; but see also Price *et al.*, 1996). A recent study of vocalizing humans may shed light on this question: this showed much more activation in superior temporal gyri when auditory feedback of the subject's own voice was altered than when it was heard normally (Hirano *et al.*, 1997; McGuire *et al.*, 1996). This raises the possibility that, at least once speech is acquired, Wernicke's and other high-level speech processing areas may be more active when detecting mismatched as opposed to expected auditory feedback of self. In birds as well, it will be important to test neuronal responses when auditory feedback of the bird's own voice is altered.

As with primates, comparisons of songbirds with closely related species that are not vocal learners have the potential to provide insights into the steps that led to song learning. For instance, the suboscine birds such as flycatchers and phoebes, which are close relatives of the passerine (or oscine) songbirds, sing but show no evidence of dialects or individual variations and produce normal song even when deafened young (Kroodsma, 1985; Kroodsma and Konishi, 1991). These birds also show no evidence of a specialized forebrain song control system, which suggests that another crucial step in the appearance of specialized song control areas may have been the acquisition of auditory input by pre-existing forebrain motor control areas. Likewise, in humans, the capacity to learn speech and the development of specialized cortical systems for its control may have resulted from close interaction of motor control areas for orofacial movements with a variety of areas involved in processing and memorizing complex sounds (Ojemann, 1991). Despite its clear importance, the link between perception and production is surprisingly ill understood in both speech and song systems, and further understanding of how motor control and auditory feedback interact at the neural level will be crucial for progress in both fields.

SENSITIVE PERIODS FOR SPEECH AND SONG LEARNING

A critical period for any behavior is defined as a specific phase of the life cycle of an organism in which there is enhanced sensitivity to experience, or to the absence of a particular experience. One of the most universally known and cited critical periods is that for human language acquisition. Songbirds also do not learn their vocalizations equally well at all phases of life. In this final section we review the evidence suggesting that sensitive periods for vocal learning in these two systems are indeed very similar, and we examine and compare possible underlying mechanisms.

The term critical was initially coined in the context of imprinting on visual objects early in life, in which sensitivity to experience is short-lived and ends relatively abruptly. Many critical periods, however, including those for vocal learning, begin and end less abruptly and can be modulated by a variety of factors, so the term now preferred by many investigators is sensitive or impressionable period. Because critical period is such a commonly recognized term, we use these terms interchangeably, but with the caveat that this does not necessarily imply a rigidly regulated and complete loss of sensitivity to experience.

BASIC EVIDENCE FOR SENSITIVE PERIODS IN BIRDS AND HUMANS

Humans

Lenneberg (1967) formulated the strongest claims for a critical or sensitive period for speech learning, stating that after puberty it is much more difficult to acquire a second language. Lenneberg argued that language learning after puberty was qualitatively different, more conscious and labored, as opposed to the automatic and unconscious acquisition that occurs in young children as a result of mere exposure to language.

Evidence for a sensitive period for language acquisition has been derived from a variety of sources: (1) classic cases of socially isolated children show that early social isolation results in a loss of the ability to acquire normal language later (Fromkin *et al.*, 1974; Lane, 1976); (2) studies of patients who suffer cerebral damage at various ages provide evidence that prognosis for language recovery is much more positive early in life as opposed to after puberty (e.g. Duchowny *et al.*, 1996; Bates, 1992); and (3) studies of second-language learning indicate that there are differences in the speed of learning and ultimate accuracy of acquisition in language learning at different stages of life (Johnson and Newport, 1991; Oyama, 1978; Snow, 1987).

It has been known for a long time that children recover better from focal brain injury than do adults with analogous lesions. Moreover, after major damage to left frontal or parietal lobes, or even hemispherectomies for intractable epilepsy, children can still develop language using the right hemisphere (e.g. Dennis and Whitaker, 1976; Woods, 1983). There is an upper limit to this extreme plasticity, however, with studies suggesting the cut-off occurs sometime between 3 and 6 years of age. In cases of less severe injury, the period after 6–8 years, but before puberty, is still more likely to support learning of speech than the period after puberty (Vargha-Khadem *et al.*, 1997).

Studies on the acquisition of a second language offer the most extensive data in support of the idea that language learning is not equivalent across all ages. For instance, second languages learned past puberty are spoken with a foreign accent, in other words with phonetics, intonation, and stress patterns that are not appropriate for the new language. Comprehension of spoken speech and grammar, as well as grammatical usage, are also poorer for languages learned later in life. Numerous studies show that all these aspects of language are performed poorly by immigrants who learn a second language after the ages of 11–15 years, independent of the length of time the learner has been in the new country (Oyama, 1976, 1978; Johnson and Newport, 1991; Newport, 1991). Even when adults initially appear to acquire certain aspects of language faster than children, they do not end up as competent as children after equivalent amounts of training (Snow, 1987).

Moreover, the capacity to learn may decline in several stages. A number of studies suggest that children who have been exposed to and learned a new language at a very young age, between 3 and 7 years of age, perform equivalently to native speakers on various tests. After 6–8 years of age, performance seems to decline gradually but consistently, especially during puberty, and after puberty (after approximately 15–17 years of age), there is no longer any correlation between age of exposure and performance, which is equally poor in all cases (Tahta *et al.*, 1981; Asher and Garcia, 1969; Flege, 1991). A similar pattern of results is shown in deaf adults who are native-language signers but have learned American Sign Language (ASL) at different ages: a comparison of subjects who had learned either from birth, from 4 to 6 years of age, or after the age of 12 showed a clear progression in both production and comprehension of the grammar of ASL that indicated that earlier learners signed more accurately than later learners (Newport, 1991).

Could the critical period simply be a limitation in learning to produce speech, while perceptual learning is not limited? Studies suggest that the accents adult learners use when attempting to produce a foreign language are not attributable to simple motoric failures in learning to pronounce the sounds of the new language but also involve perceptual difficulties. When students were tested on a foreign language 9–12 months after their first exposure to it, those with the best pronunciation scores also showed the best performance on the discrimination test (Snow and Hoefnagel-Höhle, 1978). Moreover, the numerous studies of perception reviewed earlier (Werker and Polka, 1993; Kuhl, 1994) indicate that adults have difficulty discriminating phonetic contrasts not used systematically in their native language. Interestingly, although the effects of experience on perception are evident early in life (6 months to 1 year of age), these studies of second language learning show that these effects are also reversible, and that plasticity remains enhanced, for a relatively long period. Moreover, even a modest amount of exposure to a language in early childhood has been shown to produce a more native-like perception of its syllable contrasts in adulthood (Miyawaki *et al.*, 1975). Consistent with the idea that perception as well as production is altered, brain mapping studies show that different cortical areas of the brain are activated by the sound of native and second languages when the second language is learned later in life, whereas similar brain regions are activated by both languages if the two are learned early (e.g. Kim *et al.*, 1997). As suggested earlier, perceptual learning may in fact constrain which sounds can be correctly produced. Regardless of whether perception or production is primary, both production and perception of phonology, as well as grammar and prosody, provide strong data in support of sensitive periods for speech.

SONGBIRDS

It has long been realized that songbirds have a restricted period for memorization of the tutor song (Thorpe, 1958; Marler, 1970b). Now that studies of humans show that early perceptual capacities narrow with experience, the parallels between songbird and human critical periods are even more compelling. Despite numerous anecdotal accounts, the number of carefully studied songbird species remains small. The classical study is that of the white-crowned sparrow by Marler (1970b), which shows that as in humans, sparrows have an early phase of extreme plasticity (around 20–50 days of age) with a later gradual decline in openness, with some acquisition possible up to 100 or 150 days (Nelson *et al.*, 1995). After the age of 100–150 days, in most cases, birds did not learn new songs from sensory exposure to new tutors, regardless of whether they had had normal tutor experience or had been isolated. Birds with a classical critical period like this are often called closed learners. Some birds are open-ended learners: that is, their ability to learn to produce new song either remains open or reopens seasonally in adulthood (e.g. canaries [Nottebohm *et al.*, 1986] and starlings [Chaiken *et al.*, 1994; Mountjoy and Lemon, 1995]), although it is still unclear in many cases whether the reopening is sensory or motor in nature. Comparing the brains of these birds with those of closed learners should provide an opportunity to elucidate what normally limits the capacity to learn.

Does the capacity to produce sounds also have a critical period, independent of sensory exposure? That is, if correct motor learning is not accomplished by a certain age, despite timely sensory exposure, or is not closely linked in time with perceptual learning, can it ever be completed or corrected? The studies of tracheostomized children suggest that vocal motor learning may indeed also be developmentally restricted, but this is another question more easily addressed in songbirds than in humans. Songbird experiments provide conflicting evidence, however. One line of evidence comes from hormonal manipulations of birds. Singing of adult male birds is enhanced by androgen, and castration markedly decreases (but does not eliminate) song output in adult birds. Sparrows castrated as juveniles learn from tutors at the normal time, and produce good imitations in plastic song, but fail to crystallize song (Marler *et al.*, 1988a). When given testosterone as much as a year later, however, these birds then rapidly crystallize normal song, which suggests that the transition from plastic to more stereotyped crystallized song does not have to occur within a critical time window. These experiments do not perfectly address the question of a critical period for sensorimotor learning, however, because all young birds vocalized somewhat around the normal time of song onset, giving them some normal experience of sensorimotor matching. Similarly, a castrated chaffinch that had not sung at all during its first year still developed normal song when given hormone later (Nottebohm, 1981). In both these experiments, the absence of androgen, which dramatically decreases singing, might also have delayed motor development and motor sensitive period closure.

Other nonhormonal manipulations suggest that disruptions of motor learning at certain ages are in fact critical. For example, song lateralization primarily involves motor learning and production. Although this lateralization seems to have quite different mechanisms than that of speech, it shares with speech an early sensitive period for recovery from insults to the dominant side. Left hypoglossal dominance in canaries can be reversed if the left tracheosyringeal nerve is cut or the left HVC lesioned prior to the period of vocal motor plasticity when song production is learned, but not thereafter (Nottebohm *et al.*, 1979). This provides evidence that at least in canaries some organization occurs during motor practice that cannot be reversed later. An experiment to address this more directly might be to eliminate or disrupt all vocal normal practice until the usual time of crystallization and then to allow the birds to recover. Recent experiments in songbirds with transient botulinum toxin paralysis of syringeal muscles in zebra finches during late plastic song do suggest that critical and irreversible changes occur during late sensorimotor learning (Pytte and Suthers, 1996).

TIMING AND THE ROLE OF EXPERIENCE: WHAT CLOSES THE SENSITIVE PERIOD?

The question raised by the data on the difficulties of late learning is: what accounts for differential learning of language at different periods in life? By the classical critical period argument, it is time or development that are the important variables. Late experience has missed the window of opportunity for language learning, making it more difficult, if not impossible, to acquire native language patterns of listening and speaking, or of normal birdsong. This time-limited window presumably reflects underlying brain changes and maturation, which are as yet poorly understood, especially in humans. Lenneberg (1967) thought that at puberty the establishment of cerebral lateralization was complete and that this explained the closing of the sensitive period. The data reviewed in the previous section suggest, however, that the capacity for speech learning declines gradually throughout early life, or at least has several phases prior to adolescence. More striking, both song and speech studies are increasingly converging on a role for learning and experience itself in closing the critical period, as described below.

HORMONES

The approximate coincidence of puberty with closure of the sensitive period points to hormones as some of the maturational factors that limit learning. Surprisingly little has been done to examine this, however, for instance by comparing second language acquisition in boys and girls, or by investigating language development in human patients with neuroendocrine disorders (McCardle and Wilson, 1990). Because dyslexia and stuttering are 10 times more common in boys than in girls, testosterone has been hypothesized to play a role in some forms of dyslexia (Geschwind and Galaburda, 1985), but by their nature, studies of language disabilities may not address normal learning. Recent imaging data suggest that lateralization for speech is less strong in human females than in males (Shaywitz *et al.*, 1995), although whether the origin of this difference is hormonal is unclear, as is its relationship to critical period closure.

Male songbirds provide much more evidence for hormonal effects on learning. The earliest studies of song learning showed that the period of maximum sensitivity was not strictly age dependent but could be extended by manipulations (such as light control or crowding) that also delayed its onset (Thorpe, 1961). Just as in humans, these manipulations suggested a role for hormones, especially sex steroids, in closure of the critical period. This idea was further strengthened by work by Nottebohm (1969). He found that a chaffinch castrated in its first year, before the onset of singing, did not sing and subsequently learned a new tutor song in the second year, when it received a testosterone implant. Although this experiment did not indicate what ended the readiness to learn song, it certainly showed that it could be extended. Because singing often begins in earnest around the time that testosterone rises, and because it can be delayed or slowed by castration, a reasonable possibility is that these developmental increases in male hormones to a high level are also involved in closing the critical period. In an experimental manipulation to test this hypothesis, Whaling and colleagues (1998) castrated white-crowned sparrows at 3 weeks of age and then tutored them long after the normal 100-day close of the critical period. There was a small amount of learning evident in some animals subsequently induced to sing by testosterone replacement, which suggests that the critical period had indeed been extended. The effect was weak, however, perhaps indicating that a single hormone is unlikely to control normal learning.

ACTIVITY-DEPENDENCE: ADEQUATE SENSORY EXPERIENCE OF THE RIGHT TYPE

Although there is much to support a timing or maturational explanation for loss of the capacity for vocal learning, an alternative account is emerging in both humans and songbirds, which suggests that learning itself also plays a role in closing the critical period. In humans, this alternative account has been developed at the phonetic level, where the data suggesting a sensitive period are strongest (Kuhl, 1994). As described earlier, work on the effects of language experience suggest that exposure to a particular language early in infancy results in a complex mapping of the acoustic dimensions underlying speech. This warping of acoustic dimensions makes some physical differences more distinct whereas others, equally different from a physical standpoint, become less distinct; this may facilitate the perception of native-language phonetic contrasts and appears to exert control on how speech is produced as well (Kuhl and Meltzoff, 1997).

By this hypothesis, speech maps of infants are incomplete early in life, and thus the learner is not prevented from acquiring multiple languages, as long as the languages are perceptually separable. As the neural commitment to a single language increases (as it would in infants exposed to only one language), future learning is made more difficult, especially if the category structures of the primary and secondary languages differ greatly (for discussion see Kuhl, 1998). In this scenario, for example, the decline in infant performance is not due to the fact that American English /r/ and /l/ sounds have not been presented within a critical window of time, but rather that the infant's development of a mental map for Japanese phonemes has created a map in which /r/ and /l/ are not separated. This effect of the learning experience could be thought of as operating independently of, and perhaps in parallel with, strict biological timing, as stipulated by a critical period. By analogy to studies in other developing systems, this model might be called experience-dependent.

This view of early speech development incorporates some of the new data demonstrating that children with dyslexia, who have language and reading difficulties and are past the early phases of language development, can nonetheless show significant improvements in language ability after treatment with a strategy that assists them in separating sound categories (Merzenich *et al.*, 1996; Tallal *et al.*, 1996). These children and others with language difficulties (Kraus *et al.*, 1996) often cannot separate simple sounds such as /b/ and /d/. By the activity-dependent model, these children have either not been able to separate the phonemes of language and thus have not developed maps that define the distinct categories of speech, or they have incorrect maps, producing difficulties with both spoken language and reading. The treatment was computer-modified speech that increased the distinctiveness of the sound categories and may have allowed the children to develop for the first time a distinct and correct category representation for each sound, and to map the underlying space. Although children were doing this well after the time at which it would have occurred normally in development, their ability to do so may have depended on the fact that they had not previously developed a competing map that interfered with this new development. This hypothesis suggests that even dyslexic adults might benefit from such treatment, if the lack of normal mapping effectively extended the critical period.

Another test of the experience-dependent hypothesis for critical period closure might be to study congenitally deaf patients, not exposed to sign language, who have been outfitted with cochlear implants at different ages. If the critical period closes simply because of auditory input creating brain maps for sound, the complete absence of input might leave the critical period as open in 8- or 18-year-olds as in newborns. Alternatively, if some maturational process is also occurring, and/or if complete deprivation of inputs has negative effects, the critical period might close as usual, or be extended, but not indefinitely. To date, insufficient data are available to address these issues because cochlear implants of excellent acoustical quality have only recently become available, and relatively few children have been implanted (Owens and Kessler, 1989). However, even though deaf children who learn sign language at different ages are presumably not mapping any other languages prior to acquiring ASL, their decreasing fluency in ASL as a function of the age of learning does suggest that the capacity to learn shows at least some decline with age, even without competing sensory experience (Newport, 1991).

The end of the sensitive period may not be characterized by an absolute decrease in the ability to learn but rather by an increased need for enhanced and arousing inputs. In other systems, such as the developing auditory-visual maps of owls, the timing and even the existence of sensitive periods have been found to depend on the richness of the animal's social and sensory environment (Brainard and Knudsen, 1998). In speech development, the inputs provided by adults who produce exaggerated, clear speech ("parentese" or infant-directed speech) when speaking to infants may be crucial. This speech, which provides a signal that emphasizes the relevant distinctions and increases the contrast between phonetic instances, could be related to the kind of treatment that is effective in treating children with dyslexia. This raises the possibility, for example, that Japanese adults might also be assisted in English learning by training with phonemes that exaggerate the differences between the categories /r/ and /l/. These adults have a competing map, but exaggerated sounds might make it easier to create a new map that did not interfere or could coexist with the original one formulated for Japanese Studies also show that training Japanese adults by using many instances of American English /r/ and /l/ improves their performance (Lively *et al.*, 1993). Thus, both exaggerated, clear instances and the great variability characteristic of infant-directed speech may promote learning after the normal critical period.

In the songbird field, it has been known for some time that the nature of the sensory experience affects the bird's readiness to learn song. For instance, early exposure to conspecific song gradually eliminated the willingness of chaffinches to learn heterospecific song, or conspecific song with unusual phrase order (Thorpe, 1958). Similarly, birds born late in the breeding season of a year, when adults have largely stopped singing, were able to acquire song later than siblings born earlier in the season and thus exposed to much more song (Kroodsma and Pickert, 1980). More specific demonstrations that the type of auditory experience can affect or delay the closure of the critical period come from studies of other species, especially zebra finches. Immelmann (1969) and Slater *et al.* (1988) showed that zebra finches tutored with Bengalese finches were able to incorporate new zebra finch tutors into their songs at a time when zebra finches reared by conspecifics would not. This suggested that the lack of the conspecific input most desirable to the brain left it open to the correct input for longer than usual. Even more deprivation, by raising finches only with their nonsinging mothers or by isolating them after 35 days of age, gives rise to finches that will incorporate new song elements or even full songs when exposed to tutors as adults (Eales, 1985; Morrison and Nottebohm, 1993). This is reminiscent of activity dependence in other developing systems, such as the visual system, in which a lack of the appropriate experience can delay closure of the critical period. Although unresolved in birds, it seems likely that the critical period can be extended in this way, although not indefinitely (except perhaps in open learners).

A caveat is that it may not be the sensory experience but the motor activity associated with learning (or, as always, both the sensory and motor activity interwoven) that decreases the capacity to learn. This has been little studied in humans, but in chaffinches, crystallization was associated with the end of the ability of the birds to incorporate new song (Thorpe, 1958). This is also a possibility suggested by Nottebohm's experiment, in which castrated birds that had not yet sung were still able to learn new tutor song. Because testosterone induces singing, perhaps it is not a direct effect of hormones that closes the critical period, but some consequence of the motor act of singing. To dissociate these possibilities will require more experiments, because under normal conditions androgens invariably cause singing and song crystallization (Korsia and Bottjer, 1991; Whaling *et al.*, 1995). Zebra finches raised in isolation do incorporate at least some new syllables as adults even though they have already been singing (isolate) song (Morrison and Nottebohm, 1993; Jones *et al.*, 1996); these studies do not settle the issue, however, because the birds that showed the most new learning were also the least crystallized (Jones *et al.*, 1996).

SOCIAL FACTORS

Closure of the critical period is also affected by social factors. Although young white-crowned sparrows learn most of their conspecific song from either tapes or live tutors heard between days 14 and 50, Baptista and Petrinovich (1986) showed that these birds will even learn from a heterospecific song sparrow after 50 days of age if they are exposed to a live tutor. In zebra finches, social factors interacting with auditory tutoring may explain some of the conflicting results on whether and for how long the critical period can be kept open: birds raised with only their mothers showed extended critical periods, whereas birds raised with both females and (muted) males, or with siblings, did not show late learning (Aamodt *et al.*, 1995; Volman and Khanna, 1995; Wallhausser-Franke *et al.*, 1995). Jones *et al.* (1996) directly tested the effect of different social settings on learning in finches. They showed that major changes of song in adulthood were rare and were found only in the more socially impoverished groups. It will be crucial to try to tease social and acoustic factors apart. Although the neural mechanisms of social factors (perhaps hormonal in nature) remain unclear, their effects are certainly potent: merely the presence of females caused males to have larger song nuclei than males in otherwise identical photoperiodic conditions (Tramontin *et al.*, 1997).

In both songbirds and humans, it seems likely that a number of factors act in concert to gradually close the critical period, just as a number of factors control the selectivity of learning. Maturation, auditory experience, social factors, and hormones (which could be the basis for the maturational or social effects) can all be shown to affect the onset and offset of learning. When learning occurs in normal settings, these factors all propel learning in the same direction. When some or all of these factors are disrupted, the critical period can be extended, although probably not indefinitely.

CONCLUSIONS

Recurrent themes emerge when the comparable features of birdsong and speech learning are studied: innate predispositions, avid learning both perceptually and vocally, critical periods, social influences, and complex neural substrates. The parallels are striking, although certainly there are differences. Both the commonalities and the differences point to the gaps in our knowledge and suggest future directions for both fields.

The grammar and other aspects of meaning in human speech are the most obvious differences between birdsong and speech. These differences suggest that although human speech is undoubtedly built on pre-existing brain structures in other primates, there must have been an enormous evolutionary step, with convergence of cognitive capacities as well as auditory and motor skills, in order to create the flexible tool that is language. In contrast, it seems a smaller jump from the suboscine birds that produce structured song but do not learn it, to songbirds. Nonetheless, a critical step shared by avian vocal learners with humans must have been to involve the auditory system, both for learning of others and for allowing the flexibility to change the vocal motor map. In fact, the existence of closely related nonlearners as well as of numerous different species that learn is one of the features of birdsong that has allowed more dissection of innate predispositions than is possible with humans. It may seem that more is prespecified in songbirds, with their learning preferences and isolate songs. Many of the analogous experiments, however, cannot be done or simply have not yet been done in humans, for instance examining whether newborn monkeys and humans prefer conspecific sounds over other vocalizations, and if so, what acoustic cues dictate this preference. Neurophysiological analysis of high-level auditory areas in young members of both groups (using microelectrodes in songbirds and perhaps event-related potentials in humans), and comparisons with nonhuman primates and other birds, should provide insight into what the brain recognizes from the outset, how it changes with experience, and how it differs in nonlearners.

The early perceptual learning in both humans and songbirds seems different from many other forms of learning: it does not require much if any external reinforcement and it occurs rapidly. What mechanisms might underlie this? In songbirds it is known that songs can be memorized with just a few experiences, whereas in humans this area is as yet unexplored. Although human vocal learning seems to be rapid, it is not known if it takes 10 minutes or 4 hours a day to induce the kind of perceptual learning seen in infants, or whether the input has to be of a certain quality or even from humans. Both groups seem to have enormous attentiveness to the signals of their own species and, in most cases, choose to learn the right things. This could be due to a triggering of a prespecified vocal module, as has often been suggested, it could be that auditory or attentional systems have innate predispositions that guide them, or it could simply be that learning in each case is specific to sounds with some regular feature that is as yet undiscovered. The learned "template" of songbirds continues to be much sought after, but it is clear from numerous behavioral studies that the idea of a single sensory template is much too simple: Many birds memorize and produce multiple songs, at least during song development. This presumably means that they have multiple learned templates, or perhaps some more complex combinatorial memory mechanism. And what mediates their ultimate selection of a subset of those songs as adults? It may be guided by a combination of genetics and experience, including potent social effects of conspecifics. Social effects on learning seem crucial in humans as well, but in both cases how social influences may act on the brain is poorly understood. Understanding how to mobilize these, however, could have profound implications for treatment of communication disorders, at any age.

Our understanding of neural substrates of both speech and birdsong should continue to improve as methods for exploring the brain advance, although the study of an animal model such as songbirds will always have a certain advantage. The question of how auditory feedback acts during vocalization is a persisting and important puzzle shared by both fields, and more insight into this would shed light not only on general issues of sensorimotor learning but also on human language disabilities such as stuttering. Finally, brain plasticity and the critical period remain fascinating and important issues. What is different at the neural level about language learning before and after puberty? How does the brain separate the maps of the sounds of languages such as English and Japanese in infants raised in bilingual families? Understanding what governs the ability to learn at all ages may not only advance our basic knowledge of how the brain changes with time and experience, but could also be of practical assistance in the development of programs that enhance learning in children with hearing impairments, dyslexia, and autism, and might aid in the design of programs to teach people of any age a second language. Clearly, studies of songbirds with different types of learning have a remarkable potential to reveal possible neural mechanisms underlying the maintenance and loss of brain plasticity, although this area is as yet largely unmapped. These issues all raise more questions than they answer, but research in both fields is progressing rapidly. Continuing to be aware of and to explore the parallels, as well as admitting when they fail, should be helpful to both fields.