

THE FUNCTION AND SIGNIFICANCE OF INTER-SPECIES ACOUSTIC CUES IN THE TRANSFORMATION OF BUDGERIGAR (*MELOPSITTACUS UNDULATUS*) SOUNDS INTO “SPEECH”

Jim Scanlan

University of New England, AUSTRALIA

ABSTRACT: Analysis of speech-imitation sounds produced by pet budgerigars (*Melopsittacus undulatus*) reveals a consistent focus on acoustic components of speech which have counterparts in the species-specific vocalizations of budgerigars. These budgerigar vocalizations include whistle sequences (which, with their rapid glides in pitch contour, need only slight modification to constitute a second-formant representation of speech), the contact call (which is acoustically similar to the second-formant transition of high front vowels), and sounds with harmonic spectra (which can be modified to represent the formant structure of certain vowels). This transformation of species-specific sounds into speech-imitation sounds by focusing on shared acoustic features prompts the hypothesis that, with the appropriate social stimulation, pet budgerigars perceive human speech as modified budgerigar sounds. The hypothesis is supported by the fact that the shared acoustic features are particularly important cues in the perception of species-specific communication sounds by both humans and budgerigars. Such inter-species vocal communication cues, having a common origin somewhere in vertebrate phylogeny, would help to explain the many reported examples of human-like speech perception by nonhuman vertebrates. The shared neural mechanisms which correspond to these shared acoustic patterns could constitute a phylogenetically conservative level of auditory perception which is communication-sound-specific but not species-specific.

INTRODUCTION

Some birds, when kept as pets, copy the communication sounds (i.e. speech) of their human companions. Although it has been valued as an entertaining curiosity for thousands of years, this remarkable behaviour has received relatively little scientific attention. Such attention is warranted, however, because copies of speech sounds by birds, when subjected to acoustic analysis, open new perspectives on

Address correspondence to Dr Jim Scanlan, c/- School of Biological Sciences, University of New England, Armidale, NSW 2351, Australia.

avian vocal communication. Avian "speech" (by which term I refer both to speech-copying behaviour and to its acoustic manifestations) represents human speech after it has been filtered through an avian perceptual system, incorporated into an avian behaviour pattern, and reproduced by means of an avian vocal mechanism. It can thus provide a unique vantage point for the study of those perceptual, behavioural and vocal mechanisms.

Also, by presenting us with sounds which we perceive as "speech" but which differ, physically, from speech in a variety of ways, it can provide insight into aspects of our own communication sounds which are inaccessible to the self-scrutiny of conventional linguistics and phonetics. Finally, in demonstrating that congruent vocal mechanisms are not necessary for acoustic communication between two individuals, it emphasises the predominance of auditory over articulatory factors in vocal communication generally.

This paper will argue, on the basis of acoustic data and behavioural observations, that speech-trained, pet budgerigars (*Melopsittacus undulatus*) perceive speech sounds as modifications of species-specific budgerigar sounds, and are able to copy those "modifications" in communicative behaviour directed towards their human companions. This behaviour is facilitated principally by three factors: communicatively significant acoustic similarities between budgerigar sounds and speech (explained below); the budgerigar's adoption of those similar - albeit "modified" - communication sounds because of social bonding with a human being; the budgerigar's ability to continue modifying its communication sounds throughout life (Farabaugh *et al.*, 1994; Brittan-Powell *et al.*, 1997) to match those of its social companions.

The behavioural and acoustic evidence that birds perceive non-species-specific communication sounds (e.g. those of human speech) as if they were species-specific suggests that the former are being processed by neural pathways normally dedicated to the perception of the latter. This could be facilitated by acoustic patterns and corresponding neural mechanisms which are common to the vocal communication systems of birds and humans.

The published literature has established a rigidly dichotomous perceptual model according to which, at any particular point along the human auditory pathway, neural processing of speech is either "acoustic" or "phonetic". Within the framework of this model, human-like processing of some speech sounds by nonhuman animals is adduced as evidence that aspects of speech perception previously thought to be "phonetic" must, in fact, be "acoustic". This paper is based on analyses of budgerigar "speech" which suggest that the birds

perceive certain acoustic features of speech as being functionally identical to acoustically similar features of budgerigar sounds. It hypothesises an intermediate level of analysis which is communication-sound-specific but not species-specific. Further testing of the hypothesis could help us to escape between the horns of the dilemma posed by the artificial “acoustic-phonetic” dichotomy. It could also lead to new insights into the evolution of vocal communication - including human speech.

The aims of the paper are thus to present some unprecedented observations of budgerigar “speech”, to propose an explanatory hypothesis and to show how that hypothesis relates to current knowledge and theory. In pursuing the last-mentioned aim, the remaining Introduction selectively draws on the extensive literature relating to avian vocal communication and the Discussion draws on the even more extensive speech-perception literature. While neither section is intended as an exhaustive review, some detailed discussion of this literature was considered necessary.

Acoustic similarities

Bertram (1970) and Nottebohm (1976) drew attention to acoustic similarities between human speech sounds and certain species-specific vocalizations of the Indian hill mynah (*Gracula religiosa*) and the orange-winged Amazon parrot (*Amazona amazonica*) respectively. Each author associated these similarities with the ability of his subject species to imitate speech sounds. In both cases, the speech-like species-specific vocalizations are used in “close-range social contact” (Bertram, 1970), and their acoustic structure resembles that of human vowels in their low-frequency periodicity and apparent formant structure.

Some calls and song syllables in the complex vocal repertoire of the budgerigar display these and other speech-like features. The budgerigar repertoire consists of at least 11 (and possibly as many as 14) call types (Higgins, 1999, p. 515) as well as the complex and variable vocal pattern known as “warble” (Brockway, 1964, who distinguished between “loud warble” and “soft warble”) or “warble song” (Farabaugh *et al.*, 1992). The latter term is used here, as it emphasises the important fact that this vocal pattern is in some respects analogous to passerine song. Indeed, Ferrell and Baptista (1982) maintain that budgerigar warbling is “functionally equivalent” to passerine song. Warble song consists of call notes as well as song-specific syllables (Wyndham, 1980; Farabaugh *et al.*, 1992; Brittan-Powell *et al.*, 1997), providing psittacine support for the theory that bird song evolved from calls (Thorpe, 1958; Thielcke, 1966; Catchpole, 1979).

Brockway (1964) described "loud warble" as "the chatter so commonly associated with this species". The use of the word "chatter" suggests the speech-like quality of budgerigar warble song, and the spectrograms of Farabaugh *et al.* (1992) and Wyndham (1980) confirm this similarity. Some warble song syllables are broad-bandwidth sounds with formant-like concentrations of energy - particularly in the region of human second-formant frequencies (about 800-2,500 Hz).

Behavioural parallels between warble song and speech imitation

Brockway (1964), Farabaugh *et al.* (1992) and Wyndham (1980) all remark on the highly variable nature of warble song. In spite of its wide range of syllable types and the high variability of its sequencing, Farabaugh *et al.* (1992) detected the possible existence of combinatorial rules and therefore of a learnt "syntax" in the structure of warble song.

Warble song is primarily a courtship vocalization (Brockway, 1964; Wyndham, 1980; Farabaugh *et al.*, 1992), but can be used in other forms of social interaction (Farabaugh *et al.*, 1992). Wyndham's (1980) observations, and the studies of Farabaugh *et al.* (1992, 1994), indicate that warble song contains cues for individual (and group) identification. Warble song is most commonly performed by males, but can also be performed by females (Ferrell and Baptista, 1982). Its performance is highly "mimetic" (Brockway, 1964; Ferrell and Baptista, 1982): one budgerigar is stimulated to warble by hearing another. However, a warbling performance is not obviously directed at any other individual (Brockway, 1969).

Vocal imitation plays an important role in warble song acquisition and modification, with social factors influencing the pattern of imitative learning. Farabaugh *et al.* (1992) found that males in the same social group shared a significantly greater percentage of syllable types than males in different social groups. The contact call, which is one of the most common components of warble song, is both learnt from adults and matched among siblings. Even if the adult model is aberrant (due to deafening) or of another species (e.g. zebra finch, *Taeniopygia guttata*), its influence is evident in the contact calls of juvenile budgerigars (Brittan-Powell *et al.*, 1997).

These characteristics of the budgerigar's warble song have their parallels in the "verbal" (i.e., speech-imitation) behaviour of the birds. Male budgerigars are recognised by fanciers as the best "talkers", but females are capable of imitating speech. Verbal behaviour in this species is usually stimulated by human chatter or noise in the environment; thus it is "mimetic" in Brockway's sense, as well as being

imitative in its use of vocal copying. The budgerigars recorded during the present study showed, when “talking”, no interactive orientation towards any human companion; their verbal utterances were seldom in direct response to utterances of a human speaker.

In contrast to the verbal behaviour of budgerigars, that of some other “talking” species, such as some larger parrots and the Indian hill mynah, is often an overt interaction with human companions. Whereas the budgerigar, once in the necessary psycho-physical state of arousal, tends to run right through the verbal repertoire in a fixed order, these other species are more inclined to produce particular verbal utterances as a direct and specific response to verbal or other environmental stimuli. I have argued [Scanlan, 1988] that this behavioural distinction reflects the modelling of budgerigar “speech” on *singing* behaviour, in contrast to the modelling of mynah “speech”, or that of the larger parrots, on *calling* behaviour.

The following three sections discuss the ontogeny and neuroethology of learnt avian vocalizations with relation to budgerigar “speech” and human speech. For a review of the literature relating to ontogenetic and neuroethological correspondences between human speech and oscine song, see Doupe and Kuhl (1999).

The development of avian “speech”: ontogeny of song in birds parallels that of speech in humans

The two *Melopsittacus undulatus* subjects of this study had been carefully trained by their owners to repeat a fixed repertoire of phrases and verses. Both these birds, before entering upon their repertoires, would go through a preparatory performance which began as species-specific warble song and included an ever-increasing number of speech-like sounds until the repertoire itself was begun. This sequence, which I have called “pre-performance warble” (Scanlan, 1988), contains speech phrases and syllables which are not part of the performance repertoire. This conforms with the eclectic nature of warble song as described by Brockway (1964), Wyndham (1980), and Farabaugh *et al.* (1992). It also suggests a possible connection with the plastic song of songbirds, which includes a wide variety of vocalizations and, in some cases, imitations of other species. (See Thorpe, 1955, and Marler, 1956, although these early studies had not yet distinguished between “plastic song” and the ontogenetically earlier “subsong”.) Physically, oscine plastic song as presented in the literature has a broader - and therefore more speech-like - spectral envelope than full song. (See spectrograms in, for example, Thorpe and Pilcher, 1958, and Marler and Peters, 1982.) Like warble song and speech imitation in budgerigars, it is

mainly performed by males, but can be performed by females.

The "pre-performance" or speech-like warble of pet budgerigars has its parallels in the vocal behaviour of other "talking" species. Groups of captive Indian hill mynahs (Foss, 1964) and European starlings, *Sturnus vulgaris* (West *et al.*, 1983), exposed daily to the speech of their human carers, developed vocal timbres and cadences which resembled those of speech.

Many authors, including Charles Darwin in *The Descent of Man* (1871), have compared subsong and plastic song in songbirds with babbling in human infants. (See, e.g., Thorpe, 1961; Kuhl, 1989; Locke, 1993.) Menyuk and Menn (1979), in discussing the babbling of infants, describe the boundary between babble and speech as "in general a fuzzy one". At this stage in the child's development, they say, there are "recurrent entities" in a child's production that cannot be classified unequivocally as either "babble" or "speech". In some children these entities "occupy a pivotal position in language development, and careful consideration of their form and function illuminates the nature of the transition from babble to speech in the general case" (Menyuk and Menn, *ibid.*). This paper examines some analogous "entities" in the pre-performance warble of "talking" budgerigars. These "entities" are, in their acoustic structure, transitional between species-specific sounds and speech-imitation sounds. In this, as well as in their loose overall organization and their preparatory, exploratory function, they resemble avian plastic song. Thus they provide clues to the motivation and the mechanisms involved in the transformation of budgerigar sounds into "speech". (The word "transformation" in this paper refers both to modifications of acoustic structure, and to the - much more dramatic - change in the corresponding aural impression.)

Social and sequential factors involved in avian vocal learning

At least some songbirds - e.g. zebra finches (Slater *et al.*, 1988) and white-crowned sparrows, *Zonotrichia leucophrys* (Marler, 1970) - appear to be genetically predisposed to learn species-specific song, but cross-fostering experiments with both finches (Immelmann, 1969; Eales, 1987, 1989) and white-crowned sparrows (Baptista and Petrinovich, 1984, 1986) have demonstrated that any such predisposition can be overridden by social factors. Furthermore, in laboratory finches, social interaction - visual as well as vocal - between tutor and pupil is necessary for song learning to occur with a precision comparable to that observed in the wild (Eales, 1989).

Although, in captivity, some birds can learn species-specific song from tape recordings (Marler and Tamura, 1964; Adret, 1993), the

learning of *species-specific* calls and songs by a bird typically involves a familial (e.g. Clayton, 1987), social (e.g. Payne, 1981) or antagonistic (e.g. Baptista and Petrinovich, 1984) relationship with the model/tutor. In the earlier study on which this paper is based (Scanlan, 1988), it was argued that learning of *non-species-specific* vocalizations by birds is of two distinct kinds - that in which there is such a relationship with the model, and that in which there is no such relationship. In the former case, the overall behaviour pattern (including ontogeny, neurophysiology, and social context) is exactly the same as that in the learning of species-specific sounds and should thus be referred to by the same term ("vocal imitation"). The term "vocal mimicry", then, should be restricted to those instances of non-species-specific vocal copying in which no familial/social/antagonistic relationship is involved (and in which, as a corollary, ontogenetic and neurophysiological patterns differ from those associated with "imitation"). The argument of this paper involves classifying speech imitation by budgerigars as "vocal imitation".

Social factors determine the species-specific call-learning behaviour of budgerigars. Farabaugh *et al.* (1994) found that the contact calls of unrelated budgerigars converged on a common acoustic pattern soon after they had been confined together in the same cage. (Nowicki, 1989, reported a similar result with the flock-recognition call of the black-capped chickadee, *Parus atricapillus*.) Evidence presented below will demonstrate that modification of the contact call is an important ingredient in the process of a budgerigar's transforming species-specific sounds into "speech".

Early experiments on song learning in captive birds (e.g. white-crowned sparrows - Marler, 1970, and zebra finches - Immelmann, 1969) established the concept of a "sensitive period", after which normal learning could not occur. However, further experimental work with both these species (white-crowned sparrows - Baptista and Petrinovich, 1984, 1986; zebra finches - Eales, 1985, 1987) made it clear that the sensitive period was not immutable, but could be extended if a suitable model had not been found by the end of the usual song-learning period.

There is a growing body of evidence that, in the wild, vocal learning in some birds is open-ended - i.e., it is not restricted to a sensitive period early in life. (See reviews in Mundinger, 1982; Farabaugh *et al.*, 1994; Doupe and Kuhl, 1999. The learning of novel vocalizations should be distinguished from the process of selective matching with an early-acquired set of models revealed, for example, in the white-crowned sparrow by Nelson and Marler, 1994.) In terms of the traditional distinction between calls and song (Thorpe, 1958;

Thielcke, 1966; Nottebohm, 1972), open-ended learning strategies seem to apply particularly to call learning (Farabaugh *et al.*, 1994). However, modification of song in adulthood has been documented in several species - e.g. in the canary, *Serinus canarius* (Nottebohm and Nottebohm, 1978), and especially in species in which song has an important social as well as sexual/territorial function, such as the European starling, (Adret-Hausberger *et al.*, 1990; Chaiken *et al.*, 1994; Mountjoy and Lemon, 1995), the Australian magpie, *Gymnorhina tibicen* (Brown *et al.*, 1988), and the Indian hill mynah (Bertram, 1970). Even in species - such as the zebra finch - in which song learning normally occurs during a clearly-defined period (Immelmann, 1969), learning can be delayed indefinitely in captivity if no suitable model is available during the usual sensitive period (Eales, 1985).

Farabaugh *et al.* (1994) showed that the learning and modification of budgerigar contact calls is an open-ended process. An earlier study (Farabaugh *et al.*, 1992) had shown that contact calls - and other calls - are important components of budgerigar warble song. To this extent at least, then, warble song is subject to learnt modification throughout adulthood. The extent of such modification to the highly variable overall structure - both acoustic and syntactic - of budgerigar warble song is as yet unknown. It will be argued below, however, on the basis of acoustic analyses presented here, that speech imitation is a learnt modification of warble song. Budgerigar "speech" would thus demonstrate this species' ability to modify its song extensively and open-endedly in adjusting to a novel social environment.

In songbirds that modify their song repertoires as adults there is evidence that plastic song occurs in adulthood (reviewed in Margoliash *et al.*, 1991), suggesting a connection between plastic song and not only song learning in juveniles but also song modification in adults (Marler and Peters, 1982; Margoliash *et al.*, 1991). European starlings have a "warbling song" (Adret-Hausberger and Jenkins, 1988) which appears to be analogous to budgerigar warble song in both structure and function. Hausberger *et al.* (1991) remarked on the plastic-song-like nature of the starling warble song; the same could be said of the warble song of budgerigars. It will be shown below that every time a budgerigar performs a speech-imitation sequence it does so by modifying some of its species-specific sounds. The association of each "speech" sequence with a sequence of "pre-performance warble" thus emphasises the analogy between "pre-performance warble" in speech-trained budgerigars and plastic song in songbirds. Considered as a whole, a bout of pre-performance warble by a pet budgerigar, which becomes increasingly more speech-like throughout, seems to reproduce, in a condensed form, the "speech"-learning process in the bird.

Neuroethological factors

Patterns of neural activity are fundamental ingredients of the behaviour patterns observed by ethologists. The integrated study of ethological and neural events has developed into the discipline of "neuroethology" (see, e.g., Suga, 1988; Brenowitz and Kroodsma, 1996). Over the past 20 years, researchers have uncovered neural activity patterns which form an integral part of audio-vocal communication behaviour in birds. In songbirds, some of these patterns are associated with neural pathways reserved for the process and/or the product of vocal learning. These pathways include feedback circuits from auditory to vocal telencephalic centres (Kelley and Nottebohm, 1979; Doupe and Konishi, 1991), pathways devoted to the production of learnt song components as distinct from unlearnt components (Simpson and Vicario, 1990), and an "anterior forebrain pathway" apparently crucial to the process of vocal learning (Bottjer *et al.*, 1984, 1985; Doupe and Konishi, 1991; Herrmann and Arnold, 1991; Mooney and Doupe, 1991; Scharff and Nottebohm, 1991; Doupe, 1997). The vocal control nuclei along these pathways contain auditory units which are maximally responsive to autogenous (i.e., the bird's own) song (Margoliash, 1983, 1986; Margoliash and Fortune, 1992; Vicario and Yohay, 1993; Doupe, 1997; Whaling *et al.*, 1997). This selectivity applies even if autogenous song is abnormal as a result of experimentally manipulated learning (Margoliash, 1983). Thus, in the learning and production of non-species-specific sounds, if environmental, social and metabolic aspects of the behaviour pattern are preserved, neural aspects of the pattern may also be preserved.

Physical similarities between the vocalizations of some avian species and human speech could facilitate this process. For example, Chew *et al.* (1996) found neurons in the caudomedial neostriatum of the zebra finch which respond strongly to human speech sounds as well as to species-specific calls. These neurons respond less strongly to the calls of other, related, songbird species. In this case, the strength of the response appears to be directly related to acoustic similarities between zebra finch calls and speech - similarities not shared by the alien bird sounds.

It can be hypothesized, therefore, that a "talking" songbird (e.g. a speech-trained Indian hill mynah) employs its specialised vocal-communication neural pathways in the perception of speech and the production of speech-like sounds and that consequently, in the vocal control nuclei of such a bird, there would be units maximally responsive to its own speech-imitation sounds.

Parrots differ from songbirds not only in the organization of

telencephalic vocal control centres, but also in the systems of projections from auditory to vocal centres (Striedter, 1994). The one parrot species examined so far - *Melopsittacus undulatus* - has three recurrent vocal-control pathways in the anterior forebrain (Durand *et al.*, 1997). These pathways may be, anatomically, either analogous to (Brauth *et al.*, 1994) or homologous with (Durand *et al.*, 1997) the anterior forebrain pathway in songbirds; in either case, the corresponding role of vocal learning in songbird and budgerigar ontogeny reinforces the conclusion that in budgerigars, too, they have a vocal-learning function (Brauth *et al.*, 1994; Durand *et al.*, 1997). Indeed, their greater neurological complexity in the budgerigar could reflect that species' greater flexibility in the acquisition of new vocalizations (Durand *et al.*, 1997). Preliminary reports indicate that Anna's hummingbird (*Calypte anna*) which, like at least some other hummingbirds (Gaunt *et al.*, 1994), develops its song through vocal learning (Baptista and Schuchmann, 1990), has forebrain nuclei similar in appearance to those of songbirds (Brenowitz and Kroodsma, 1996). Thus the hypothesis in the previous paragraph relating to songbirds may be extended to apply to budgerigars (as well as to other parrots and hummingbirds).

In summary, neurological data indicate the existence, in songbirds and budgerigars, of complex, specialised audio-vocal neural pathways for the processing of learnt communication sounds. They also indicate that these pathways can operate whether the learnt vocalizations are normal (i.e. species-specific) or abnormal (e.g. allospecific). The hypothesis explored in this paper is that there are acoustic cues, common to a range of vertebrate communication sounds, which, when experienced in conjunction with the necessary social stimuli, allow certain vocalizations - whether species-specific or allospecific - access to these special pathways. Experimental evidence for the existence of such acoustic cues is presented below.

METHODS

What follows is an analysis of several sequences from the pre-performance warble of two pet budgerigars (Budgerigar A and Budgerigar B - each trained by a different adult female owner) which illustrate the transformation of bird sounds into "speech". Each bird had been trained, over a number of years, by its owner's repeating a sequence of words and phrases (having a total duration, in each case, of 30-45 sec) in a stereotyped order and intonation pattern. In each case, too, training occurred in training sessions which were clearly defined by

time of day, the orientation of the owner in relation to the bird's cage (e.g. closer than at other times), and the distinctly pedagogical attitude of the owner during these sessions (e.g. demanding the bird's attention). Each bird was recorded in its cage in its owner's home, and each on a different day. On each of those days, the respective human trainer was recorded repeating her bird's repertoire in the manner (i.e., using the stereotyped intonation pattern) employed during training sessions.

Budgerigar A ("Lucky") was recorded with a Uher Report tape recorder, and Budgerigar B ("Joey") with a Uher CR 240 cassette recorder. The microphone used for both recordings was a Sony Electret condenser microphone, ECM-30. In both cases, the microphone was suspended from the wire forming one side of the bird's cage, and it took both budgerigars about one hour to become comfortable enough with this novel object (and the presence of the investigator) to resume normal vocalization.

Both birds began vocalization with sequences of "pre-performance warble" which led into repetitions of their repertoires; however, the distinction between pre-performance warble and performance proper was more marked in Budgerigar A than in Budgerigar B. The tape recorder was switched on near the beginning of each sequence of "pre-performance warble". Each recording session lasted about four hours: in that time, Budgerigar A produced seven clearly-defined sequences of "pre-performance warble" plus repertoire, and Budgerigar B several sequences, in each of which it was possible to distinguish an initial "pre-performance" and a subsequent "performance" component. Both owners confirmed that, despite the presence of the microphone and the investigator, the sequences were typical of the birds' vocalizations. The spectrograms were produced on a Kay Sona-Graph 6061B, using the 300 Hz bandwidth filter.

RESULTS

Harmonic spectra

A common syllable in budgerigar warble song is a strident-sounding array of harmonically related components. This sound is designated "brr" by Brockway (1964) and "weep" by Farabaugh *et al.* (1992), but in this paper I will continue to use my term "harmonic warble" (Scanlan, 1988), which is descriptive of its acoustic structure. Amplitude spectra of harmonic warble sounds confirm that their components are, indeed, harmonically related; moreover, the corresponding sound pressure waveforms, strikingly similar to those of

human vowels, indicate that the budgerigar could be using a sound-production mechanism similar to that of human phonation, as explained in my original study (Scanlan, 1988). In a non-speech-imitation context it has a fundamental frequency, producing an intense spectral peak, of 1-2 kHz (see spectrograms in Farabaugh *et al*, 1992) - i.e., the range of the second formant, rather than the fundamental, of human vowels. Within the pre-performance warble of the two budgerigars under investigation here, however, fundamental frequencies of harmonic warble varied between the 1-2 kHz typical of a non-speech-imitation context and frequencies below 1 kHz in syllables which sounded more speech-like. In other words, an initial mechanism employed by these budgerigars in transforming harmonic warble to "speech" is a simple lowering of fundamental frequency.

Figure 1, a sequence of Budgerigar B's pre-performance warble, demonstrates further stages in this transformation. This sequence shows a progression from speech-oriented, low-frequency harmonic warble to actual speech imitation in which the main factors in the transformation are the spectral and temporal qualities of the initial burst of noise, and the pitch contour of the succeeding vocalic segment. The harmonic warble syllable (labelled "1" at the top of the spectrogram) is followed by an intermediate stage (2), and then by the speech imitation proper (3). It is obvious from the spectrogram that the bird has modified 1 only minimally to produce the very different perceptual effect of 3. This, however, is not the only mechanism available to a budgerigar for the imitation of vowels: an alternative mechanism involves the amplitude modulation of a sinusoidal carrier (Scanlan, 1988; Banta Lavenex, 1999), the carrier and modulation frequencies being within the ranges of human second-formant and fundamental frequencies respectively (Scanlan, 1988).

While there is no formant pattern in 1, there are discernible spectral peaks in 2 and 3, the first vocalic segment of 3 having a "formant transition" from about 2,000 Hz to 1,000 Hz. These are frequencies appropriate for the transition from an initial palatal fricative [dʒ] to a mid vowel such as [ə]. (Compare the speech pattern of the bird's trainer in pronouncing the same sound sequence in Figure 2.) The fricative noise at the beginning of (3) has a similar frequency range (mainly between 2,000 Hz and 4,500 Hz), intensity, and duration (46 msec) to those in the same word - the bird's name, "Joey" - pronounced by his trainer (Fig. 2). The overall contour of the harmonic components has been modified in (3) to resemble the falling and then rising contour of F2 in Figure 2.

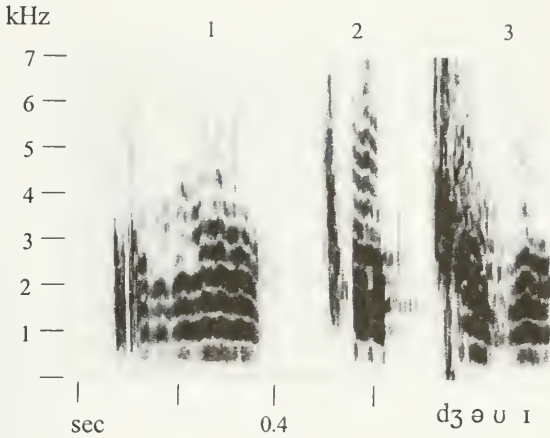


Figure 1. A sound sequence produced by Budgerigar B illustrating the transformation of a species-specific syllable (of the “harmonic warble” type - see text) into an imitation of the name Joey [dʒəuɪ].

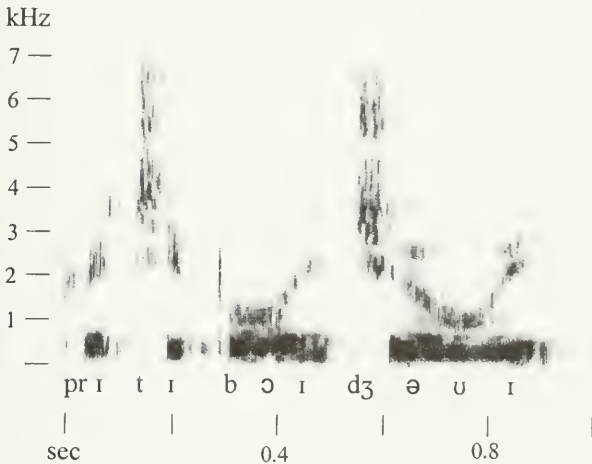


Figure 2. The phrase “pretty boy Joey” [prɪtɪ bɔɪ dʒəuɪ] produced by the woman trainer of Budgerigar B, for comparison with Figure 1 [dʒəuɪ], Figures 3 and 4 [prɪtɪ bɔɪ], and Figures 6 and 7 [bɔɪ].

Whistles and the second formant

Amplitude modulation is a common feature of bird sounds (Greenewalt, 1968; Stein, 1968), and the reported range of avian amplitude-modulation frequencies (90-300 Hz - Stein, *ibid.*) roughly corresponds with the range of fundamental frequencies in adult human speech (Lieberman, 1977). For human listeners, amplitude-modulation frequency is the acoustic correlate of perceived pitch (Ritsma, 1962; Schouten *et al.*, 1962).

Amplitude modulation is a dominant feature of both species-specific contact calls and speech-imitation sounds produced by budgerigars (Banta Lavenex, 1999). Budgerigars are more sensitive than humans to changes in amplitude-modulation rate (Dooling and Searcy, 1981). Furthermore, budgerigars are better able than humans to detect a tone of 2-4 kHz within a band of noise (Dooling and Saunders, 1975; Saunders *et al.*, 1979; Dooling and Searcy, 1980, 1985). This range (2-4 kHz) also represents the peak in the budgerigar's audibility curve (Dooling, 1982; Dooling and Saunders, 1975; Dooling *et al.*, 1978, 1979). These findings suggest that budgerigars may perceive some vowel sounds as a "carrier" (i.e. the peak component of F2) surrounded by "modulation-induced sidebands" (i.e. harmonic components). The hypothesized use by birds of equally-spaced spectral components in the perception of pitch is comparable to the use of harmonic spacing by human listeners in the perception of fundamental frequencies in speech (Remez and Rubin, 1984).

My acoustic analysis of budgerigar "speech" has revealed that F2 is, indeed, the acoustic axis of these vocalizations (Scanlan, 1988). It has also revealed consistent variation of modulation patterns - both AM and FM - within a budgerigar's "speech" repertoire which indicate that such patterns, which may be communicative features of species-specific vocalizations, have been imposed with communicative intent on the appropriated speech-imitation sounds. This is consistent with the finding of Lohr and Dooling (1998) that budgerigars are much more sensitive than humans to anomalies in harmonic patterns.

Sequences of whistles with patterns of rapidly changing frequency within the range of human F2 are common components of budgerigar warble song (related to the "contact", "chirly" and "widdly" calls of Farabaugh *et al.*, 1982). These whistled sequences resemble, in both frequency range and frequency modulation (Dooling *et al.*, 1995), individual components of the synthetic "sinewave speech" used in speech-perception experiments (Remez *et al.*, 1981, 1994; Best *et al.*, 1989). My study showed that, by amplitude modulating whistles such as these, a bird can create the auditory impression of low-frequency

“voicing” (Scanlan, 1988). At the pre-performance warble stage, however, patterns of amplitude modulation can still be more bird-like than voice-like, and the main adaptation towards “speech” is a matching of whistle frequency contours with the changing contours of spectral peaks - principally those representing F2 - in the speech signal.

Figure 3, taken from the pre-performance warble of Budgerigar A, shows that only minimal alterations are required to transform species-specific whistles into crude representations of speech sounds. The spectrogram represents the bird’s imitation of a phrase - “pretty pretty boy” [prɪtɪ prɪtɪ bɔɪ] - followed by a whistle sequence which bears a strong resemblance to the phrase. The plosive burst and upward “second-formant transition” of the first [prɪ] have their counterparts in the rapid release and initial upward glide of the first whistle (at 1.2 sec on the spectrogram). The plosive burst, aspiration noise, and falling second-formant transition of [tɪ] are precisely mirrored - in both frequency range and transition slope - in the rapid release and downward glide in the first whistle. The second [prɪtɪ] is a simplified version of the first. In the third word, the rising second-formant transition of [ɔɪ], at 0.8 sec, has its counterpart in an excursion of similar slope over an identical frequency range (1,200 Hz - 3,300 Hz) at the end of the final whistle. These segments bear a strong resemblance to the second-formant transition in the human-speech version of [bɔɪ] (Fig. 2). The overall timing of the whistle sequence matches that of the “speech” sequence at about 0.9 sec.

While it is remarkable that the vocalization represented by Figure 3 should be perceived by human listeners as “speech”, experiments have confirmed that synthetic speech signals consisting simply of three sinusoidal components - representing F1, F2 and F3 - are intelligible to naive listeners (Remez *et al.*, 1981), and trained listeners to “whistled speech” (Busnel and Classe, 1976) can recover the full phonetic content of a single-sinusoid representation (Busnel *et al.*, 1970). In the example presented here, the amplitude modulation of the sinusoidal F2 representation is sufficient to create an impression of “voicing”.

Figure 4 was derived from the recording of Budgerigar B, which produced the same phrase as Budgerigar A [prɪtɪ prɪtɪ bɔɪ] in pre-performance warble and, like Budgerigar A, followed the “speech” sequence by a series of whistles resembling it in both spectral and temporal dimensions. The overall sequence has a broader and lower-frequency spectrum than that of Figure 3, and in those respects is more speech-like than the earlier example. In Figure 4 the tripartite temporal structure of the whistled sequence, and its frequency contour pattern,

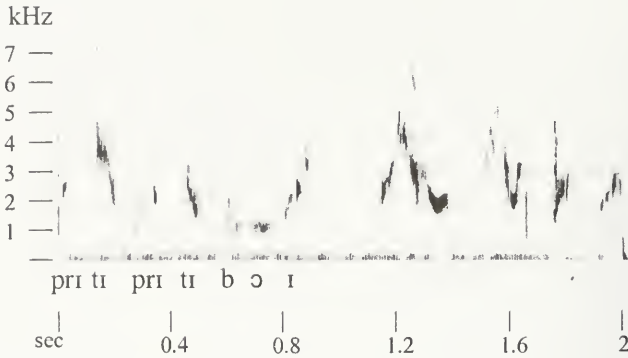


Figure 3. A sound sequence produced by Budgerigar A during “pre-performance warble”: the phrase “pretty pretty boy” [pri ti pri ti b o i] followed immediately by a series of species-specific whistles (at 1.2 - 2.0 sec) mirroring several acoustic features (see text) of the speech-imitation phrase.

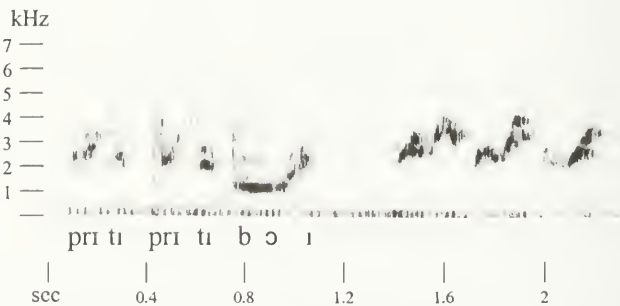


Figure 4. The same speech-imitation phrase as in Figure 3, this time produced by Budgerigar B during “pre-performance warble”. In this instance, too, an immediately-following series of species-specific whistles reproduces several spectral and temporal features (see text) of the speech-imitation phrase.

follow the “speech” sequence even more closely than their counterparts in Figure 3. In Figure 4 the rising contours associated with the high front vowels in [pɪɪ] and [bɔɪ] have particularly obvious correlates in the whistled sequence.

Experiments by Dooling *et al.* (1995) have shown that budgerigars can discriminate among three-formant synthetic /ra/ and /la/ syllables using the same cue - F3 transition - as human listeners. Furthermore, they can discriminate in a similar fashion among single-sinewave stimuli which follow the frequency modulation patterns of F3 in the /ra/-/la/ continuum. (The F3 frequency range in these stimuli was about 1,700-2,800 Hz.) These results suggest that budgerigars may hear single-sinusoid signals as speech-like if those signals incorporate cues which are sufficient for discrimination among speech sounds. The corollary, of course, is that they may reproduce speech as a single-sinusoid signal incorporating those critical cues.

The examples of Figure 3 and Figure 4 suggest that, for both birds, there is a strong association between the speech-imitation and species-specific sounds. Further analysis - some of which is presented below - demonstrates that this association is at least partly based on acoustic similarities between the high front vowels /i/ and /ɪ/ (and their transitions to adjacent speech sounds) and certain narrow-bandwidth elements of budgerigar warble song.

The contact call and high front vowels

Perhaps the most convincing evidence for the transformation of a natural sound pattern into an imitation of speech is to be found in those sequences in which speech-like sounds have replaced only part of the original species-specific pattern or, conversely, in which a species-specific element has been reintroduced into an imitated word or phrase. In the pre-performance warble of budgerigars, this is most common in the case of the high front vowels /i/ and /ɪ/, which share some important acoustic characteristics with the most common call of the budgerigar, the contact call (Park and Dooling, 1985; Farabaugh *et al.*, 1992, 1994, 1998; Farabaugh and Dooling, 1996; Brittam-Powell *et al.*, 1997). These vowels, like the contact call, have a strong concentration of energy (high F2 + F3) at about 2,000-3,500 Hz. The contact call is also characterised by rapid glides which, in speech imitation, can be adapted to serve as transitions to a high second formant. Indeed the second-formant transitions of normal human speech, if isolated, sound so like bird vocalizations such as budgerigar contact calls (labelled “chrps” by Wyndham, 1980) that they are called “chirps” in the literature of

phonetics (Mattingly *et al.*, 1971).

The finding of Farabaugh *et al.* (1998) that the spectral structure of contact calls in wild and domesticated budgerigars is essentially the same, emphasises the typical, species-specific nature of the pet budgerigar's "chrp". Domesticated budgerigars are able, however, to discriminate among contact calls which have been experimentally highly distorted (Park and Dooling, 1986).

Figure 5 shows a sequence of pre-performance warble by Budgerigar A in which a contact-call-like vocalization (subsequently referred to as a "chrp") has taken the place of the [ɪ] sound in the speech-imitation phrase "clever boy" [klevə bɔɪ]. Comparison of this sequence of pre-performance warble, having a mixture of speech-like and avian elements, with a completely "spoken" version (Fig. 6), produced by the same budgerigar (Budgerigar A) during a full performance, demonstrates the similarity between the rapid upward glide at the beginning of the "chrp" and the second-formant transition from [ɔ] to [ɪ] in the word [bɔɪ]. The energy in the [ɪ] sound in the "spoken" phrase is centred on about 2,800 Hz, as is that in the "chrp".

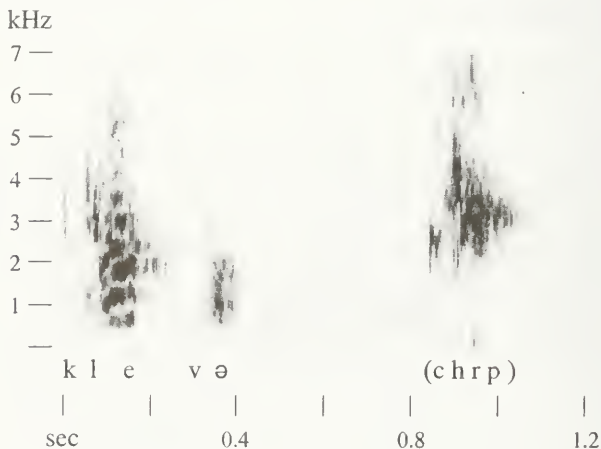


Figure 5. Budgerigar A, in a bout of "pre-performance warble", begins to produce the phrase "clever boy" [klevə bɔɪ], which forms part of his speech-imitation repertoire, but substitutes for the word [bɔɪ] a species-specific contact call. This call is acoustically similar to the vowel [ɪ] which ends the word [bɔɪ]. In timing, the contact call occurs precisely when the [ɪ] sound of the complete phrase (see Figure 6) is due.

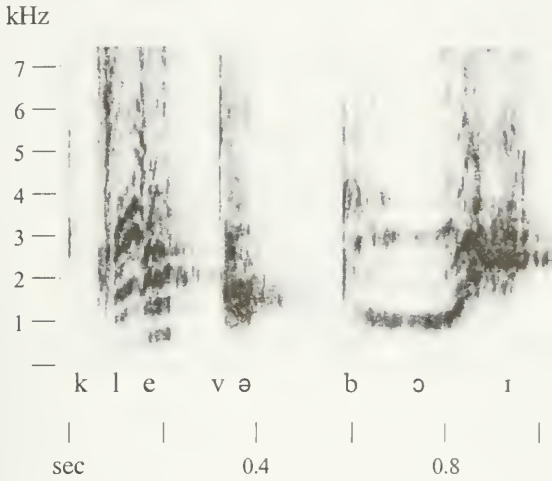


Figure 6. The phrase “clever boy” [klevə bɔɪ] from a full performance of Budgerigar A’s repertoire, for comparison with the “pre-performance warble” version of the phrase in Figure 5, and of the word [bɔɪ] in Figures 3, 4 and 7.

Compare Figure 6 with the human-voice version of [bɔɪ] in Figure 2, and note the similarity of the F2 transition patterns.

Further evidence that the “chrp” in Figure 5 represents the final section of the word [bɔɪ] is the presence on the original spectrogram (although not clearly visible on its reproduction here) of a click, like the plosive burst of [b], about 0.13 sec before the onset of the “chrp”. Relative timing is also an important clue: the onset of the “chrp” in Figure 5 occurs 0.83 sec after the initial plosive burst in [klev-], and exactly the same interval separates the beginning of the [ɪ] sound in Figure 6 from the beginning of the phrase. The durations of the [ɪ] sound and the “chrp” are similar.

Figure 7, another sequence from the pre-performance warble of Budgerigar A, shows the word [bɔɪ] followed immediately by a “chrp” which retains many of the acoustic features of the [ɔɪ] sound. The main area of intensity is within the same frequency range (2,000-3,500 Hz, the F2-F3 region of high front vowels) and the contour of the second-formant transition in [ɔɪ] resembles, albeit in a lower frequency range, the steep upward glide of the “chrp”. As in the whistle sequences, this juxtaposition of acoustically similar species-specific and speech-like sounds suggests that this similarity has significance for the bird.

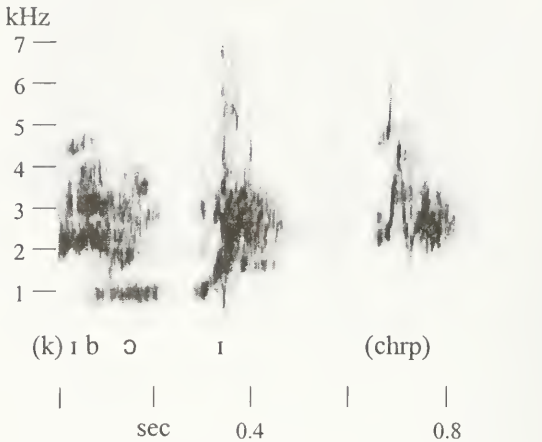


Figure 7. A sequence of “pre-performance warble” from Budgerigar A in which part of the phrase “Lucky boy” [(lʌk)ɪ bɔɪ] is followed immediately by a species-specific contact call which is acoustically similar to the [ɔɪ] section of [bɔɪ].

DISCUSSION

The analysis above has identified components of budgerigar and human communication sounds which are similar in acoustic structure. It indicates that, in copying speech, budgerigars focus on components of the speech signal which resemble components of budgerigar warble. These “shared” components (e.g. the frequency modulations typical of budgerigar contact calls and human second-formant transitions) are of particular communicative significance in both human and budgerigar vocalizations. Suga (1988) has suggested that if animals and humans produce similar acoustic patterns, they could share similar neural mechanisms for perceiving them. The corollary to Suga’s hypothesis is that, under the right circumstances, a neural mechanism designed for “special” perceptual processing of a species-specific vocal pattern could afford such “special” processing to an acoustically similar non-species-specific vocalization. The data presented here support this view: the incorporation of speech sounds into the budgerigar’s communicative

repertoire indicates that these non-species-specific sounds are being processed by such “special” perceptual mechanisms. (The perception of a sound as being an appropriate model for imitation is an essential ingredient of the overall perceptual-motor pattern of vocal learning. Extrapolation from observed vocal production to inferred auditory perception is thus a valid procedure.) This perceptual phenomenon could help to explain the human-like results of speech-perception experiments using avian (and other non-human) subjects.

Levels of analysis

One of the basic concerns of speech perception research has been to investigate the progressive analysis of the speech signal as it ascends from the periphery. The tendency has been to attempt to isolate discrete “levels” of analysis, the most commonly-drawn distinction being that between an “acoustic” or “auditory” level and a “phonetic” level (e.g. Pisoni, 1973). The experimental results of Cutting (1976) and Samuel and Kat (1996) even prompted those authors to propose the existence of several “auditory” levels at progressively higher stages of abstraction and integration. It has been axiomatic in the literature that, when a non-human animal successfully discriminates among speech sounds according to human-like categories, the categorisation occurs at an “auditory” level of perception because, by definition, it cannot happen at a “phonetic” (i.e., speech-specific) level (Kuhl, 1981, 1988; for a more recent review see Fitch *et al.*, 1997). Such animal results have been regarded, by analogy, as evidence that the speech-sound discriminations in question must rely on “auditory” processes in human listeners as well (Kuhl, 1987).

The term “levels of analysis” as used in the phonetics literature can be confusing: it seems to imply an anatomical correlate to what is usually no more than a simple conceptual scheme imposed on a highly complex neurological process. While there are undoubtedly hierarchical patterns within this process, references to “levels of analysis” in the present paper should *not* be understood as referring to anatomically identifiable centres along the neural pathway for the perception and recognition of communication sounds.

This paper argues, on the basis of the data presented above, that in humans and at least some other vertebrates speech sounds pass through a level (or levels) of perceptual analysis which is neither simply auditory nor specifically phonetic, but which is specially designed to process the complex acoustic cues which are common and peculiar to the communication sounds of many vertebrate species. In other words, vertebrates (including humans) share neural mechanisms capable of

processing an *interspecific* range of communicatively significant acoustic signals. These mechanisms are at levels higher than the purely auditory, but lower than the species-specific-vocal (e.g. the “phonetic” level in humans). The term “bio-communication” (Scanlan, 1988) is used below in referring to such levels of analysis and to the acoustic cues which receive special attention there. The involvement of “bio-communication” mechanisms in the perception of certain speech sounds could explain similar patterns of categorisation by both human and non-human listeners. The arguments against such categorisation occurring at a more peripheral “auditory” level are detailed below.

This “neuroethological” approach to speech perception (Suga, 1988) has the potential to reveal some aspects of the phylogeny of speech. For example, within the framework of a divergent evolutionary model, the data presented here could indicate a common antecedent for the budgerigar form (FM pattern of contact calls) and the human form (FM pattern of formant transitions) of a particular “bio-communication cue”. Focusing on the present study of human/budgerigar vocal interaction, a shared level of communication-sound processing would explain not only the budgerigar’s ability to substitute components of budgerigar warble into speech imitation sequences (and components of speech imitation into warble sequences), but also the human listener’s ability to interpret these performances as “speech”.

Formant transitions

A major problem for speech researchers is the apparently invariant perception of phonemes represented by highly variable acoustic patterns. A classic case of such acoustic variability is that of stop consonants in consonant-vowel (CV) syllables. Perceptual experiments such as those by Liberman *et al.* (1954) and Liberman (1970) showed that formant transitions - particularly F2 transitions - could convey sufficient information to distinguish place of production (for example, to distinguish a bilabial stop like /b/ from an alveolar stop like /d/ in the contrasting pair /ba/- /da/). In two-formant syntheses of pairs like /di/- /du/, however, in which the same consonant is followed by different vowels, the F2 transitions cueing the identical-sounding /d/ segments differ greatly between the two syllables, their contours being determined by the F2 frequencies of the following vowels (Liberman *et al.*, 1967). Is there, then, an invariant property of the signal corresponding to the invariant perception of /d/?

The traditional approach to this problem in relation to stop-consonant place of production (e.g. by Blumstein and Stevens, 1979) has been to focus - with inconclusive results - on spectral features,

which are presumed to engage purely auditory mechanisms of perception. The discovery that non-human subjects (Japanese quail, *Coturnix coturnix* - Kluender *et al.*, 1987; rhesus monkeys, *Macaca mulatta* - Morse and Snowdon, 1975; other macaques, *Macaca fasciata* and *M. nemistrina* - Kuhl and Padden, 1983; chinchillas, *Chinchilla laniger* - Miller and Kuhl, 1976) could also discriminate among stop-consonant stimuli according to place-of-production categories reinforced the view that such discrimination relied on purely auditory processes. However, continuing research (e.g. by Tomiak *et al.*, 1987; Kewley-Port *et al.*, 1988) has indicated that, at least in humans, there is no "special" processing of speech (such as the imposition of perceptual categories) at auditory levels. Indeed, there is strong neurophysiological evidence that, at these levels, processing of speech and non-speech sounds is identical. Sharma *et al.* (1993) recorded an event-related potential which is sensitive to fine acoustic differences while subjects listened to stimuli comprising a /da/-/ga/ continuum. They found that the response was the same for stimulus pairs within and between categories, indicating that there is no categorical effect at auditory levels. The experiments of Kluender and Diehl (1987) and Kluender *et al.* (1987), which could identify no invariant acoustic correlates of the speech categories perceived by Japanese quail, support the view that in some non-human species, too, processing of such categories does not occur at purely auditory levels.

This, then, is the situation: there is a large body of experimental evidence which indicates that some communicatively significant components of speech are processed similarly in humans and some non-humans. Also, there is a range of evidence which refutes the view that such processing occurs at purely "auditory" levels. The conclusion from these findings is that this processing must occur at more central levels - levels which afford some "special" treatment to communication sounds. These levels must be intermediate between "auditory" and, e.g., (in humans) "phonetic" or (in birds) "song-specific" levels, because they are capable of processing *interspecific* communication sounds. This is possible because some communicatively significant acoustic cues are common to the vocalizations of a variety of vertebrates. These acoustic cues, which allow components of one species' communication sounds to be processed as "communicatively significant" by an individual of another species, can be isolated by the study of phenomena such as avian "speech". The "bio-communication cue" examined in this paper is a particular pattern of frequency modulation which is a communicatively essential component of both the /i/ vowel in human speech and the budgerigar contact call. A budgerigar's processing of this cue at the "bio-communication" level allows it to identify the /i/

sound with the species-specific call in communicative behaviour with a human companion.

At the same time as the speech scientists were investigating the putative "auditory discontinuities" of speech perception by perceptual experiments with animals, comparative neurologists were beginning to develop models of speech perception based on findings relating to central perceptual mechanisms in non-human species. A concept basic to these models is that of the "neural map". One such neuroethological model has been constructed - significantly for the present argument - by analogy with an avian perceptual mechanism. Sussman (1989) has proposed an analogy between the categorical perception of speech sounds by humans and sound localization in the barn owl (*Tyto alba*). The essence of his argument is that, by the co-activation of an *array* of neurons, individual speech sounds can be located within a multi-dimensional neural map of speech-sound categories in much the same way as perceived sounds can be located within the barn owl's neural map of external space (Knudsen and Konishi, 1978).

This model of speech perception is in line with the hypotheses of Suga (1988), who discovered arrays of neurons in the auditory cortex of the moustached bat (*Pteronotus parnellii*) which responded to combinations of biologically significant parameters in the bat's echolocation vocalisations (Suga *et al.*, 1978). Suga hypothesised that speech perception could be based on a spatio-temporal pattern of neural activity which, by such "neural mapping" of communicatively significant acoustic signals, operates in a fashion analogous to the perception of echolocation calls by bats. He (Suga, 1988) endorsed a neuroethological view of acoustic communication in which the uniqueness of a species' system for the perception of biologically significant signals is, nevertheless, based on the specialization of shared mechanisms. He pointed out that many bio-communication sound systems, including human speech, are composed of three types of information-bearing elements: constant frequency (CF), noise burst (NB) and frequency-modulated (FM) components. For example, the consonant-vowel syllables of English consist of a plosive burst (NB), formant transitions (FM) and vowel formants (CF). The CF and FM components of these sounds have a similar acoustic structure to the sonar signals of moustached bats, and could therefore be processed by analogous mechanisms.

Suga (1988) also points out that the classical methods of auditory physiology, in which neural responses to pure tones and clicks are elicited, has contributed little to our understanding of central auditory mechanisms. He advocates a neuroethological approach - using biologically significant sounds - for the study of the central auditory

system. In the spirit of this plea, Cynx *et al.* (1990) related CF and FM cues in the harmonic spectra of zebra finch song syllables to the formant structure of human speech, and suggested that this acoustic correspondence might be related to central rather than peripheral auditory mechanisms.

Thus the human-like results of speech-perception tests with non-human animals could result from the higher-order processing of these sounds in auditory centres designed for the processing of species-specific signals similar to speech in their information-bearing (or “bio-communication”) parameters. The demonstrated capacity of a budgerigar to substitute communicatively significant elements of speech for acoustically similar elements of species-specific song in its vocal repertoire indicates the possibility of such common auditory mechanisms for the processing of communication sounds. (In the case of the budgerigar, however, even the highest-level centres have been able to adapt, as a result of vocal learning, to process speech as if it were budgerigar sounds. See Margoliash, 1983, for evidence that single units in songbird HVC - the “higher vocal centre” of the telencephalon - respond to specific song syllables which the bird has learnt as part of its repertoire even when, as a result of contrived learning conditions, those syllables are untypical of its natural song.)

The hypothesis presented here is that there is a level of analysis which handles bio-communication sounds in an integrative, categorical fashion - extracting information from parameters which vertebrate communication sounds have in common before the sounds proceed to the highest, species-specific levels of analysis. In species which learn their vocalizations, like some birds, these highest levels, too, can be modified by learning experience to process alien sounds as if they were species-specific. This is possible, however, only because of the lower-level function of interspecific bio-communication cues.

The voicing distinction

The most intensively-studied example of “special” processing in speech perception - by both human and non-human subjects - is the categorical discrimination of voiced and voiceless initial stop consonants in CV syllables. A synthetic continuum of sounds between English /ba/ and /pa/, for example, is perceived by English speakers as consisting of a group of virtually indistinguishable /ba/s and a group of virtually indistinguishable /pa/s, with a marked contrast between the two groups. Stimuli adjacent to each other in the continuum, but which fall on opposite sides of the perceptual boundary, are clearly distinguishable. The physical variables in this case are connected with a

contrast in the onset of laryngeal vibration, which occurs near the beginning of the sound in /ba/ and after a delay of more than 20 msec in /pa/. (See Liberman, 1977, for a discussion and review.)

The discovery of this phenomenon was initially regarded as evidence of speech-specific mechanisms in human auditory perception, and was even embraced by proponents of the "motor theory" of speech perception (Liberman *et al.*, 1967; Mattingly *et al.*, 1971; Liberman and Mattingly, 1985; Mattingly and Liberman, 1988), which seeks to link the perception of speech sounds with knowledge of the articulatory gestures involved in their production. Experimental evidence that prelingual infants imposed the same perceptual categories on the stimuli (reviewed by Kuhl, 1987, 1988, 1989) countered the arguments of the motor theorists, while studies with infants exposed only to languages which do not employ the voiced/voiceless contrast provided evidence for an innate, universal sensitivity to this distinction (Lasky *et al.*, 1975; Streeter, 1976; Eilers *et al.*, 1979).

Comparable results from several species of nonhuman vertebrate (rhesus monkeys - Waters and Wilson, 1976; Japanese macaques, *Macaca fuscata* - Kuhl and Padden, 1982; chinchillas - Kuhl, 1976, 1979, 1981, Kuhl and Miller, 1975a, 1975b, 1978; budgerigars - Dooling *et al.*, 1987, 1989, 1995, Dooling and Brown, 1990) have indicated that the phenomenon is not even *homo*-specific. When both identification and discrimination of the stimuli have been tested, the animal results have dovetailed in the fashion considered necessary and sufficient evidence for the categorical perception of speech (Studdert-Kennedy *et al.*, 1970) - the peak in the discrimination function corresponding with the discontinuity in the labelling function (Kuhl, 1981; Kuhl and Padden, 1983). (These results are reviewed, and their implications discussed, by Miller, 1977; Kuhl, 1979, 1986a, 1987, 1988, 1989; Snowdon, 1979, 1990; Moody *et al.*, 1990.)

Finally, correspondences between VOT as *produced* by speech-imitating birds and by their human trainers have contributed to speculation about common underlying perceptual mechanisms (Scanlan, 1988; Patterson and Pepperberg, 1998).

The difference in timing between the onset of laryngeal vibration and the plosive noise burst in syllable-initial prestressed stop consonants is known as "voice onset time" or VOT (Lisker and Abramson, 1964). On spectrograms of English stop consonants the beginning of laryngeal vibration is most clearly reflected in the onset of first-formant (F1) excitation, as there is little pre-voicing aspiration noise at first-formant frequencies. A voiced/voiceless continuum can, therefore, be synthesised simply by "cutting back" the first formant as the stimuli move from voiced to voiceless (Liberman *et al.*, 1958). The

stimuli in all the animal experiments have been of the “F1 cutback” variety - in which, as in speech itself, the onset frequency of F1 is a spectral correlate of the temporal cue. (Soli, 1983, has argued that this spectral feature is critical for human subjects. See Rosen and Howell, 1987, for a review of the debate about spectral and temporal cues for VOT.) Thus these experiments, while demonstrating categorical perception of voiced/voiceless consonant continua in non-human vertebrates, do not isolate timing differences as the sole criteria for the distinction.

Sinex *et al.* (1991) emphasised the importance of spectral cues in the categorical response of auditory-nerve fibres to an “F1 cutback” /da/-/ta/ continuum which they recorded in chinchillas. An analogous peripheral response to the same stimuli in humans has been suggested by the recording of cortical auditory evoked potentials (Sharma and Dorman, 1999). The integration of cues resulting from auditory sensitivities at the periphery could occur at the hypothesised “bio-communication” level of processing.

The variable, inverse relationship between spectral and temporal cues has become known as “trading relations” (e.g. in Best *et al.*, 1981; Repp, 1982; Sinnott and Brown, 1997). Best *et al.* (1981) interpreted this “perceptual equivalence” of cues from different acoustic dimensions as indicating that these acoustic cues are processed phonetically, and found that such processing did not occur when listeners perceived the acoustic variations as nonspeech contrasts. They went on to argue that this “equivalence” of the acoustic cues at the phonetic level indicates that, at that level, the common articulatory origin of the two cues is taken account of. This view (in accord with that of Repp, 1982) falls within the ambit of the “motor theory” of speech perception: see, e.g., Liberman and Mattingly (1985). Unfortunately for the motor theory, however, it has been shown that Japanese quail have a human-like compensatory response to covariations in F1 onset frequency and F1 cutback duration in the perception of CV syllables (Kluender, 1991; Kluender and Lotto, 1994). Similarly, Sinnott and Brown (1997) have found comparable trading relations between spectral and temporal cues in the perception of /ra/-/la/ continua by Japanese macaques.

Recently, Lotto *et al.* (1997) have tackled the problem by hypothesizing that, as uniform physical laws operate on all the vocal mechanisms in the animal kingdom, individuals are specially sensitive to the manifestation of certain physical constraints on vocalization - even in the sounds of an alien species. (They were discussing their finding that in Japanese quail the perceptual boundary within a /da/-/ga/ continuum changes, in response to changes in a preceding syllable, in

the same way as it does for human subjects. In humans, this effect is known as “perceptual compensation for coarticulation”.) The approach of Lotto *et al.* is, in effect, an attempt to accommodate the animal results within a kind of species-general motor theory. Considering the fundamental differences between the mechanisms of vocalization in birds and humans, however, it is not possible to invoke even the most generalised motor theory in explaining inter-specific neural patterns for the perception of human and avian vocalizations. In this case, at least, an emphasis on auditory rather than motor factors in explaining the phenomena seems inevitable. The adult-like perceptual performance of prelingual infants (e.g. Kuhl, 1989) is strong evidence that here, too, auditory factors must be predominant.

An attempt could be made to incorporate arguments such as those of Lotto *et al.* into a “direct realist” theory of speech perception, which argues that perception of all sounds - speech included - can be considered as *direct experience* of the sound source rather than (as in the motor theory) a *cognitive reconstruction* of the sound source (Fowler, 1980, 1986, 1990, 1992). The attempt would probably founder, however, in the conundrum of “direct” perceptual experience of a songbird’s syringeal mechanism.

Alternatively to the motor theory or direct realist interpretations, and more in accord with the body of evidence adduced in this paper, trading relations (and perceptual compensation for coarticulation) could be seen as reflecting the complex activity of a multi-dimensional “neural map” of communicatively significant acoustic cues similar to that hypothesised by Suga (1988) and discussed above. In this case, processing at intermediate “bio-acoustic” levels could be similar in humans and animals, and explain the human-like categorical response of animals to some speech-sound stimuli.

Vowels

Vowel sounds have traditionally been regarded as discriminable on the basis of static spectral features (Peterson and Barney, 1952; see Rosen and Fourcin, 1986, for a review and critique of this position, and Miller, 1989, for its elaboration into an “auditory-perceptual theory” of the vowel). A long-established principle of acoustic phonetics is that the *relationships* among formant frequencies are of prime significance in vowel discrimination (Potter and Steinberg, 1950). Scheich *et al.* (1977) remarked on the significance of “formant” structure (relative position of spectral envelope peaks) in vertebrate vocalizations generally, and, in a later paper (Scheich, Bonke and Langner, 1979), focused on the formant structure of human vowels and the ability of

individuals of other species (chinchillas - Burdick and Miller, 1975; dogs - Baru, 1975; cats - Dewson, 1964; monkeys - Dewson *et al.*, 1969) to discriminate among them. Such discrimination had been shown to be based, as in human vowel perception, on formant-frequency relationships (Baru, 1975; Burdick and Miller, 1975). This has been corroborated by more recent work with baboons, *Papio anubis* (Hienz and Brady, 1988) and macaques and vervet monkeys, *Cercopithecus aethiops* (Sinnott, 1989). On the other hand, the work of Kuhl (1991) has suggested that, for human listeners but not for rhesus monkeys, there may be within-category distinctions relating to "prototypical" vowel sounds. Recordings of event-related potentials reflect the multi-faceted nature of human speech perception (Aaltonen *et al.*, 1987; Naatanen *et al.*, 1997), as incorporated in recent attempts at "connectionist" (or neural network) modelling of such perception (reviewed in Protopapas, 1999).

Scheich and his colleagues recorded neuronal responses to both natural and synthetic vowel sounds in tonotopically organised layers of telencephalic Field L of speech-trained Indian hill mynahs. They reported (Langner *et al.*, 1979) that some units responded selectively to only one or two vowels, while many units could distinguish between two vowels which shared a common formant frequency. This approach revealed a variety of excitatory-inhibitory interaction patterns in the selective responses of individual neurons to two-formant stimuli. When plotted in relation to F1 and F2, these neuronal responses (Langner *et al.*, *ibid.*) corresponded with discrimination patterns derived from human psychophysical responses to the same vowels (Hose *et al.*, 1983). This indicates a similar neural perception mechanism for these human sounds in both man and mynah (Hose *et al.*, *ibid.*). The hypothesis of Langner *et al.* (1981) that this vowel-specific response in the birds' forebrains was because of their speech-learning experience, while in line with the argument of this paper, remains to be tested. Meanwhile, my own analyses of speech-imitation sounds by several species of "talking" birds (Scanlan, 1988), and analyses by Patterson and Pepperberg (1994) of the "speech" of an African grey parrot (*Psittacus erithacus*), indicate that formant frequency relationships are distinguishing features of vowel imitations *produced* by birds, and that therefore they are *perceived* as such.

The ascending perceptual pathway in the Guinea fowl (*Numida meleagris*) involves auditory midbrain nuclei whose neurons are responsive to complex acoustic cues used in species-specific communication sounds but are unable to discriminate among such sounds (Scheich *et al.*, 1977). The Guinea fowl's telencephalic auditory nucleus (Field L), on the other hand, contains units which can

discriminate among call types (Bonke *et al.*, 1979; Scheich, Langner and Bonke, 1979). These findings suggest a hierarchical pattern of auditory perception in which "bio-communication cues" are identified before the complete communicative analysis of the signal. (Significantly, the Guinea fowl vocalization which was the subject of these studies, the "iambus call", has an acoustic structure similar to that of human vowels - Scheich, Langner and Bonke, *ibid.*) Field L of the Indian hill mynah is similar in structure and function to that of the Guinea fowl (Langner *et al.*, 1981), which suggests that there could be a corresponding identification of "bio-communication cues", such as the presence of formant-like peaks in the spectrum, at midbrain levels in the mynah before the complete analysis of, e.g., vowel sounds in the forebrain. That the response of non-human animals to vowel formant patterns is not the result of simply acoustic neural processing is suggested by the finding that in some birds (Hienz *et al.*, 1981) and monkeys (Hienz and Brady, 1988; Sinnott and Kreiter, 1991) acuity of vowel discrimination is far greater than that expected on the basis of pure-tone discrimination ability.

The patterns of neuronal response to species-specific calls in Guinea fowl, and human vowels in Indian hill mynahs, correspond to those (discussed above) which Knudsen and Konishi (1978) found in sound-localizing barn owls, as well as those which Suga discovered in echolocating bats (Suga *et al.*, 1978) and those which he (Suga, 1988) hypothesised as being involved in human speech perception. The results of psychophysical tests of human vowel perception by Hose *et al.* (1983) support Suga's hypothesis.

Scheich *et al.* (1983) considered the presence of such neural auditory response patterns in a diversity of vertebrates, in which "each individual call will lead to a distinct spatial pattern of excitation of small ensembles of neurons", as being evidence of their phylogenetic stability. These homologous auditory mechanisms could have resulted in the evolution of a variety of species relying on similar acoustic cues in vocal communication (Scheich, 1985). Reciprocally, the communicative effectiveness of these acoustic cues, and therefore their phylogenetic conservation, could explain the remarkable stability of the auditory system throughout vertebrate evolution (Scheich *et al.*, 1983).

GENERAL DISCUSSION

The evolution of speech

Assumptions about the evolution of speech underlie most

theorising on the basic mechanisms of speech perception. Advocates of “auditory sensitivities” (e.g. Kuhl, 1988) regard audition as primary, and the evolution of acoustic communication systems - including human speech - among the vertebrates as having been influenced by auditory factors which have been relatively stable, phylogenetically (Scheich *et al.*, 1983; Dooling *et al.*, 1995), and thus as having generally preceded and guided the evolution of vocal mechanisms. The “motor theory”, on the other hand, with its emphasis on the primacy of production, regards the evolution of special auditory mechanisms for speech perception as having occurred subsequently to the evolution of the human vocal mechanism (Mattingly and Liberman, 1988), and the ontogeny of speech perception as being dependent on vocal experience (Liberman *et al.*, 1967).

Apart from the apparently insuperable problems for the motor theory posed by the animal and prelingual-infant experiments, the “auditory sensitivities” approach has the advantage of parsimony. Indeed, the idea that auditory systems involved in acoustic communication have evolved independently in various vertebrate classes and orders - and even genera and species (e.g. *Homo sapiens*) - is difficult to accommodate to any version of evolutionary theory.

The question then arises: do these “sensitivities” in humans relate to the general auditory system, or are they specially evolved for processing speech? Experiments with nonspeech sounds (e.g. those by Hirsh, 1959, Stevens and Klatt, 1974, Miller *et al.*, 1976, Pisoni, 1977, Divenyi and Sachs, 1978, and Jusczyk *et al.*, 1980, which all indicate a critical interval - analogous to VOT - in the perception of temporal order) have shown that categorical perception can apply to nonspeech stimuli. As Kuhl (1986b, 1987) has pointed out, however, these results could be interpreted in relation to the “tuning” of special speech-processing mechanisms - i.e., they may not be so narrowly tuned to speech as to exclude nonspeech sounds having the relevant features. Indeed Mattingly (1972), in comparing speech cues with sign stimuli such as the red belly of the male stickleback (see Tinbergen, 1951), emphasised the increased effectiveness of “supernormal” (i.e. artificially enhanced) cues in stimulating both speech-sound discrimination by humans and reaction to sign stimuli by animals.

Mattingly’s insight, however, did not go beyond the level of comparing human speech cues with non-human acoustic signals such as avian alarm calls (which elicit stereotyped responses). The present paper argues, rather, that speech cues should be compared with features of avian vocalizations, such as certain frequency modulations in budgerigar warble song, which function contrastively within highly variable acoustic patterns to communicate specific information - such

as personal and group identity. (It is possible, nevertheless, that the communication cues of both human speech and budgerigar warble evolved from acoustic "sign stimuli" such as alarm calls.)

While animal speech-perception experiments demonstrate that mechanisms such as those underlying the perception of F2 transitions are not specific to humans, the research results presented in this paper indicate that, in budgerigars and humans at least, such shared mechanisms correspond to acoustic cues which have a shared *communicative* function. This conclusion is in line with those of Suga (1988), Cynx *et al.* (1990) and Dooling *et al.* (1995). The perspective advocated here is thus aligned to the "auditory sensitivities" view, with its Darwinian approach to the evolution of speech, but sees the phylogenetically stable element of the perceptual mechanism not as purely auditory, but as specially evolved to filter communicative components from incoming acoustic signals. The phylogenetic stability of this filter, and the conservation of acoustic "bio-communication cues" throughout phylogeny are, of course, complementary.

In focusing on their common, inherited features, this perspective adds new depth to a holistic view of vertebrate (including human) communication sounds.

ACKNOWLEDGEMENTS

The acoustic analysis was done in the Department of Phonetics and Linguistics, University College London, with the encouragement and under the supervision of Professor A. J. Fourcin, and with the generous help of his colleagues. The theoretical arguments were developed and the paper written during a period as Honorary Research Associate in Professor Lesley Rogers's neuroscience and animal behaviour group within the Department of Zoology at the University of New England, New South Wales. I am grateful to Professor Rogers and Professor Nihal Agar for providing work space and facilities, to Professor Rogers and her research group for stimulating discussions, and to Professor Rogers, Jan Patterson, and Dr Helen Fraser for helpful comments on the manuscript. The figures were produced with the generous assistance of Craig Lawlor and Ian Truswell, both of the University of New England.

REFERENCES

- Aaltonen, O., Niemi, P. & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, 24, 197-207.
- Adret, P. (1993). Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Animal Behaviour*, 46, 149-159.
- Adret-Hausberger, M., Guttinger, H.-R & Merkel, F.W. (1990). Individual life history and song repertoire changes in a colony of starlings (*Sturnus vulgaris*). *Ethology*, 84, 265-280.
- Adret-Hausberger, M. & Jenkins, P.F. (1988). Complex organization of the warbling song in the European starling *Sturnus vulgaris*. *Behaviour*, 107, 138-156.
- Banta Lavenex, P. (1999). Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): The presence and implications of amplitude modulation. *Journal of the Acoustical Society of America*, 106, 491-505.
- Baptista, L.F. & Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Animal Behaviour*, 32, 172-181.
- Baptista, L.F. & Petrinovich, L. (1986). Song development in the white-crowned sparrow: social factors and sex differences. *Animal Behaviour*, 34, 1359-1371.
- Baptista, L.F. & Schuchmann, K.-L. (1990). Song learning in the Anna hummingbird (*Calypte anna*). *Ethology*, 84, 15-26.
- Baru, A.V. (1975). Discrimination of synthesized vowels /a/ and /i/ with varying parameters in dog. In G. Fant & M.A.A. Tatham (Eds.), *Auditory Analysis and Perception of Speech* (pp. 91-101). New York and London, Academic Press.
- Bertram, B. (1970). The vocal behaviour of the Indian hill mynah, *Gracula religiosa*. *Animal Behaviour Monographs*, 3, 79-192.
- Best, C.T., Morriongiello, B. & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 29, 191-211.
- Best, C.T., Studdert-Kennedy, M., Manuel, S. & Rubin-Spitz, J. (1989). Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics*, 45, 237-250.
- Blumstein, S.E. & Stevens, K.N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66, 1001-1017.
- Bonke, D., Scheich, H. & Langner, G. (1979). Responsiveness of units in the auditory neostriatum of the Guinea fowl (*Numida meleagris*) to species-specific calls and synthetic stimuli I: tonotopy and functional zones of field L. *Journal of Comparative Physiology*, 132, 243-255.
- Bottjer, S.W., Glaessner, S.L. & Arnold, A.P. (1985). "Ontogeny of brain nuclei controlling song learning and behavior in zebra finches. *Journal of Neuroscience*, 5, 1556-1562.
- Bottjer, S.W., Miesner, E.A. & Arnold, A.P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, 224, 901-903.
- Brauth, S.E., Heaton, J.T., Durand, S.E., Liang, W. & Hall, W.S. (1994). Functional anatomy of forebrain auditory pathways in the budgerigar (*Melopsittacus undulatus*). *Brain, Behavior and Evolution*, 44, 210-233.
- Brenowitz, E.A. & Kroodsma, D.E. (1996). The neuroethology of birdsong. In D.E. Kroodsma & E.H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 285-304). Ithaca and London, Cornell University Press.

- Brittan-Powell, E.F., Dooling, R.J. & Farabaugh, S.M. (1997). Vocal development in budgerigars (*Melopsittacus undulatus*): contact calls. *Journal of Comparative Psychology*, *111*, 226-241.
- Brockway, B.F. (1964). Ethological studies of the budgerigar: reproductive behaviour. *Behaviour*, *23*, 294-324.
- Brockway, B.F. (1969). Roles of budgerigar vocalization in the integration of breeding behaviour. In R.A. Hinde (Ed.), *Bird Vocalizations* (pp. 131-158). Cambridge, UK, Cambridge University Press.
- Brown, E.D., Farabaugh, S.M. & Veltman, C.J. (1988). Song sharing in a group-living songbird, the Australian magpie, *Gymnorhina tibicen*. 1: vocal sharing within and among groups. *Behaviour*, *104*, 1-28.
- Burdick, C.K. & Miller, J.D. (1975). Speech perception by the chilchilla. *Journal of the Acoustical Society of America*, *58*, 415-427.
- Busnel, R.-G. & Classe, A. (1976). *Whistled Languages*. Berlin, Heidelberg and New York, Springer-Verlag.
- Busnel, R.-G., Leroy, C. & Moles, A. (1970). Etude sur la langue sifflee de Kuskoy en Turquie. *Revue de Phonetique Appliquee*, *14-15*, 1-184.
- Catchpole, C.K. (1979). *Vocal Communication in Birds*. London, Edward Arnold.
- Chaiken, M., Bohner, J. & Marler, P. (1994). Repertoire turnover and the timing of song acquisition in European starlings. *Behaviour*, *128*, 25-39.
- Chew, S.J., Vicario, D.S. & Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 1950-1955.
- Clayton, N.S. (1987). Song tutor choice in zebra finches. *Animal Behaviour*, *35*, 714-721.
- Cutting, J.E. (1976). Auditory and linguistic processes in speech perception: inferences from six fusions in dichotic listening. *Psychological Review*, *83*, 114-140.
- Cynx, J., Williams, H. & Nottebohm, F. (1990). Timbre discrimination in zebra finch (*Taeniopygia guttata*) song syllables. *Journal of Comparative Psychology*, *104*, 303-308.
- Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London, John Murray.
- Dewson, J.H. (1964). Speech sound discrimination by cats. *Science*, *144*, 555-556.
- Dewson, J.H., Pribram, K.H. and Lynch, J.C. (1969). Effects of ablations of temporal cortex upon speech sound discrimination in the monkey. *Experimental Neurology*, *24*, 579-591.
- Divenyi, P.L. & Sachs, R.M. (1978). Discrimination of time intervals bounded by tone bursts. *Perception & Psychophysics*, *24*, 429-436.
- Dooling, R.J. (1973). Behavioural audiometry with the parakeet (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, *53*, 1757-1758.
- Dooling, R.J. (1982). Auditory perception in birds. In D.E. Kroodsma & E.H. Miller (Eds.), *Acoustic Communication in Birds*, Vol. 1 (pp.95-130). New York and London, Academic Press.
- Dooling, R.J., Best, C.T. & Brown, S.D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *Journal of the Acoustical Society of America*, *97*, 1839-1846.
- Dooling, R.J. & Brown, S.D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels. *Perception & Psychophysics*, *47*, 568-574.
- Dooling, R.J., Okanoya, K. & Brown, S.D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced-voiceless distinction. *Perception & Psychophysics*, *46*, 65-71.

- Dooling, R.J., Peters, S.S. & Searcy, M.H. (1979). Auditory sensitivity and vocalizations of the field sparrow (*Spizella pusilla*). *Bulletin of the Psychonomic Society*, 14, 106-108.
- Dooling, R.J. & Saunders, J.C. (1975). Hearing in the parakeet (*Melopsittacus undulatus*). *Journal of Comparative and Physiological Psychology*, 88, 1-20.
- Dooling, R.J. & Searcy, M.H. (1980). Forward and backward auditory masking in the parakeet (*Melopsittacus undulatus*). *Hearing Research*, 3, 279-284.
- Dooling, R.J. & Searcy, M.H. (1981). Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*). *Journal of Comparative Physiology*, 143, 383-388.
- Dooling, R.J. & Searcy, M.H. (1985). Nonsimultaneous auditory masking in the budgerigar (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 99, 226-230.
- Dooling, R.J., Soli, S.D., Kline, R.M., Park, T.J., Hue, C. & Bunnell, T. (1987). Perception of synthetic speech sounds by the budgerigar (*Melopsittacus undulatus*). *Bulletin of the Psychonomic Society*, 25, 139-142.
- Dooling, R.J., Zoloth, S.R. & Baylis, J.R. (1978). Auditory sensitivity, equal loudness, temporal resolving power and vocalizations in the house finch (*Carpodacus mexicanus*). *Journal of Comparative and Physiological Psychology*, 92, 867-876.
- Doupe, A.J. (1997). Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *Journal of Neuroscience*, 17, 1147-1167.
- Doupe, A.J. & Konishi, M. (1991). Song-selective auditory circuits in the vocal control system of the zebra finch. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 11339-11343.
- Doupe, A.J. & Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567-631.
- Durand, S.E., Heaton, J.T., Amateau, S.K. & Brauth, S.E. (1997). Vocal control pathways through the anterior forebrain of a parrot (*Melopsittacus undulatus*). *Journal of Comparative Neurology*, 377, 179-206.
- Eales, L.A. (1985). Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Animal Behaviour*, 33, 1293-1300.
- Eales, L.A. (1987). Song learning in female-raised zebra finches: another look at the sensitive phase. *Animal Behaviour*, 35, 1356-1365.
- Eales, L.A. (1989). The influence of visual and vocal interaction on song learning in zebra finches. *Animal Behaviour*, 37, 507-508.
- Eilers, R.E., Gavin, W. & Wilson, W.R. (1979). Linguistic experience and phonemic perception in infancy: a cross-linguistic study. *Child Development*, 50, 14-18.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague, Mouton.
- Farabaugh, S.M., Brown, E.D. & Dooling, R.J. (1992). Analysis of warble song of the budgerigar *Melopsittacus undulatus*. *Bioacoustics*, 4, 111-130.
- Farabaugh, S.M., Dent, M.L. & Dooling, R.J. (1998). Hearing and vocalizations of wild-caught Australian budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 112, 74-81.
- Farabaugh, S.M. & Dooling, R.J. (1996). Acoustic communication in parrots: laboratory and field studies of budgerigars, *Melopsittacus undulatus*. In D.E. Kroodsma & E.H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 97-117). Ithaca and London, Cornell University Press.
- Farabaugh, S.M., Linzenbold, A. & Dooling, R.J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108, 81-92.
- Ferrell, R. & Baptista, L.F. (1982). Diurnal rhythms in the vocalizations of budgerigars. *Condor*, 84, 123-124.

- Fitch, R.H., Miller, S. & Tallal, P. (1997). Neurobiology of speech perception. *Annual Review of Neuroscience*, 20, 331-353.
- Foss, B.M. (1964). Mimicry in mynahs. *British Journal of Psychology*, 55, 85-88.
- Fowler, C. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113- 133.
- Fowler, C. (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics*, 14, 3-28.
- Fowler, C. (1990). Sound-producing sources as objects of perception: rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, 88, 1236-1249.
- Fowler, C. (1992). Vowel duration and closure duration in voiced and unvoiced stops: there are no contrast effects here. *Journal of Phonetics*, 20, 143-165.
- Gaunt, S.L., Baptista, L.F., Sanchez, J.E. & Hernandez, D. (1994). Song learning as evidenced from song sharing in two hummingbird species (*Colibri coruscans* and *C. thalassinus*). *Auk*, 111, 87-103.
- Greenewalt, C.H. (1968). *Bird Song: Acoustics and Physiology*. Washington DC, Smithsonian Institution Press.
- Hausberger, M., Jenkins, P.F. & Keene, J. (1991). Species-specificity and mimicry in bird song: are they paradoxes? A re-evaluation of song mimicry in the European starling. *Behaviour*, 117, 53-81.
- Herrmann, K. & Arnold, A.P. (1991). The development of afferent projections to the robust archistriatal nucleus in male zebra finches: a quantitative electron microscopic study. *Journal of Neuroscience*, 11, 2063-2074.
- Hienz, R.D. & Brady, J.V. (1988). The acquisition of vowel discriminations by nonhuman primates. *Journal of the Acoustical Society of America*, 84, 186-194.
- Hienz, R. D., Sachs, M.B. & Sinnott, J.M. (1981). Discrimination of steady-state vowels by blackbirds and pigeons. *Journal of the Acoustical Society of America*, 70, 699-706.
- Higgins, P.J. (Ed.) (1999). *Handbook of Australian, New Zealand & Antarctic Birds. Volume 4: Parrots to Dollarbird*. Melbourne, Oxford University Press.
- Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31, 759-767.
- Hose, B., Langner, G. & Scheich, H. (1983). Linear phoneme boundaries for German synthetic two-formant vowels. *Hearing Research*, 9, 13-25.
- Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In R.A. Hinde (Ed.), *Bird Vocalizations*. Cambridge, UK, Cambridge University Press.
- Jusczyk, P.W., Pisoni, D.B., Walley, A. & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America*, 67, 262- 270.
- Kelley, D.B. & Nottebohm, F. (1979). Projections of a telencephalic auditory nucleus - field L - in the canary. *Journal of Comparative Neurology*, 183, 455-470.
- Kewley-Port, D., Watson, C.S. & Foyle, D.C. (1988). Auditory temporal acuity in relation to category boundaries: speech and nonspeech stimuli. *Journal of the Acoustical Society of America*, 83, 1133-1145.
- Kluender, K.R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *Journal of the Acoustical Society of America*, 90, 83-96.
- Kluender, K.R. & Diehl, R.L. (1987). Use of multiple speech dimensions in concept formation by Japanese quail. *Journal of the Acoustical Society of America*, 82, Suppl. 1, S84.
- Kluender, K.R., Diehl, R.L. & Killeen, P.R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195-1197.

- Kluender, K.R. & Lotto, A.J. (1994). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. *Journal of the Acoustical Society of America*, 95, 1044-1052.
- Knudsen, E.I. & Konishi, M. (1978). A neural map of auditory space in the owl. *Science*, 200, 795-797.
- Kuhl, P.K. (1976). Speech perception by the chinchilla: categorical perception of synthetic alveolar plosive consonants. *Journal of the Acoustical Society of America*, 60, S81(A).
- Kuhl, P.K. (1979). Models and mechanisms in speech perception: species comparisons provide further contributions. *Brain, Behavior and Evolution*, 16, 374-408.
- Kuhl, P.K. (1981). Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, 70, 340-349.
- Kuhl, P.K. (1986a). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology*, 45, 233-265.
- Kuhl, P.K. (1986b). Reflections on infants' perception and representation of speech. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 19-30). Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- Kuhl, P.K. (1987). The special mechanisms debate in speech: categorisation tests on animals and infants. In S. Harnad (Ed.), *Categorical Perception: the Groundwork of Cognition* (pp. 355-386). Cambridge, Mass., Cambridge University Press.
- Kuhl, P.K. (1988). Auditory perception and the evolution of speech. *Human Evolution*, 3, 19-43.
- Kuhl, P.K. (1989). On babies, birds, modules, and mechanisms: a comparative approach to the acquisition of vocal communication. In R.J. Dooling & S.H. Hulse (Eds.), *The Comparative Psychology of Audition: Perceiving Complex Sounds* (pp. 379-419). Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- Kuhl, P.K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93-107.
- Kuhl, P.K. & Miller, J.D. (1975a). Speech perception by the chinchilla: voiced voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
- Kuhl, P.K. & Miller, J.D. (1975b). Speech perception by the chinchilla: phonetic boundaries for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 57, S549(A).
- Kuhl, P.K. & Miller, J.D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63, 905-917.
- Kuhl, P.K. & Padden, D.M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, 32, 542-550.
- Kuhl, P.K. & Padden, D.M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, 73, 1003-1010.
- Langner, G., Bonke, D. & Scheich, H. (1979). Selective responses of neurons to vowel sounds in the auditory neostriatum of the mynah bird. In O. Creutzfeld, H. Scheich & Chr. Schreiner (Eds.), *Experimental Brain Research Supplement 2: Hearing Mechanisms and Speech*. Berlin and Heidelberg, Springer-Verlag.
- Langner, G., Bonke, D. & Scheich, H. (1981). Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. *Experimental Brain Research*, 43, 11-24.
- Lasky, R.E., Syrdal-Lasky, A. & Klein, R.E. (1975). VOT discrimination by four to six month and a half old infants from Spanish environments. *Journal of Experimental Child Psychology*, 20, 215-225.

- Liberman, A.M. (1970). Some characteristics of perception in the speech mode. *Perception and its Disorders*, 48, 238-254.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Liberman, A.M., Delattre, P.C. & Cooper, F.S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1, 153-167.
- Liberman, A.M., Delattre, P.C., Cooper, F.S. & Gerstman, L.J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 68, 1-13.
- Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Liberman, A.M. & Mattingly, I.G. (1989). A specialization for speech perception. *Science*, 243, 489-494.
- Lieberman, P. (1977). *Speech Physiology and Acoustic Phonetics*. New York, Macmillan.
- Lisker, L. & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.
- Locke, J.L. (1993). Learning to speak. *Journal of Phonetics*, 21, 141-146.
- Lohr, B. & Dooling, R.J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 112, 36-47.
- Lotto, A.J., Kluender, K.R. & Holt, L.L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102, 1134-1140.
- Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *Journal of Neuroscience*, 3, 1039-1057.
- Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *Journal of Neuroscience*, 6, 1643-1661.
- Margoliash, D. & Fortune, E.S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch's Hvc. *Journal of Neuroscience*, 12, 4309-4326.
- Margoliash, D., Staicer C.A. & Inoue, S.E. (1991). Stereotyped and plastic song in adult indigo buntings, *Passerina cyanea*. *Animal Behaviour*, 42, 367-388.
- Marler, P. (1956). The voice of the chaffinch and its function as a language. *Ibis*, 98, 231-261.
- Marler, P. (1970). A comparative approach to vocal learning: song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, 71 (Monograph 2[2]), 1-25.
- Marler, P. & Peters, S. (1982). Subsong and plastic song: their role in the vocal learning process. In D.E. Kroodsma & E.H. Miller (Eds.), *Acoustic Communication in Birds*. Vol. 2 (pp. 25-50). New York and London, Academic Press.
- Marler, P. and Tamura, M. (1964). Culturally transmitted patterns of vocal behaviour in sparrows. *Science*, 146, 1483-1486.
- Mattingly, I.G. (1972). Speech cues and sign stimuli. *American Scientist*, 60, 327-337.
- Mattingly, I.G. & Liberman, A.M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G.M. Edelman, W.E. Gall & W.M. Cowan (Eds.), *The Comparative Psychology of Audition: Neurobiological Bases of Hearing* (pp. 775-793). New York, Wiley.
- Mattingly, I., Liberman, A., Syrdal, A. & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 2, 131-157.
- Menyuk, P. & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language Acquisition*. (pp. 49-70). Cambridge, UK, Cambridge University Press.

- Miller, J.D. (1977). Perception of speech sounds in animals: evidence for speech processing by mammalian auditory mechanisms. In T.H. Bullock (Ed.), *Recognition of Complex Acoustic Signals* (pp. 49-58). Berlin, Dahlem Konferenzen.
- Miller, J.D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85, 2114-2134.
- Miller, J.D. & Kuhl, P.K. (1976). Speech perception by the chinchilla: a progress report on syllable-initial voiced-plosive consonants. *Journal of the Acoustical Society of America*, 59, S554(A).
- Miller, J.D., Wier, C.C., Pastore, R.E., Kelly, W.J. & Dooling, R. (1976). Discrimination and labelling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *Journal of the Acoustical Society of America*, 60, 410-422.
- Moody, D.B., Stebbins, W.C. & May, B.J. (1990). Auditory perception of communicatory signals by Japanese monkeys. In W.C. Stebbins & M.A. Berkley (Eds.), *Comparative Perception: Vol.II - Complex Signals* (pp. 311-344). New York, Wiley.
- Mooney, R. & Doupe, A.J. (1991). Neurobiology of birdsong: circuits, synapses, and development. *Discussions of Neuroscience*, 7, 100-111.
- Morse, P.A. & Snowden, C.T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics*, 17, 9-16.
- Mountjoy, D.J. & Lemon, R.E. (1995). Extended song learning in wild European starlings. *Animal Behaviour*, 49, 357-366.
- Mundinger, P.C. (1982). Microgeographic and macrogeographic variation in the acquired vocalizations of birds. In D.E. Kroodsma & E.H. Miller (Eds.), *Acoustic Communication in Birds*, Vol. 2 (pp. 147-208). New York and London, Academic Press.
- Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J. & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Nelson, D.A. & Marler, P. (1994). Selection-based learning in bird song development. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 10498-10501.
- Nottebohm, F. (1972). The origins of vocal learning. *American Naturalist*, 106, 116-140.
- Nottebohm, F. (1976). Phonation in the orange-winged Amazon parrot (*Amazona amazonica*). *Journal of Comparative Physiology*, 108, 157-170.
- Nottebohm, F. & Nottebohm, M. (1978). Relationship between song repertoire and age in the canary, *Serinus canarius*. *Zeitschrift für Tierpsychologie*, 46, 298-305.
- Nowicki, S. (1989). Vocal plasticity in captive black-capped chickadees: the acoustic basis and rate of call convergence. *Animal Behaviour*, 37, 64-73.
- Park, T.J. & Dooling, R.J. (1985). Perception of species-specific contact calls by budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 99, 391-402.
- Park, T.J. & Dooling, R.J. (1986). Perception of degraded vocalizations by budgerigars (*Melopsittacus undulatus*). *Animal Learning and Behavior*, 14, 359-364.
- Patterson, D.K. & Pepperberg, I.M. (1994). A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels. *Journal of the Acoustical Society of America*, 96, 634-648.
- Patterson, D.K. & Pepperberg, I.M. (1998). Acoustic and articulatory correlates of stop consonants in a parrot and a human subject. *Journal of the Acoustical Society of America*, 103, 2197-2215.

- Payne, R.B. (1981). Song learning and social interaction in indigo buntings. *Animal Behaviour*, 29, 688-697.
- Peterson, G.E. & Barney, H.L. (1952). Control methods used in the study of the vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253-260.
- Pisoni, D.B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61, 1352-1361.
- Potter, R.K. & Steinberg, J.C. (1950). Toward the specification of speech. *Journal of the Acoustical Society of America*, 22, 807-820.
- Protopapas, A. (1999). Connectionist modeling of speech perception. *Psychological Bulletin*, 125, 410-436.
- Remez, R.E. & Rubin, P.E. (1984). On the perception of intonation from sinusoidal sentences. *Perception & Psychophysics*, 35, 429-440.
- Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S. & Lang, J.M. (1994). On the perceptual organization of speech. *Psychological Review*, 101, 129-156.
- Remez, R.E., Rubin, P.E., Pisoni, D.B. & Carrell, T.D. (1981). Speech perception without traditional cues. *Science*, 212, 947-950.
- Repp, B. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92, 81-110.
- Ritsma, R.J. (1962). Existence region of the tonal residue. *Journal of the Acoustical Society of America*, 34, 1224-1229.
- Rosen, S. & Fourcin, A. (1986). Frequency selectivity and the perception of speech. In B.C. J. Moore (Ed.), *Frequency Selectivity in Hearing*. London, Academic Press.
- Rosen, S. & Howell, P. (1987). Auditory, articulatory and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical Perception: the Groundwork of Cognition* (pp. 113-160). Cambridge, UK, Cambridge University Press.
- Samuel, A.G. & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology (Human Perception and Performance)*, 22, 676-694.
- Saunders, J.C., Rintelmann, W.F. & Bock, G.R. (1979). Frequency selectivity in bird and man: a comparison among critical ratios, critical bands and psychophysical tuning curves. *Hearing Research*, 1, 303-323.
- Scanlan, J.P. (1988). *Analysis of Avian "Speech": Patterns and Production*. PhD thesis, University of London.
- Scanlan, J.P. (1994). The transformation of bird sounds into 'speech'. In R. Togneri (Ed.), *Proceedings of the 5th Australian International Conference on Speech Science and Technology*, Vol. 1 (pp. 81-86). Perth, University of Western Australia.
- Scharff, C. & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *Journal of Neuroscience*, 11, 2896-2913.
- Scheich, H. (1985). Auditory brain organization of birds and its constraints for the design of vocal repertoires. In B. Holldobler & M. Lindauer (Eds.), *Experimental Behavioral Ecology and Sociobiology* (pp. 195-209). Stuttgart and New York, Fischer Verlag.
- Scheich, H., Bock, W., Bonke, D., Langner, G. & Maier, V. (1983). Acoustic communication in the Guinea fowl (*Numida meleagris*). In J.-P. Ewert, R.R. Capranica & D.J. Ingle (Eds.), *Advances in Vertebrate Neuroethology* (pp. 731-780). New York, Plenum Press.

- Scheich, H., Bonke, D. & Langner, G. (1979). Tonotopy and analysis of side-band calls in field L of the Guinea fowl. In O. Creutzfeldt, H. Scheich & Chr. Schreiner (Eds.), *Experimental Brain Research Supplement 2: Hearing Mechanisms and Speech* (pp. 94-109). Berlin and Heidelberg, Springer-Verlag.
- Scheich, H., Langner, G. & Bonke, D. (1979). Responsiveness of units in the auditory neostriatum of the Guinea fowl (*Numida meleagris*) to species-specific calls and synthetic stimuli II: discrimination of iambus-like calls. *Journal of Comparative Physiology*, *132*, 257-276.
- Scheich, H., Langner, G. & Koch, R. (1977). Coding of narrow-band and wide-band vocalizations in the auditory midbrain nucleus (MLD) of the Guinea fowl (*Numida meleagris*). *Journal of Comparative Physiology*, *117*, 245-265.
- Schouten, J.F., Ritsma, R.J. & Cardozo, B.L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America*, *34*, 1418-1424.
- Sharma, A. & Dorman, M.F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, *106*, 1078-1083.
- Sharma, A., Kraus, N., McGee, T., Carrell, T. & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology*, *88*, 64-71.
- Simpson, H.B. & Vicario, D.S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *Journal of Neuroscience*, *10*, 1541-1556.
- Sinex, D.G., McDonald, L.P. & Mott, J.B. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *Journal of the Acoustical Society of America*, *90*, 2441-2449.
- Sinnott, J.M. (1989). Detection and discrimination of synthetic English vowels by Old World monkeys (*Cercopithecus*, *Macaca*) and humans. *Journal of the Acoustical Society of America*, *86*, 557-565.
- Sinnott, J.M. & Brown, C.H. (1997). Perception of the American English liquid /ra-la/ contrast by humans and monkeys. *Journal of the Acoustical Society of America*, *102*, 588-602.
- Sinnott, J.M. & Kreiter, N.A. (1991). Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans. *Journal of the Acoustical Society of America*, *89*, 2421-2429.
- Slater, P.J.B., Eales, L.A. & Clayton, N.S. (1988). Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. *Advances in the Study of Behavior*, *18*, 1-33.
- Snowdon, C.T. (1979). Response of nonhuman animals to speech and to species-specific sounds. *Brain, Behavior and Evolution*, *16*, 409-429.
- Snowdon, C.T. (1990). Language capacities of nonhuman animals. *Yearbook of Physical Anthropology*, *33*, 215-243.
- Soli, S.D. (1983). The role of spectral cues in discrimination of voice onset time differences. *Journal of the Acoustical Society of America*, *73*, 2150-2165.
- Stein, R.C. (1968). Modulation in bird sounds. *Auk*, *85*, 229-243.
- Stevens, K.N. & Klatt, D.H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, *55*, 653-659.
- Streeter, L.A. (1976). Kikuyu labial and apical stop discrimination. *Journal of Phonetics*, *4*, 43-49.
- Striedter, G.F. (1994). The vocal control pathways in budgerigars differ from those in songbirds. *Journal of Comparative Neurology*, *343*, 35-56.
- Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. & Cooper, F.S. (1970). Motor theory of speech perception: a reply to Lane's critical review. *Psychological Review*, *77*, 234-249.

- Suga, N. (1988). Auditory neuroethology and speech processing: complex-sound processing by combination-sensitive neurons. In G.M. Edelman, W.E. Gall & W.M. Cowan (Eds.), *The Comparative Psychology of Audition: Neurobiological Bases of Hearing* (pp. 679-720). New York, Wiley.
- Suga, N., O'Neill, W.E. & Manabe, T. (1978). Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. *Science*, *200*, 778-781.
- Sussman, H.M. (1989). Neural coding of relational invariance in speech: human language analogs to the barn owl. *Psychological Review*, *96*, 631-642.
- Thielcke, G.A. (1966). Ritualized distinctiveness of song in closely related sympatric species. *Philosophical Transactions of the Royal Society of London (Series B)*, *251*, 493-497.
- Thorpe, W.H. (1955). Comments on *The Bird Fancier's Delight*, together with notes on imitation in the subsong of the chaffinch. *Ibis*, *97*, 247-251.
- Thorpe, W.H. (1958). The learning of song patterns by birds, with especial reference to the song of the chaffinch. *Ibis*, *100*, 535-570.
- Thorpe, W.H. (1961). *Bird-song*. Cambridge, UK, Cambridge University Press.
- Thorpe, W.H. & Pilcher, P.M. (1958). The nature and characteristics of sub-song. *British Birds*, *51*, 509-514.
- Tinbergen, N. (1951). *The Study of Instinct*. Oxford, UK, Oxford University Press.
- Tomiak, G.R., Mullenix, J.W. & Sawusch, J.R. (1987). Integral processing of phonemes: evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, *81*, 755-764.
- Vicario, D.S. & Yohay, K.H. (1993). Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. *Journal of Neurobiology*, *24*, 488-505.
- Waters, R.A. & Wilson, W.A. (1976). Speech perception by rhesus monkeys. *Perception & Psychophysics*, *19*, 285-289.
- West, M.J., Straud, A.N. & King, A.P. (1983). Mimicry of the human voice by European starlings: the role of social interaction. *Wilson Bulletin*, *95*, 635-640.
- Whaling, C.S., Solis, M.M., Doupe, A.J., Soha, J.A. & Marler, P. (1997). Acoustic and neural bases for innate recognition of song. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 12694-12698.
- Wyndham, E. (1980). Diurnal cycle, behaviour and social organization of the budgerigar (*Melopsittacus undulatus*). *Emu*, *80*, 25-33.