ANALYSIS OF AVIAN "SPEECH": PATTERNS AND PRODUCTION

A thesis submitted to
the University of London
for the degree of Ph.D.

by James Patrick Scanlan

Dept of Phonetics
and Linguistics,
University College
London

1988

ABSTRACT

Acoustic analysis of avian "speech" is used as a means of approaching various questions concerning vocal communication in birds. In conjunction with anatomical studies, such analysis reveals a sound production mechanism in the psittacine syrinx analogous in some respects to the mammalian larynx. It also reveals, however, that this is not the only syringeal mechanism at the parrot's disposal: vocal redundancy is an important aspect of avian sound production.

Comparison of avian "speech" with species-specific vocalizations shows that some "speech" sounds are adapted (sometimes only slightly) from species-specific sounds. Acoustic features common to bird sounds and human speech are an important factor in this inter-species copying behaviour.

Much avian "speech" concentrates on the second-formant region of the spectrum. This region seems to coincide roughly with the range of resonant frequencies of the syringeal membranes. Lower-frequency information in the "speech" signal is often produced by modulating a second-formant carrier. These findings are correlated with published data on avian sound perception.

In the first chapter, a discussion (and re-definition) of vocal "imitation" and "mimicry" establishes a conceptual framework for the study of avian "speech".

For my parents

Owls have very expressive notes; they hoot in a fine vocal sound, much resembling the *vox humana*, and reducible by a pitch-pipe to a musical key. This note seems to express complacency and rivalry among the males: they use also a quick call and an horrible scream; and can snore and hiss when they mean to menace. *Ravens*, besides their loud croak, can exert a deep and solemn note that makes the woods to echo; the amorous sound of a *crow* is strange and ridiculous; *rooks*, in the breeding season, attempt sometimes in the gaiety of their hearts to sing, but with no great success; the *parrot*-kind have many modulations of voice, as appears by their aptitude to learn human sounds; *doves* coo in an amorous and mournful manner, and are emblems of despairing lovers; the *woodpecker* sets up a sort of loud and hearty laugh; the *fern-owl*, or *goat-sucker*, from the dusk till day-break, serenades his mate with the clattering of castanets. All the tuneful *passeres* express their complacency by sweet modulations, and a variety of melody.

* * * * *

My musical friend, at whose house I am now visiting, has tried all the owls that are his near neighbours with a pitch-pipe set at concert-pitch, and finds they all hoot in B flat. He will examine the nightingales next spring.

- Gilbert White's *The Natural History of Selborne*

Table of Contents

CHAPTER 1:
PROLEGOMENON TO A STUDY OF "TALKING" BIRDS

Page 18. *AIMS OF THIS STUDY*

(a) Sound production
(b) Auditory perception
(c) Central processing of vocal communication sounds

Page 25. *EVOLUTION OF VOCAL LEARNING*

Phylogenetic considerations
Intense speciation
Morphological variation and divergence
Syringeal morphology
Vocal plasticity and neural complexity

Page 35. *VOCAL-LEARNING VARIABLES*

Audition-based control of syringeal muscles
The role of improvisation in birdsong
Auditory feedback
The variable copying/improvisation ratio

Page 45. *DISTINCTIONS AND DEFINITIONS*

Calls and song
"Imitation" and "mimicry"
An interpretation of subsong
An evolutionary hypothesis
Avian "speech" is vocal imitation

Page 57. *AUDITORY AND SOCIAL FACTORS IN IMITATION*

The nature of the auditory template
Ambiguity of the "template" hypothesis
Redundancy of the "template" hypothesis
A species-specific auditory factor
Social factors
Duetting

Page 68. *PSYCHOLOGY OF "TALKING" BIRDS*

Social bonding theories of avian "speech"
Psychological kinship theories of avian "speech"

Page 76. *HUMAN/AVIAN SIMILARITIES IN VOCAL ONTOGENY*

Vocal ontogeny in general
The timing of vocal learning
Possible implications for neurology
Vocal control

Page 84. *PSYCHOLOGY OF "TALKING" BIRDS: A UNIFIED VIEW*

Page 88. *CONCLUSIONS*

Page 89. *NOTES TO CHAPTER 1*

CHAPTER 2:
ANATOMY AND PHYSIOLOGY

Page 92. *INTRODUCTION*

Page 97. *VIBRATILE ELEMENTS*

Page 105. *AIR FLOW AND PRESSURE*

Page 109. *SYRINGEAL MUSCULATURE*

Page 111. *INNERVATION OF THE SYRINGEAL MUSCLES*

Page 112. *FUNCTION OF THE SYRINGEAL MUSCLES*

- *Extrinsic syringeal muscles*
- *Intrinsic syringeal muscles*
- *Psittaciformes*

Page 123. *THEORIES OF SYRINX FUNCTION*

- *Source vibrations*
  - Aerodynamic
  - Mechanical
  - Phonation
  - Membrane resonances
  - Amplitude-frequency coupling
  - Membrane response curve
- *Modulating vibrations*

Page 136. *SYRINX STRUCTURE AND FUNCTION IN PARROTS*

- *Structure*
- *Function*

Phonation mechanism
Comparison with human vocal mechanism

Page 149. *INVOLVEMENT OF THE LARYNX IN VOCALIZATION*

Published observations
Evidence from X-ray films
Laryngeal movements
Ventro-caudad movement of the larynx
Suction-drinking analogy
Glottis closure
Synchronized larynx movements during whistles
Preparatory phase in vocalization
Dorso-caudad larynx mvements

Page 168. *ACOUSTIC RESONANCES OF THE VOCAL TRACT*

Wind-instrument and human-voice analogies
Relationship of fundamental and formant frequencies
Which avian "speech formants" are true formants?

Page 180. *SUMMARY AND CONCLUSIONS*

Page 182. *NOTES TO CHAPTER 2*

CHAPTER 3:
VOCAL REDUNDANCY

Page 187. *INTRODUCTION*

Page 188. *REDUNDANCY IN SOUND PRODUCTION*

Apparent anatomical redundancy
Complexity, redundancy and plasticity
The budgerigar's [ɔ]
Sound-production hypotheses: general principles
Hypothetical model for the production of both sounds
An inter-species comparison

Page 206. *THE PERCEPTION OF MODULATED SOUNDS*

Modulation in bird sounds
"Harmonic" relationships
Perception of low pitch in "harmonic" complexes
Anharmonic relationship
Perception of low pitch in anharmonic complexes
Frequency modulation
A possible FM mechanism
"Primary" and "secondary" modulations
Combinations of AM and FM
A species-specific acoustic pattern contrast?

Page 219. *CONCLUSIONS*

Intrinsic syringeal muscles and vocal plasticity

Page 222. *SUMMARY*

Page 223. *NOTES TO CHAPTER 3*

CHAPTER 4:
PERCEPTION OF COMMUNICATION SOUNDS

Page 226. *INTRODUCTION*

Page 228. *PERCEPTUAL COMPLEXITY*

The avian auditory system
- Introduction
- Physical constraints
- The basilar membrane
- The basilar papilla
- Cochlear vibrations
- Acoustic analysis in the cochlea

Avian hearing: data from behavioural tests
- Audibility curves
- Hearing range
- Frequency discrimination
- Intensity discrimination
- Temporal resolution

Frequency selectivity
- Introduction
- Critical bandwidths
- Nonsimultaneous masking
- Critical ratios

Patterns of auditory projection
- The cochlear nuclei
- The mesencephalon and telencephalon
- Voco-auditory integration

Patterns of neuronal response

Relationship to cochlear events

Periodicity analysis

"Simple" and "complex" neurons

"Formant" detectors

Response to vowel stimuli

Page 253. *SPEECH SOUND PATTERNS AND THEIR PROCESSING*

Vowels

Spectral features

Animal studies

Ambiguities in interpretation: spectral envelope shape

Ambiguities in interpretation: fine structure

Patterning and redundancy in vowel sounds

Stop consonants

Complexity and redundancy in the temporal domain

Voice onset time

Spectral cues for the voicing distinction

Avian preference for the temporal voicing cue

Systematic VOT variations

Cross-language and infant studies

Animal studies

Perception of non-speech stimuli

Transition slope and direction

Transition length

Discrete and graded signals

Acoustic grading and vocal plasticity

Continuous and categorical perception

Feature detection

Non-human perception and reproduction of speech

Neural lateralization

Lateralized perception of communication sounds

Lateralization of temporal auditory processing

Lateralization of transitional cues

Single-formant avian "speech"
Sufficiency of F2
Avian perception and reproduction of F2
Single-formant human speech
Whistled languages
Whistles as "bio-communication"

Page 293. *SUMMARY*

Page 295. *NOTES TO CHAPTER 4*

CHAPTER 5

TRANSFORMATION OF BIRD SOUNDS INTO "SPEECH"

Page 299. *INTRODUCTION*

Page 302. *ONTOGENY OF AVIAN "SPEECH"*

Page 305. *TRANSFORMATION OF NATURAL SOUNDS INTO "SPEECH"*

Voicing
Spectral envelope shape and Fx contour
Formant pattern
Single formant "talking"
High front vowels

Page 318. *PERCEPTION OF AVIAN "SPEECH" BY HUMAN LISTENERS*

Importance of context: an exploratory experiment
Listeners' expectations: further experiments

Page 328. *SUMMARY*

Page 329. *NOTES TO CHAPTER 5*

CHAPTER 6:

"COMPARATIVE PHONETICS"

Page 332. *INTRODUCTION*

Detailed budgerigar-human comparison
Energy distribution
Vibration patterns
Timing
Formant structure
Formant transitions

Page 352. *PERCEPTION OF SPECH PATTERNS IN MAN AND BIRD*

Relationship of perceived pitch and F2 frequency
Nymphicus hollandicus
Psittacus erithacus
Relationship of perceived pitch and F1 frequency
Ara chloroptera
Whispering

Page 364. *AVIAN INTER-SPECIES COMPARISONS*

The parrot's vocal mechanism
The mynah's vocal mechanism
"Speech" patterns in mynah and parrot

Page 376. *SUMMARY*

Page 378. *NOTES TO CHAPTER 6*

Page 380. *APPENDIX TO CHAPTER 6*

Page 388. *GENERAL CONCLUSION*

Page 399. *REFERENCES*

Table of plates

PLATE 2.1 Page 94
PLATE 2.2 Page 102
PLATE 2.3 Page 142
PLATE 2.4 Page 144
PLATE 2.5 Page 150
PLATE 2.6 Page 153
PLATE 2.7 Page 164
PLATE 3.1 Page 191
PLATE 3.2 Page 207
PLATE 3.3 Page 210
PLATE 5.1 Page 306
PLATE 5.2 Page 309
PLATE 5.3 Page 315
PLATE 5.4 Page 325
PLATE 6.1 Page 333
PLATE 6.2 Page 336
PLATE 6.3 Page 340
PLATE 6.4 Page 353
PLATE 6.5 Page 358
PLATE 6.6 Page 365
PLATE 6.7 Page 367
PLATE 6.8 Page 372
PLATE 6.9 Page 382

ACKNOWLEDGMENTS

My thanks go first of all to Professor Adrian Fourcin for providing me with the opportunity for this study, and for his advice, interest and encouragement throughout. I am also greatly indebted to the encouragement and support of Professor R.-G. Busnel and Dr Marie-Claire Busnel, and to their kindness and hospitality during several visits to Paris.

This work would not have been possible without the help of many people in several fields of interest. Those who assisted in the preparation of the anatomical chapter include Vicki Aitken, Glen Baggott, Philip Burton, Irene Christie, Professor George Du-Boulay, Frances MacCurtain, Nick Noscoe, Stuart Rosen, Barry Salt, George Smith, Nancy Thomas, and Robert Warner.

I would like to thank everyone in the Phonetics Department, University College London, for their kindness and ready assistance, and for providing a pleasant atmosphere in which to work. Likewise, my thanks to everyone at the Laboratoire D'Acoustique Animale, Jouy-en-Josas, for their hospitality during a three-months' study program there in 1983, under the supervision of Professor and Marie-Claire Busnel. The program was funded by Naturalia et Biologia, Paris.

Indispensable to this work have been the birds themselves, and their owners who welcomed me into their homes.

SUBJECTS AND INSTRUMENTS

The avian subjects of this research were all pets, and were recorded in their homes with the generous assistance of their owners. About 20 birds were recorded in all. The "speech" of eight of these provided the material for the photographic plates. These were: four budgerigars, two grey parrots, a green-winged macaw, and a mynah. The recording of the cockatiel was provided by Dr Hans-Heiner Bergmann, and those of the glaucous-winged gull and the raven by Cecil Hyndman.

Recordings were made either with a Uher Report tape recorder or a Uher CR 240 cassette recorder, using a Sony Electret condenser microphone, ECM-30. Spectrograms were produced on a Kay Sona-Graph 6061B, using both 300 Hz and 45 Hz bandwidth filters. Amplitude spectra were produced on an Ono Sokki dual channel analyser, CF-910, and a Brüel and Kjaer spectrum analyser type 2131. Waveforms were drawn by a Hewlett Packard X-Y plotter after sampling of recorded material by a PDP12 computer.

CHAPTER 1:
PROLEGOMENON TO A STUDY OF "TALKING" BIRDS

AIMS OF THIS STUDY

Avian "speech"[1] is a familiar example of inter-species vocal copying. Acoustically, it is a unique source of data on the audio-vocal processing of speech sounds in a group of non-human organisms. Thus, although this study is founded principally on acoustic data, its five main aims may be organized under these headings: *(a) production*, *(b) peripheral reception*, and *(c) central neural processing* of human speech sounds by birds.

(a) Sound production

AIM 1: Anatomy and physiology. In the perception of complex sounds there is no one-to-one relationship between a particular auditory impression and a particular acoustic signal: acoustic signals differing radically in physical characteristics can produce identical auditory impressions (including those of pitch and timbre, as discussed in Chapter 3) in an individual listener.

This contributes to the possibility of vocal copying between species (e.g. avian and primate) differing significantly in vocal anatomy and physiology, and thus in the physical properties of all complex sounds they may produce - including interspecies copies.

The *first aim* of this study is both to investigate the anatomy and physiology of non-human vocal systems which are capable of producing in a human listener the auditory impression of "speech" and, with the aid of acoustic analysis, to explain this feat of aural illusion. Chapter 2 is devoted to anatomical and physiological investigations, the results of which are subsequently related to the acoustic analyses of Chapters 3 and 6.

AIM 2: Vocal plasticity. The acoustic phenomenon of a vocal copy is a product of the sensory-motor process of vocal learning.

Vocal learning can be defined as the use of auditory information (including feedback) in the formation of an individual's vocal repertoire (Kroodsma, 1982; Nottebohm, 1972). Thus, in vocal learning, sound production is guided by perception: i.e., the animal matches its vocal output, using auditory feedback, with its auditory impression of the model sound. This matching process must involve a degree of vocal flexibility (or "plasticity").

From these propositions - and in the light of studies (reviewed in Chapter 3) which indicate that in birds, as in man, there is no unique relationship between auditory impression and acoustic signal - it follows that a learnt vocalization could be produced in an individual by differing vocal mechanisms, and thus have differing acoustic properties, while being perceived by the copier (and/or the model) as the "same" sound[2]. Such apparent interchangeability of sound-production gestures in an individual has been referred to as "vocal redundancy" (e.g. by Gaunt and Gaunt, 1977).

The *second aim* of this study is to investigate the concepts of vocal plasticity and redundancy in a context - the copying of highly complex, allospecific communication sounds by birds - in which the dissociation of particular vocal gestures from particular auditory impressions in the copier should be most pronounced. Chapter 3 examines in some detail an example of such vocal redundancy in avian "speech".

(b) Auditory perception

AIM 3: Avian hearing. It follows from the definition of vocal learning above that interspecies vocal copies must bear the imprint of the copier's species-specific auditory system. Physical comparisons of interspecies copies with the sounds copied can thus help specify such functions of the copier's auditory system as frequency analysis, discrimination and selectivity; audibility thresholds; intensity discrimination; temporal resolution. Such observations can be correlated with the results of psychophysical and electrophysiological tests of these auditory functions.

The *third aim* of this study is to present - chiefly in Chapter 6 - comparative data from acoustic analyses of human speech (the models) and avian "speech" (the copies), and to interpret these in the light of available data - reviewed in Chapter 4 - from auditory tests on birds. Avian "speech" can thus be used to investigate the auditory world of the bird.

(c) Central processing of vocal communication sounds

AIM 4: Pattern recognition and reproduction. For the (free-living) bird, the most important components of this auditory world are its species-specific communication sounds.

In birds (and in mammals) perceptual processing of these species-specific sounds does not occur at the periphery - as it does, for example, in frogs (Frishkopf and Goldstein, 1963) - but at more central levels of perception; i.e., avian recognition of species-specific sounds involves central pattern processing of a complex auditory input rather than a mechanical filtering of acoustic signals at the periphery.

The extent to which such patterning may be involved in the perception of allospecific communication sounds is an

important question which can be elucidated by the study of interspecies vocal copies. For example, Klatt and Stefanski (1974) and Thorpe (1969) presented preliminary spectrographic evidence that the Indian hill mynah (*Gracula religiosa*) reproduces formant frequency relationships (i.e. patterns), rather than absolute frequency values, in its copies of human vowels. (Psychophysical tests have shown that discrimination of vowel sounds by some non-human mammals is based, as in human listeners, on the relationships of formant frequencies: see J. D. Miller, 1977, for a review.)

Single-unit recording by Langner *et al.* (1979, 1981) has given some insight into the neural mechanisms involved in the perceptual processing of vowels by *Gracula*. Their data show that some neurons in the mynah's telencephalon respond selectively to - i.e. discriminate among - two-formant synthetic vowels. This and related work will be reviewed in Chapter 4.

Hulse (1987) and Hulse *et al.* (1984) discovered similarities - and differences - between avian and human performance in the serial processing of rhythmic and pitch-contour patterns. They chose the European starling (*Sturnus vulgaris*) as their subject species because of its ability to copy non-species-specific sounds - including speech.

The *fourth aim* of this study is to contribute to an assessment of the nature of patterning in avian copies of human speech sounds - and, therefore, in the perceptual processing of these allospecific vocalizations by birds.

AIM 5: Acoustic cues in vocal learning. As, in vocal-learning animals, sound production is guided by perception (and must therefore be relatively plastic), the transmission and maintenance of vocal species-specificity in such animals is more likely to be ensured by perceptual filtering for the selection of appropriate models than by stereotypy of production.

Thus, if perceptual mechanisms control the maintenance of vocal species-specificity, the learning of allospecific vocalizations appears to involve a perceptual anomaly . . . an apparent anomaly which could be a key to understanding those very perceptual mechanisms.

Those mechanisms, in man and songbird, have been referred to, collectively, as the language/song-learning "template". The "template" is conceived of as a perceptual filter which, during the period of vocal ontogeny, selects species-specific vocalizations from the range of audible phenomena as the models to be copied, and provides an innate organizational framework for learning them.

The acoustic correlates of these neural mechanisms are acoustic cues. Considered from an auditory standpoint, an acoustic cue is a simple (e.g. single-parameter) acoustic feature of a complex communication sound to which - and to variations of which - the nervous system of the receiver appears to be particularly sensitive. Chapter 4 will review the evidence - still somewhat controversial - for believing that such cues relate to psychophysical properties of the nervous system.

A cue's function in acoustic communication is, firstly, in signalling to (i.e. "cueing") the receiver that the sound is communicatively significant and, secondly, in providing a set of maximally perceptible acoustic contrasts for the encoding of information.

Psychophysical, electrophysiological, and auditory-evoked-response testing, reviewed in Chapter 4, has shown that the response patterns of non-human animals to some of the acoustic cues of speech (and to their variation) is similar to the response patterns of human subjects. These results suggest that a speech cue could also have communicative significance in the species-specific vocalizations of a non-human animal.

Such inter-specificity of acoustic communication cues

could help to explain the apparent perceptual anomaly of interspecies vocal copying - the hypothesis being that the copier recognizes and learns to produce, in its own way, acoustic cues which correspond to perceptual mechanisms used in its species-specific vocal learning.

My *fifth aim* is to test this "bio-communication-cue hypothesis".

An obvious test is to make a physical comparison of any "speech cues" to be found in avian copies both with their human models and with possible cognates in the species-specific vocalizations of the copier. Reports on such comparative acoustic analyses are to be found in Chapters 4, 5 and 6.

A corollary of the bio-communication-cue hypothesis is that the "speech" of an avian copier should bear a strong physical resemblance (at least in respect of some important acoustic cues) to its species-specific vocalizations. Only minimal modification of the species-specific sound would then be needed for the human listener, sensitive to some of the same acoustic communication cues as the bird, to have the auditory impression of "speech". Chapter 5 tests the bio-communication-cue hypothesis from this angle: firstly by a study of "verbal" development in the budgerigar (*Melopsittacus undulatus*), and secondly by a series of perceptual experiments in which human listeners are asked to categorize budgerigar "speech" and non-speech sounds.

Integration of aims. Finally, it must be emphasised that none of the three aspects of interspecies vocal copying discussed above - plasticity of production, acoustic cues and perceptual patterning - is sufficient to account for the occurrence of avian "speech". The *overall aim* of this thesis is to indicte how the three may interact - through a complex network of intraspecific and interspecific connexions - to enable birds to copy speech sounds and, reciprocally, human listeners to

perceive these avian copies as "speech".

The rest of Chapter 1 is an attempt to establish a conceptual framework for the study of "talking" birds (beginning with some evolutionary considerations).

This involves a survey of avian vocal copying behaviour generally, and an attempt to understand avian "speech" in terms of this broad perspective.

The discussion seeks to enumerate those physical, neuro-physiological and behavioural attributes of an avian species which contribute to "talking" ability. The argument centres on new definitions of (and distinctions between) "imitation" and "mimicry".

EVOLUTION OF VOCAL LEARNING

To define vocal learning as the guidance of sound production by auditory perception implies an hypothesis which links, in phylogenetic terms, the development of vocal learning with that of neuro-muscular vocal control systems.

This hypothesis is supported, pragmatically, by observations of a direct relationship between vocal-learning ability and complexity of syringeal musculature in birds (Gaunt, 1983; Nottebohm, 1972). It is given theoretical support by current systems of avian taxonomy and phylogeny (Van Tyne and Berger, 1959; Ames, 1971) which regard both vocal learning and complex syringeal musculature as characteristic of species "highly derived" (Kroodsma, 1982) from "the ancestral avian stock" (Nottebohm, 1972).

Thus each of the two taxonomic groups of particular interest in a study of avian "speech" - the family Psittacidae (Psittaciformes), and the suborder Passeres (Passeriformes) - is distinguished by complex syringeal musculature and by an "advanced" status (Mayr, 1958; Storer, 1960) in systems of evolutionary taxonomy.

Phylogenetic considerations

Two general evolutionary factors have been considered of particular relevance to the development of vocal learning in such "advanced" groups: (1) intense speciation, and (2) morphological variation and divergence.

Intense speciation. Nottebohm (1972) pointed out that "in 8,580 living species of birds, 4,015 are oscine songbirds, that is members of the suborder Passeres of the Passeriformes". Among orders of non-passerine birds, the two families which have most species are "the Psittacidae (parrots) and the Trochilidae (hummingbirds), with 317 and 320 species respectively" (Nottebohm, *ibid.*). Vocal learning has been reported in hummingbirds (Snow, 1968; Wiley, 1971) as well as in parrots.

These statistics led Nottebohm to believe that there may be a causal link between the evolution of vocal learning and the evolutionary mechanisms underlying - or the evolutionary consequences of - intense speciation. "In no case", he wrote, however, "is it possible to say whether vocal learning preceded and aided rapid speciation or resulted from it" (Nottebohm, 1972). Marler (1970b), on the other hand, opted for the precedence of rapid speciation, arguing that the consequent need for correspondingly rapid development of species-specific vocalizations could have favoured the emergence of a dominant role for learning in the vocal ontogenies of some birds.

(Kroodsma, 1982, in discussing the intense-speciation hypothesis of Nottebohm and Marler, suggested that the apparent correlation could simply reflect a greater probability of finding vocal learning in families with large numbers of species. Although only further study - particularly of non-passerine birds - will resolve this issue, the genetic data mentioned in the following two paragraphs certainly lend support to the intense-speciation hypothesis.)

Observations of vocal dialects in many songbird species (reviewed by Harris and Lemon, 1972; Mundinger, 1982; Nottebohm, 1969; Thielcke, 1969) are regarded as primary evidence of vocal learning in these birds.

The discovery of genetic differences among contiguous dialect populations in certain sparrows - *Zonotrichia*

leucophrys (Baker, 1974, 1975) and *Z. capensis* (Handford and Nottebohm, 1976; Nottebohm and Selander, 1972) - has supported an hypothesis that song dialects indicate "incipient speciation" (Marler and Tamura, 1962). Thus, in view of the causal connexion of song dialects with vocal learning, these genetic data provide empirical support for the hypothesized relationship between vocal learning and rapid speciation in songbirds.

Morphological variation and divergence. Nottebohm (1972) noted that, exceptionally, there appears to be no vocal learning in the passerine family Tyrannidae (suborder Tyranni) which has 374 species in 119 genera (Storer, 1971). "Despite intense speciation the vocalizations of tyrannid flycatchers reveal an impressive stereotypy and invariability" (Nottebohm, *ibid.*).

He suggested that the Tyrannidae may not have evolved sufficient interspecific morphological variation to allow the development of learning-assisted vocal ontogenies, as such ontogenies entail intraspecific variation of acoustic signals (e.g. dialects). "Closely related sympatric species [of Tyrannidae] often cannot be told apart in the field on sight. . . Perhaps learnt elaboration of auditory signals is only permissible when species identity is clearly established by other cues, such as visual ones" (Nottebohm, *ibid.*).

Plumage colour is an important visual cue to species identity. In contrast to the "very conservative colouration" (Nottebohm, *ibid.*) of the Tyrannidae, the brilliant and varied plumage colours of the Psittacidae and the Trochilidae generally allow unambiguous species identification in the field.

As well as this interspecific morphological variety, the Psittacidae and Trochilidae also exhibit a remarkable familial homogeneity and distinctiveness: in evolutionary terms, a highly-marked divergence from other avian

groups. (This is in line with a general dictum of evolutionary theory that "those characters that are highly variable within populations are usually highly divergent between populations" - Baker, 1982.)

"By definition, a family is a monophyletic group, and many families are so well marked that there is not the slightest disagreement among ornithologists about which species should be included in those families. Among such families are the parrots, the hummingbirds, and toucans" (Van Tyne and Berger, 1959).

The above quotation introduces, in connexion with familial distinctiveness in the vocal-learning parrots and hummingbirds, a third non-oscine family, the toucans (Ramphastidae), in which vocal learning has been reported. However, this report (Wagner, 1944) has yet to be substantiated by acoustic analysis.

In the parrots, morphological distinctiveness is emphasised by relationships with other avian orders which are at best "tenuous and indistinct" (Serpell, 1985). This is usually regarded as evidence of their early phyletic separation from the ancestral avian stock. (See, for example, Serpell, *ibid.*).

At the family level, then, the emergence of vocal learning in the Psittacidae (and the Trochilidae), even before the proliferation of species, could have been facilitated by early-established morphological distinctiveness from other avian groups, thus allowing the variability of vocal signals necessarily entailed by vocal learning without seriously compromising vocal recognition.

In contrast to the parrots, the songbirds are usually regarded as the most recently-evolved avian group. The plethora of taxonomic problems associated with this group - in contrast, once again, to the parrots - is thought to reflect the recent emergence of families which "have not yet evolved far enough to be sharply demarked" (Van Tyne and Berger, 1959). These problems, however, are largely

created by taxonomists in an effort to group the 4,000-odd species of songbird into a convenient number of families.

The general resemblance among songbirds is so high (Mayr, 1958) that, as Van Tyne and Berger (1959) have pointed out, "if we rigidly maintained the same standards as we work through the birds of the world, setting up family groups from the ostriches and rheas down to the modern songbirds, we would probably place all passerine birds (or at least the songbirds - the suborder Passeres) in one family."

In few - if any - morphological features is the homogeneity of the Passeres more marked than in the syrinx - the "organ of voice" (Mueller, 1878). In phylogenetic terms, "the great homogeneity of syringeal structure among nearly 4,000 oscine species strongly suggests a single origin for the group" (Ames, 1971).

Thus, in studying the songbird syrinx, the more general evolutionary factors discussed above in relation to vocal learning - intense speciation and familial divergence - can be focused on the evolution of a morphological feature particularly associated with vocal learning.

Syringeal morphology.

Nitzsch (1829), who introduced the application of syringeal morphology to the classification of birds, related the acoustic complexity of songbird vocalizations to the complex musculature of the oscine syrinx.

Müller (1847, 1878) established the subordinate taxonomy of the Passeriformes on the basis of syringeal myology - the suborder Oscines consisting of those families "endowed with the full number of five pairs of song muscles" (Gadow, 1893). Müller distinguished generally five pairs of intrinsic syringeal muscles in songbirds - as did Cuvier (1805) and Savart (1826) before

him.

A general association by ornithologists of song with intrinsic syringeal muscles (often called "song muscles" in the nineteenth-century literature, as in the quotation from Gadow, above) followed the publications of Nitzsch and Müller.

Twentieth-century developments in the study of syringeal mechanics, reviewed in Chapter 2, have revealed an essential distinction in vocal physiology between birds which have intrinsic syringeal muscles and those which lack such muscles. This distinction can be summarized by the adjectives "active" and "passive".

In active sound production, there is an integral relationship between frequencies of syringeal membrane vibration (as inferred from acoustic analysis) and frequencies of syringeal muscle activity (as determined by electromyographic recording). In passive sound production there is no such relationship, muscular contraction serving primarily to adjust the position, mass and stiffness of membranes which vibrate in response to a combination of air-pressure and tissue-elastic forces. (The "myoelastic-aerodynamic" model of human voice production - Van den Berg, 1958 - is the classic theoretical analysis of such a passive system.)

According to the available electromyographic data, only intrinsic syringeal muscles seem capable of the rapid pulsatile activity characteristic of active sound production in birds (see Chapter 2, under "Theories of syrinx function"); on the other hand, contraction of intrinsic and/or extrinsic muscles can effect the less precise preparatory and sustaining functions necessary for passive sound production (Gaunt and Gaunt, 1985).

Thus the two avian groups of principal concern in the study of "talking" birds - the suborder Passeres and the family Psittacidae, both distinguished by the possession of intrinsic syringeal muscles - are also the two groups (together, perhaps, with the Trochilidae) capable of

actively controlling mechanical vibrations within the syrinx. Chapters 3 and 6 will seek to explain how this active control enables them to generate - by highly un-speechlike means - acoustic signals which are perceived as "speech" by human listeners.

Vocal plasticity and neural complexity

The "active/passive" distinction in avian vocal physiology is in line with current theories on the evolution of intrinsic syringeal muscles. The hypothesized "pico-passerine" syrinx (Ames, 1971), ancestral to that of the songbirds, lacked intrinsic muscles, and was thus capable of only passive sound production. Subsequently, with the evolution of the intrinsic muscles, songbirds developed the ability to exert active neuro-muscular control over at least some acoustic features of their vocal output.

It may be possible, with the help of acoustic analysis, to distinguish elements of a bird's vocal output which are passively produced (and thus relate to a more "primitive" vocal mechanism) from elements which are actively produced by means of an "advanced" mechanism. Chapter 3 examines repetitions of a speech phrase by a budgerigar (*Melopsittacus undulatus*) in which the same bird uses two markedly different methods of sound production in copying a particular speech sound. Acoustic analysis, in conjunction with the physiological studies reported in Chapter 2, indicates that one of these methods is basically passive, the other active. It is suggested that this alternation of passive and active production patterns in copies of the same non-species-specific sound is a form of vocal redundancy which reflects stages in the course of avian vocal evolution.

(In this connexion, as in many others, inter-species vocal copying - particularly that of acoustically complex sounds such as speech - is an ideal context for studying

the sound-production possibilities of avian vocal mechanisms, as there is presumably an element of vocally-plastic experiment in such behaviour not to be found in the normal species-specific repertoire.)

Gaunt and Gaunt (1985) remarked on the ability of the intrinsic syringeal muscles of parrots to engage in either active or passive sound production. While relating the passive mechanism to the "primitive condition" and the active mechanism to the "derived condition", they pointed out that this evolutionary sequence in the development of sound-production techniques does not necessarily determine which of the two production patterns a parrot (or songbird) may employ in the generation of a particular acoustic pattern. In other words, the bird may use the "primitive" mechanism for producing vocalizations which are acquired by the "advanced" process of vocal learning. In relation to any acoustic pattern - innate or learnt - "the activation pattern [re]presents only one of several equally possible techniques available to a psittacid syrinx, not a stage in a sequence" (Gaunt and Gaunt, *ibid.*). This argument accounts for the phenomenon described and discussed in Chapter 3: the alternation of passive and active mechanisms in repeated copies of the same speech sound.

The general evolutionary picture which emerges is one of increasingly greater neuro-muscular control over vocalizations whose acquisition is increasingly dependent on auditory information. Together, these two complementary developments - vocal learning and precise neuro-muscular control - result in an increasing variability of motor and acoustic vocal patterns. This variability, in turn, requires the development of more complex central neural networks for voco-auditory pattern processing. The degree of complexity of these central patterning networks - rather than the degree of complexity of the syringeal musculature - should be regarded, then, as the primary index of vocal learning

ability in birds.

Gaunt (1983) pointed out that while an intrinsic syringeal musculature provides the potential for motor plasticity so important in vocal copying, that potential can be realised only after a "neurological shift" in the central nervous system. Gaunt's term "neurological shift" implies more than just the development of central control over the syringeal muscles; in the terms of the argument developed in the preceding paragraph, it can be interpreted as the crossing of a "complexity threshold" in audio-vocal neural patterning.

While the nature of this threshold awaits further investigation, there is some evidence that it may be lower than has generally been considered possible, and may even be reached in birds not possessing intrinsic syringeal muscles. For example, Figure 6.1B (Chapter 6) presents an instance of speech-sound copying by a glaucous-winged gull, *Larus hyperboreus* (family Laridae, order Charadriiformes).

The literature contains reports of vocal learning in three species of Phasianidae (order Galliformes): *Tympanuchus cupido* and *T. phasianellus* (Sparling, 1979), and *Colinus virginianus* (Baily and Baker, 1982). Vocal redundancy - the availability of alternative neuro-motor patterns in producing the one vocal signal, as mentioned above in connexion with budgerigar "speech" - has been reported in two further galliform species: the domestic chicken, *Gallus gallus* (Gaunt and Gaunt, 1977) and the Guinea fowl, *Numida meleagris* (Maier, 1982). Vocal redundancy implies a level of neural complexity and abstraction which may be adequate for some form of vocal learning.

This initial evidence of vocal learning, plasticity and redundancy in the Galliformes - an order relatively low in the taxonomic hierarchy - requires confirmation and corroboration. Meanwhile, Gaunt (1983) believes that a "latent genetic potential for a neurological shift may

be widespread" among birds.

To summarize: available data suggest that all birds capable of active neuro-muscular control of syringeal vibrations have vocal ontogenies dependent (to a variable extent) on learning. All reports of (and theories on) such active syringeal control have implicated the phylogenetically "advanced" intrinsic syringeal muscles. The capacity for vocal learning must ultimately be determined, however, not merely by the flexibility of neuro-muscular sound-production control, but by the complexity of central neural networks which serve to integrate the acquisition and production of learnt vocalizations. While the complexity of these networks necessarily increases along with the evolution of more complex production mechanisms, the neural "complexity threshold" necessary for vocal learning may be reached, in some species, as a result of evolutionary factors unrelated to such syringeal development. In these species, vocal copying may perhaps be effected by learning-based control of respiratory (i.e. abdominal/tracheal) muscles rather than syringeal muscles.

VOCAL-LEARNING VARIABLES

The discussion above has drawn attention to some myological and neurological correlates of avian vocal learning, on the basis of which it seems appropriate to look for potential "talking" birds among species having an intrinsic syringeal musculature and/or a level of audio-motor neural complexity commensurate with plastic vocal behaviour. The following sections will show that, among birds with these basic qualifications, there are complex interspecific variations in learning-based vocal ontogenies which can either inhibit or release such a potential. Discussion is restricted to songbirds, as most studies of avian vocal learning have been concerned with song.

Audition-based control of syringeal muscles

Marler (1970b, 1975a) observed that individuals of several *Zonotrichia* species - species having highly divergent song patterns in the wild - developed abnormal vocalizations which were much less mutually divergent when surgically deafened before the onset of full song. He regarded this as further evidence both of a generic relationship and of learning-assisted vocal ontogenies in *Zonotrichia* species, the vocalizations of these early-deafened (i.e. learning-deprived) songbirds being a "basic reflection of the [common] properties of the sound producing equipment" (Marler, 1970b).

In the terms of the evolutionary argument developed above, these abnormal vocalizations may represent a common, ancestral ("pico-passerine"), passively-produced substratum of *Zonotrichia* song in birds deprived of the auditory experience necessary for active control of the syringeal muscles. (This basic song pattern is often

greatly obscured in early-deafened songbirds by turbulent airflow noise, and by random vibrations of syringeal elements normally under precise muscular control.) Upon this common substratum, vocal learning has constructed the diversity of song traditions to be found among *Zonotrichia* species.

Although the results of analogous experiments on species of the genus *Junco* are somewhat different, their very difference serves to reinforce the present argument by providing another perspective on the same evolutionary process. Konishi (1964) found that early-deafened Oregon juncos (*Junco oreganus*) developed a song similar in many respects to that of their wild conspecifics. Early-deafened Mexican juncos (*Junco phaeonotus*, also referred to in the literature as "Arizona junco"), on the other hand, developed a song similar to that of wild Oregon juncos rather than to that of their own species.

The basic form of Oregon junco song is a simple repetition of a single syllable. Mexican junco song, however, is more complex: it "generally consists of three or more parts, including trills and phrases" (Konishi, *ibid.*). The apparent reversion of deafened Mexican juncos to the simpler song pattern of the Oregon junco suggests that this simpler pattern is closer to that of the common *Junco* ancestor than the more complex Mexican junco song.

One of the aberrations most noticeable to Konishi in the songs of all deafened juncos was irregular and variable patterns of amplitude and frequency modulation. The examination of songbird syrinx physiology in Chapters 2 and 6 will show that at least some of these modulations in the acoustic signal are directly linked to activity of the intrinsic syringeal muscles. On the other hand, features of junco song which relate to activity of the respiratory muscles - such as song duration, and the temporal patterning of songs into "song clusters" - remained unchanged after early deafening. Thus these results add to the body of evidence indicating that early

auditory experience is necessary for the control of the intrinsic syringeal muscles during avian vocalization, whereas lack of auditory experience does not affect the sound-production activity of respiratory muscles.

Marler *et al.* (1972) reported that early deafening of red-winged blackbirds (*Agelaius phoeniceus*) produced results analogous to those for Konishi's Oregon juncos: i.e., some of the patterning of wild song was preserved. On the other hand, results for the chaffinch, *Fringilla coelebs* (Nottebohm, 1967, 1968) and the cardinal, *Richmondena cardinalis* (Dittus and Lemon, 1970) were analogous to those for the *Zonotrichia* sparrows: i.e., there was little resemblance to natural song patterning.

The role of improvisation in birdsong

Auditory feedback. Deafening experiments can help determine, for any particular songbird species, the relative importance of innate[3] substratum and learnt elaboration in that species' song. The learnt-elaboration component of song is itself, however, subject to another interspecific variable: the relative importance of copying and improvisation. This variable can be determined, for any particular species, by the procedure of acoustic isolation.

Early deafening deprives a young bird of auditory feedback as well as the opportunity of hearing conspecific adult song. Acoustic isolation, on the other hand, deprives the bird of only the song model, while preserving its ability to monitor its own voice. Thus, in the songs of such isolates, acoustic features representing an innate substratum are refined - as they are in wild birds but not in early-deafened birds - by a complex periodicity of waveform resulting from audition-based control of the intrinsic syringeal muscles.

(The discussion in this section concentrates, for simplicity, on songbird species in which the innate

substratum is minimal - i.e., in which early deafening results in the most severe deficits to normal song patterns. In birds such as the canary, on the other hand, early deafening has a less drastic effect on song development - Marler and Waser, 1977.)

In acoustic analyses, one immediately obvious result of this learnt control over spectral composition, when compared with the production of innate patterns by early-deafened birds, is a general narrowing of overall bandwidth. Chapters 2 and 3 will deal with those syringeal mechanisms for amplitude and frequency modulation which are capable of concentrating spectral energy within a relatively narrow band of frequencies.

Another common effect of syringeal control is the lengthened duration of discrete temporal elements in the song, resulting from reduced volume velocity of airflow through the syringeal valve during sound-generating expiration (Gaunt *et al.*, 1973, 1976). The amount of acoustic energy eliminated by narrowing the bandwidth could thus be considered as proportional to the energy contained in the temporal extension of the element. A complicating factor in this equation, however, is the fact that intrinsic muscular control of the syringeal valve results in more efficient conversion of mechanical to acoustic energy (Brackenbury, 1979b) - i.e., reduced dissipation of energy as heat (Brackenbury, 1977). The comparative acoustic analyses of narrow-band and wide-band vocalizations in Chapter 3, showing relatively greater amplitude of spectral peaks in the narrow-band signals, give an indication of this increased efficiency.

<u>The variable copying/improvisation ratio.</u> These basic effects of intrinsic syringeal control seem to be common to the song of all acoustically isolated songbirds. In themselves, they tend to impose uniformity and stereotypy on the vocalizations of such an isolate. In fact, however, song structure can be quite complex and variable

in acoustic isolates of some species. In the absence of auditory models, such syllabic and phrasal complexity - when compared with the song of early-deafened conspecifics - can be regarded as the result of what Marler (1967, 1975a) has referred to as "improvisation". Comparison of acoustic isolate song and the song of wild conspecifics, therefore, reveals the ratio of improvised and copied elements in a species' song. Interspecifically, this ratio is highly variable.

Songs of acoustically isolated white-crowned sparrows (*Zonotrichia leucophrys*), which consist principally of "the sustained thin whistles [i.e., quasi-sinusoidal acoustic signals] which characterize members of the genus *Zonotrichia*" (Marler, 1967), when compared with the wide-bandwidth vocalizations of early-deafened conspecifics, represent extreme examples of an innate vocal pattern's modification by learnt control of syringeal vibration. (See Konishi, 1965b, for comparative spectrograms of early-deafened, acoustic isolate, and wild white-crowned sparrow song.)

The acoustic isolates' vocal signals are both "sustained" (long duration) and "thin" (narrow bandwidth) as a result of syringeal control. They are also extremely simple in terms of syllable and song structure - i.e., improvisation is at a minimum.

Marler (1967) commented on certain genus-specific features of these isolate songs (in the absence of species-specific features). (For example, the songs of white-crowned sparrow isolates rather closely resemble the wild song of the white-throated sparrow, *Zonotrichia albicollis*: see Brooks and Falls, 1975.) Thus a distinctive <u>generic</u> feature of *Zonotrichia* song can be learnt by self-monitoring alone. The development of the more complex song features <u>specific</u> to white-crowned sparrows, however, depends on the copying of conspecific models. (See Marler, 1970a, for further spectrographic data on *Zonotrichia* vocal ontogeny.)

Thorpe's (1958a, 1958b) careful study of vocal ontogeny in the chaffinch (*Fringilla coelebs*) provides the original and classic spectrographic analysis of vocalization by acoustically isolated songbirds. He reported (Thorpe, 1958b) that the song of each of his "Kaspar Hauser" chaffinches was "of about the right length and of about the right number of notes", and that these notes were "of fairly normal tonality in themselves, though of abnormally uniform tonal quality". Thorpe (1954) distinguished three consecutive sections in normal chaffinch song; the songs of his acoustic isolates consisted simply of "abnormally uniform" syllables similar to those in "section 2" of wild song. (See also Poulsen, 1951.)

Thus, in the chaffinch, acoustic isolation throughout vocal ontogeny produces results comparable to those in the white-crowned sparrow: gross temporal features of the song are preserved and source vibrations are quasi-sinusoidal, but syllable and phrase structure show an abnormal simplicity and stereotypy. The features of overall duration and rhythm are related to respiratory function (Calder, 1970; Brackenbury, 1980b), and that of acoustic waveform periodicity to auditory self-monitoring. These regulatory mechanisms function normally in acoustic isolates. Species-specific variations in syllable and phrase structure in birds such as the chaffinch and white-crowned sparrow, however, are learnt by copying from conspecifics.

At the other extreme of songbird vocal ontogeny from the white-crowned sparrow is the song sparrow (*Zonotrichia melodia*), in which the songs of acoustic isolates can sound virtually indistinguishable from those of wild birds. (Mulligan, 1966, reported that song developed normally in acoustically isolated song sparrows, but subsequent detailed acoustic analysis by Kroodsma, 1977, showed that there are significant abnormalities in the songs of these isolates.)

The impression of normality in the songs of acoustically isolated song sparrows is the result of a high degree of improvisation during their vocal development in the wild. Thus "each individual song sparrow possesses a large repertoire of dissimilar song patterns . . . [which] are not shared with other members of the population" (Harris and Lemon, 1972). Such improvisation occurs, too, in the vocal ontogeny of acoustically isolated song sparrows, so that the simplicity and stereotypy typical of chaffinch and white-crowned sparrow isolate song does not appear in song sparrow isolates. (Harris and Lemon, *ibid.*, reported patterns of syllable sharing in wild song sparrows, and the consequent formation of dialects. Kroodsma, 1977, showed that some shared patterns of wild song in this species are abnormal in acoustic isolates. These studies indicate that some copying does normally occur in song sparrow vocal development.)

These, then, represent the extremes of the copying/improvisation variable in songbird vocal ontogeny: extreme dependence on copying in the white-crowned sparrow, and extreme reliance on improvisation in the song sparrow.

At this point in the discussion it would be in keeping with the tradition of the past 20 years to invoke the assistance of species-specific auditory "templates" (Konishi, 1965b; Marler, 1970a, 1975a, 1976) in accounting for these interspecific differences in songbird vocal ontogeny. These species-specific "templates" are conceived of as complex neural filters blocking the central audio-vocal pathways and selectively passing only those auditory response patterns which correspond to appropriate acoustic models for vocal learning in the species concerned.

In its selective response to auditory feedback, such a template is thought to be able to guide the vocal output of a young bird until an acceptable match with the

template's specifications is achieved. In species such as the song sparrow, the template is supposed to be innately adequate to perform this guidance function - given the input of auditory feedback - without auditory experience of the model. In species such as the white-crowned sparrow, on the other hand, the innate template has to be "improved" (Marler, 1976) by auditory experience of the model before it can guide the development of normal song.

In a subsequent section of this chapter ("The nature of the auditory template") some of the inadequacies, redundancies and self-contradictions of the species-specific template concept will be criticized in the context of interspecific vocal learning. It need only be pointed out here that, given a predisposition to improvise on an inherited vocal substratum within the limits imposed by motor constraints, it seems unparsimonious to introduce the shadowy agent of a species-specific auditory template into an explanation of vocal ontogeny in acoustically isolated song sparrows.

Intermediate between the two extremes of the white-crowned sparrow and the song sparrow are songbirds such as the red-winged blackbird (*Agelaius phoeniceus*) and the European blackbird (*Turdus merula*). Comparative acoustic analyses (Marler *et al.*, 1972; Marler, 1975a) of songs from early-deafened and acoustically isolated red-winged blackbirds show the expected narrower spectral bandwidth and more periodic waveform of the isolate songs. Unlike the white-crowned sparrow, however, there is a tendency in these isolates to combine diverse elements into compound syllables and phrases. The result of these improvised combinations is a song which, although abnormal, closely resembles wild song (Marler, *ibid.*) in many respects. The improvisation is more limited in extent than that in isolated song sparrows, and a greater degree of intraspecific copying is thus required for the development of normal song patterns in red-winged blackbirds.

Turdus merula has a similar pattern of vocal development. Messmer and Messmer (1956) reported normal development in acoustically isolated blackbirds although, paradoxically, showing that in wild song some patterns are copied from conspecifics. Subsequently, acoustic analysis of blackbird isolate song by Thielcke-Polz and Thielcke (1960) demonstrated that, as in red-winged blackbirds, the organization of this song was abnormal. Comparative spectrograms published by Thielcke (1970) show that in acoustically isolated blackbirds the narrow-band, sustained-frequency elements are too long, and the frequency excursion of trills is too small - i.e., the typical isolate characteristics of narrow bandwidth and protracted duration are abnormally pronounced in isolated blackbirds. There is in blackbird isolate song, however, a degree of improvised recombination of elements comparable to that in red-winged blackbird isolates. (Hall-Craggs, 1962, has provided detailed evidence of elaborate improvisation during wild blackbird vocal ontogeny.)

Two other species which have been shown by acoustic isolation experiments to have a ratio of improvisation to copying in their vocal ontogenies comparable to that in blackbirds are the cardinal, *Richmondena cardinalis* (Dittus and Lemon, 1969; Lemon and Scott, 1966) and the indigo bunting, *Passerina cyanea* (Rice and Thompson, 1968). The cardinal, for example, when raised in acoustic isolation, develops song syllables with a "flute-like" (Dittus and Lemon, 1969) quality - i.e., having a quasi-sinusoidal waveform - and lacking the structural complexity of wild cardinal syllables.

These examples are sufficient to give an idea of the interspecific variability of the improvisation/copying ratio in songbird vocal learning. It remains to indicate the complex interrelationship of these two behaviour patterns in the vocal ontogeny of any particular species.

The tendency to improvise in acoustic isolates of some

songbird species can actually lead a bird further from the natural song pattern than does the total deprivation of auditory experience by early deafening. Marler (1967) reported that each song of acoustically isolated Oregon juncos (*Junco oreganus*) typically consists of more than one syllable type, whereas most songs of wild (and early-deafened) Oregon juncos are simple repetitions of a single syllable type. (See above, under "Audition-based control of syringeal muscles". Marler, 1975a, provides comparative spectrograms of early-deafened, acoustic isolate, and wild Oregon junco song.)

On the other hand, improvisation in acoustically isolated Mexican juncos (*J. phaeonotus*) results in a closer approximation to the complex songs of wild Mexican juncos than are the single-syllable songs of early-deafened birds (Marler, 1967).

In the ontogeny of Mexican junco song, therefore, hearing the complex songs of conspecifics reinforces the tendency to improvise apparent in acoustic isolates; in the ontogeny of Oregon junco song, hearing the simple song of conspecifics inhibits the tendency to improvise. Thus the juncos provide perhaps the most striking example of the complex interaction of copying and improvisation in songbird vocal ontogeny.

At this point it seems evident that, as far as songbirds are concerned, potential "talkers" should be sought among species in which (1) the learnt component of the song pattern predominates over the innate component and (2), within that learnt component, copying predominates over improvisation.

This is not, however, the end of the process of isolating "talking" potential. The following sections will discuss the relative merits of song and calls as the species-specific vocalizations upon which "speech" may be moulded, and then establish some definitions and distinctions which are essential for a discussion of further variable factors in avian "speech" potential.

DISTINCTIONS AND DEFINITIONS

Calls and song

At this point it is necessary to consider the distinction between song and calls. Although the difference is sometimes not at all clear-cut either acoustically (Thorpe, 1961) or behaviourally (Thorpe, *ibid.*; Andrew, 1969), the features of song which are generally accepted as distinguishing it from calls are its acoustic elaboration of often call-like components (Thorpe, 1958a; Thielcke, 1966), its direct interaction with the neuro-endocrinological state of the bird (Nottebohm, 1980), and its apparent primary function as a generalized male sexual/territorial advertisement (Armstrong, 1963; Catchpole, 1979).

Calls, on the other hand, while acoustically simpler than song, are more varied in terms of motivation: they are more dependent on specific behavioural contexts than on hormonal levels.

Early-deafened chickens (Konishi, 1963) and doves (Nottebohm and Nottebohm, 1971) developed normal calls - confirming that in these species vocal communication sounds are innate rather than learnt.

Calls are apparently innate, too, in some songbird species: for example, Dittus and Lemon (1970) observed normal development of calls - as distinct from song - in early-deafened cardinals. Thorpe (1958a) was convinced that some chaffinch calls are inherited. (See "A species-specific acoustic factor", below.) Nottebohm (1967), however, found that early-deafened chaffinches developed *some* abnormal calls, indicating an element of call-learning in chaffinch vocal ontogeny.

Thorpe (1958a) observed that the song of the chaffinch includes some acoustically simple syllables which are, in spectrographic analyses, identical with certain chaffinch calls. Thielcke (1966) showed that the songs of the two European treecreeper species (*Certhia familiaris* and *C. brachydactyla*) consist of stereotyped combinations and repetitions of species-specific calls. Such examples indicate the likelihood of a more general derivation of oscine song from calls, as suggested by Thielcke (*ibid.*) and Catchpole (1979). (See below, this chapter, under "A species-specific auditory factor".)

An evolutionary development of (learnt) song from (innate) calls accords well with the generally-accepted "advanced" status of songbirds in the taxonomic hierarchy. It is supported, too, by experimental studies which - considering the Passeres as a whole - indicate that all song is subject (to a variable degree) to the "advanced" ontogenetic strategy of vocal learning (Kroodsma, 1981), whereas some songbird calls develop entirely by the more "primitive" means of a genetic program. The evolution of call learning in songbirds, then, is probably contemporary with (or later than) the evolution of song.

Song - being exclusive to songbirds - is characteristically controlled by the "song muscles" (intrinsic syringeal muscles). An evolving extension of auditory influence over the ontogenesis of songbird calls implies a corresponding extension of influence by the "song muscles" over call production.

As discussed above, song is a highly ritualized (Kroodsma, 1979; Thielcke, 1966) form of vocal behaviour, whereas calls are more flexibly responsive to the bird's changing environment. Thus in a repertoire of learnt calls there is a combination of flexible neuro-muscular activity patterns and flexible behaviour patterns. This combination allows - in species possessing such a learnt-call repertoire - a versatility of vocal behaviour which

is often regarded both as evidence of advanced evolutionary derivation and as an indication of relatively high general intelligence.

For example, many taxonomists (as discussed by Storer, 1960) believe the Corvidae to be the most highly evolved passerine family - usually on the ground of a complexity of behaviour (including vocal behaviour) which indicates high intelligence (Storer, *ibid.*).

The distinctions and relationships between calls and song which have just been discussed are of great significance to the study of "talking" birds.

A short list of those oscines which are most willing and able to "talk" reveals that most of them - principally members of the families Sturnidae and Corvidae - are virtually songless, with vocal repertoires consisting entirely of calls.

For the most celebrated "talker" among these - the hill mynah (*Gracula religiosa*) - there is evidence (Bertram, 1970) that the large and varied call repertoire is acquired entirely by learning. (Bertram, *ibid.*, in defining mynah vocalizations as calls, drew attention to their relative acoustic brevity and syllabic simplicity, their use by both sexes throughout the year, and their essential independence from hormonal levels.)

Thus *Gracula religiosa* is one of those songbird species in which a variety of (learnt) calls has evolved instead of - or has perhaps replaced - typical oscine song. This means that vocal behaviour in *Gracula* involves active neuro-muscular control over the production of signals which are responsive more to the social environment than to the neuro-endocrinological state of the individual. Such a combination of vocal and social flexibility provides optimum potential for vocal interaction by a bird with a human social partner.

The Corvidae are a family of species with similarly call-oriented vocal repertoires. The survey by Kroodsma and Baylis (1982) lists eight species of Corvidae -

including the jackdaw (*Corvus monedula*), the raven (*C. corax*) and the crow (*C. corone*) - which are known to learn at least some of their calls. Lanyon (1960) quoted a report that ravens raised in isolation from species-specific sounds "failed to develop their entire vocal repertoire". One manifestation of vocal learning in the Corvidae is, of course, interspecies vocal copying, including that of the human voice.

In considering the songless, call-learning, "talking" mynah and crows as among the most highly evolved and intelligent of the Passeres, we are faced with the analogy of the similarly songless, call-learning (Kroodsma and Baylis, *ibid.*) and intelligent "talking" parrots.

The "talking" behaviour of a parrot such as an African grey (*Psittacus erithacus*) displays a level of flexibility similar to that of a mynah. This similarity can be traced to their vocalizations in the wild: for instance, in both mynahs (Bertram, 1970) and some parrots (Nottebohm, 1970; Thorpe and North, 1965) mutual calling - or duetting - of mated pairs is an important aspect of vocal communication. These similar behaviours, with their similarly learning-based vocal ontogenies and comparable neuro-motor systems, have evolved along separate phyletic lines (Nottebohm, 1972).

Turning, now, to singing birds, the canary (*Serinus canaria*) is able to incorporate speech-like sounds into its song pattern (Greenewalt, 1968) and can even achieve considerable virtuosity as a "talker" (Maynard, 1928). There are significant differences, however, between the "speech" of singing birds such as the canary and that of non-singing oscines such as the mynah. These differences are directly related to the distinction between calls and song discussed above, and are analogous to differences - to be examined in Chapter 5 - between the "speech" of the budgerigar (*Melopsittacus undulatus*, whose warble is a kind of psittacine song) and that of the larger parrots.

Thus the "speech" of a mynah (or grey parrot), which consists principally of short, emphatic words and phrases produced in an appropriate behavioural context or during vocal exchanges with the human trainer, represents a social interaction. Budgerigar (or canary') "speech", by way of contrast, tends to be more of a performance, during which the bird, once stimulated, runs through its entire repertoire - often in a stereotyped order - regardless of context. (Social factors are of great importance in avian "speech", and will be considered in detail later in this chapter.)

Chapter 5 of this work will demonstrate that, acoustically, avian "speech" is a modification (often minimal) of the bird's species-specific vocal communication sounds. Another parameter can now be specified in such an analysis of avian "speech" ontogeny: whether the species-specific sounds thus modified can be considered as calls or song. This dimension adds depth to our understanding of the motivation for - and the performance of - this remarkable behaviour. (And, on the level of subjective experience, it adds an unprecedented immediacy and poignancy to these utterances which are all too often perceived as amusing but inconsequential curiosities.)

"Imitation" and "mimicry"

It is a long-established practice in the literature on bird sounds to distinguish between "imitation" and "mimicry" in vocal copying behaviour. The traditional distinction has been that between the copying of species-specific sounds - "imitation" - and the copying of non-species-specific sounds - "mimicry" (Thorpe, 1967; Davis, 1973).

Interest in avian vocal mimicry in the first half of this century centred on the evidence it provided of vocal learning in birds. (See Baylis, 1982, for an historical

review.) Following the experimental work of Poulsen (1951), Marler (1956), and Thorpe (1958a), however, attention shifted to intraspecific imitation in the study of avian vocal learning, and interspecific mimicry received attention mainly from ethologists who were more interested in its possible functions than in the ontogenetic mechanisms involved. Dobkin (1979) developed an elaborate scheme of distinctions and classifications of avian vocal copying behaviour based on functional considerations, and involving the introduction of two new vocal-copying categories as well as the familiar "imitation" and "mimicry".

Baylis (1982) contended that this functional approach had led Dobkin into self-contradictions as well as a cumbersome apparatus of classification. Baylis advocated a return to traditional descriptive definitions of vocal copying behaviour. "I define vocal mimicry as the resemblance of one or more vocalizations of an individual bird of one species either to the vocalizations typical of individuals of another species or to some environmental sound" (Baylis, *ibid.*).

Now, while it is granted that Dobkin's functional definitions can be misleading, it must be pointed out that definitions such as that of Baylis quoted immediately above, which describe a behaviour pattern in terms of some external model/stimulus source, can be equally misleading. The essential distinguishing feature of vocal imitation, for example, is not that the source of the acoustic model is in fact a conspecific individual, but that the acoustic model - regardless of its source - is perceived and processed according to an audio-vocal behaviour pattern appropriate for species-specific sounds. This is an important distinction, for it will be argued below that under some conditions allospecific vocalizations can be perceived and processed in this way, in which case copying behaviour should be defined as imitation rather than mimicry.

(The term "behaviour pattern", as employed here, refers to the objects of study of both ethology and neurophysiology - i.e. patterns of movement by an organism, as well as patterns of change within the neural networks which control those movements - see Fentress, 1976. In the present context, this includes both the patterned sequence of actions associated with sound reception, recognition, learning and production, and corresponding patterned changes of state within the underlying neural control networks. The entire complex is referred to as an "audio-vocal behaviour pattern".)

Thus shifting the perspective away from the external (and, ultimately, incidental) source of the acoustic model, and onto its perception and processing by the copier, would truly characterize vocal copying as an integrated audio-vocal behaviour pattern. In the light of these observations, vocal imitation can now be defined as the copying of sounds which are perceived and processed by the copier according to an audio-vocal behaviour pattern appropriate for species-specific vocalizations; complementarily, vocal mimicry is the copying of sounds which are perceived and processed by the copier according to an audio-vocal behaviour pattern appropriate for non-species-specific sounds.

These definitions, as distinct from the descriptive definitions of Baylis and the functional definitions of Dobkin, are behavioural definitions: they focus on the behaviour of the individual rather than on any external model source or supposed function of that behaviour.

Enough is now known about avian vocal learning to make feasible the application of such definitions.

At the level of neural control, special processing of species-specific vocal signals (or signals perceived as species-specific) will be discussed in Chapter 4.

At the level of the vocal-learning behaviour pattern of concern to the ethologist, the development of Bengalese finch song in zebra finches raised exclusively

by Bengalese finch foster-parents follows the complex pattern - seasonal synchrony, receptive period, dormant aural memory, auditory feedback, and the phases of subsong, plastic song and full song - typical of species-specific song development in zebra finches (Immelmann, 1969; Clayton, 1987). (For reviews of songbird vocal ontogeny see Kroodsma, 1978, 1981; Lanyon, 1960; Marler and Mundinger, 1971; Nottebohm, 1970.) Furthermore, the allospecific song patterns are a replacement of - rather than merely an addition to - species-specific song patterns (Immelmann, 1969; Clayton, 1987).

According to Baylis's (1982) definitions, this interspecific vocal-learning behaviour in Estrildidae would be labelled as "mimicry". The unsatisfactory nature of Baylis's distinction is manifest, however, when the behaviour pattern as a whole - as outlined in the previous paragraph - is considered.

According to the behavioural definitions advocated here, on the other hand, the vocal development of Immelmann's zebra finches would be labelled as "imitation". This categorization reflects the true nature of the audio-vocal behaviour pattern, which is identical with that for species-specific vocal learning.

An interpretation of subsong. The implications of this behavioural approach to vocal copying phenomena are well illustrated by the example of the chaffinch. Thorpe (1955, 1958) noted that a hand-reared chaffinch exposed to canary song in its first autumn included a phrase of canary song in its subsong the following spring. "Nevertheless, when this bird came eventually to produce its true song, it sedulously avoided contaminating it with the canary notes previously learnt, and its true song was of the simple Kaspar Hauser type" (Thorpe, 1958).

Slater (1983) cross-fostered a chaffinch chick with canary foster parents. This chaffinch developed full song

comprising canary song elements and having patterns of organisation similar to those of canary song.

The descriptive approach of Baylis (1982) would draw no categorical distinction between these two examples of canary-copying behaviour in chaffinches: both would be labelled as "mimicry"[5]. In behavioural terms, however, the two examples do exhibit a categorical distinction. In the former instance, the copying behaviour did not follow the full, complex course of vocal ontogeny, but was abandoned after the subsong phase: it must, therefore, be labelled as "mimicry". In the latter instance, the copying behaviour did follow the full, normal course of vocal ontogeny, and must be labelled as "imitation".

This comparison highlights the following behavioural distinction: while vocal mimicry involves primarily a non-discriminative facility in vocal copying, vocal imitation is characterized by an integration of this non-discriminative facility into a species-specific pattern of vocal development constraints.

In the light of this distinction, the subsong phase of vocal development can be regarded as - among other things - a period in which mimetic behaviour is gradually channelled into audio-vocal patterns appropriate for imitation. (Wildenthal, 1965, and Robinson, 1975, pointed out that most inter-specific copying in wild songbirds occurs in subsong. Robinson, *ibid.*, proposed that such subsong mimicry "is likely to be present in all passerines capable of learning and reproducing sounds".)

In these terms, then, the behavioural definitions of imitation and mimicry can provide the basis for a general approach to the phenomenon of subsong.

An evolutionary hypothesis. Marshall (1950), among others, has remarked on the particularly high quality and intensity of mimicry in Australian passerines - especially in that most celebrated family of true avian mimics, the Menuridae (lyrebirds).

The position of the Menuridae and the Altrichornithidae (scrub-birds) in the taxonomic hierarchy is still a matter of debate (see Lill, 1985, and the references quoted therein), but traditionally they have comprised a passerine sub-order (the Menurae) lower than the Passeres (Storer, 1960, 1971) - a classification based on their relatively simple syringeal myology (Newton, 1893).

Recent protein-analysis data have led to the proposal of an Australasian origin for the Eurasian songbirds (Ford, 1987).

These considerations prompt the hypothesis of an evolutionary development from avian species whose vocal ontogeny is dominated by mimicry to species in which mimicry is subordinated to imitation. (Correspondingly, most hypotheses on the origin of hominid language involve some form of mimetic copying: see, for example, the review by Hewes, 1977).

The fact that this subordination of mimicry to imitation is observable during the ontogenetic phase of subsong could contribute to the continuing debate on parallels between ontogeny and phylogeny (Gould, 1977).

Avian "speech" is vocal imitation

Now, why aren't compulsive mimics such as the Menuridae and the Mimidae prized as pets and trained to "talk"?

The foregoing discussion provides the conceptual framework for a simple answer to this question. That answer - Avian "speech" is an example of vocal *imitation* (as defined above), not vocal *mimicry* - is a proposition central to the argument of this thesis.

The first point to be made in this connexion is that all "talking" birds are representatives of species in which species-specific imitation is the principal mode of vocal learning. Writers on the subject of "talking" birds (e.g. Marler and Mundinger, 1971) have typically been puzzled by the lack of evidence for lyrebird-like vocal mimicry in wild parrots and mynahs[6]. This ceases to be a puzzle if avian "speech" is classified as imitation rather than mimicry.

(Every parrot-owner is aware, of course, that these birds do mimic in captivity: mimicry of anything from a creaking door to a dripping tap can occur. In the light of the foregoing discussion it may be argued that such mimicry represents the phylogenetically more primitive substratum of indiscriminate vocal copying upon which the patterned behaviour of speech imitation is based. An essential difference between speech imitation and the mimicry of environmental sounds is the involvement of social factors in avian "speech". These are dealt with below.)

Parallels between avian "talking" and both singing and calling behaviour were pointed out above. This is a strong argument for the categorization of "speech"-learning behaviour along with species-specific vocal imitation.

Another strong argument is provided by some remarkable acoustic similarities between avian "speech" and species-specific sounds. Chapter 5 will demonstrate how a budgerigar modifies species-specific sounds (often only minimally) in order to produce speech-like sounds. These similarities are an acoustic correlate of the patterned neural processes underlying all true vocal imitation - species-specific or interspecific - as defined above.

Chapter 5 will discuss in some detail the ontogeny of "speech" behaviour in the budgerigar, and evidence will be presented that it follows a parallel course to species-specific vocal learning in these birds. The connexion of "speech" learning with species-specific duetting behaviour in the larger parrots will be dealt with in the present chapter, below.

From now on, therefore, vocal-learning behaviour which produces the acoustic phenomena of avian "speech" will be referred to as "speech imitation".

To review the argument so far: the search for birds with "talking" potential focused to begin with on those whose vocal ontogenies involve audition-based learning. Among these, the focus narrowed to species in which the learnt component of vocal patterns predominates over the innate component. Then once again the focus was narrowed to those birds in which the learnt component shows greater evidence of copying than of improvisation. Finally, a behavioural distinction was drawn between two kinds of copying - mimicry and imitation - and it was argued that "talking" behaviour is restricted to birds whose vocal ontogenies are based primarily on imitation. (Among such birds, it was suggested that the most socially versatile "talkers" are those which base their "speech" patterns on calls rather than on song.)

The following sections will examine acoustic and social factors which determine the extent to which species-specific vocal imitation behaviour can be channelled in the direction of avian "speech".

AUDITORY AND SOCIAL FACTORS IN VOCAL IMITATION

The nature of the auditory template

The behavioural definitions of "imitation" and "mimicry" advocated here, and the consequent categorization of some inter-species vocal copying as "imitation", have far-reaching implications for theories of innate auditory mechanisms in avian vocal learning.

Marler (1970b, 1975a, 1976) conceived of the "innate auditory template" for song-learning in birds (referred to above, under "The role of improvisation in birdsong") as a neural filter which passes only species-specific patterns from the periphery to audio-vocal centres in the forebrain. He regarded such a template as being analogous to the "innate release mechanism" of Lorenz (1950) - i.e., as a mechanism which not only recognizes the appropriate stimulus, but is able to guide the organism's pattern of response to that stimulus.

The template theory, as developed principally by Marler, involves elements of confusion and redundancy which the behavioural distinctions and definitions of vocal copying can help to clarify.

Ambiguity of the "template" hypothesis. An underlying ambiguity in Marler's treatment of the subject is that although the auditory template is, according to his definition (Marler, 1970b), uniquely sensitive to species-specific song patterns, it nevertheless serves "to focus the attention of the male on patterns of external stimulation that have assumed a special valence in the course of phylogenetic history [my emphasis] of species" (Marler, 1976).

Now, the "phylogenetic histories" of closely related

songbird species such as the white-crowned sparrow and the song sparrow are, acccording to palaeontological evidence, probably identical until at least the end of the Miocene epoch (Olson, 1985). Furthermore, "it appears that all modern [avian] species have evolved within the last one million years" (Macphail, 1982). Thus, if the auditory template were primarily a neural representation of sound patterns preserved throughout the ramifications of phylogeny, it would not be expected to show such exclusive responsiveness to acoustic patterns which have evolved within the last one million years; indeed, it might be expected to facilitate rather than inhibit interspecies vocal copying - particularly among such close relatives as the *Zonotrichia* sparrows.

This problem was recognized by Marler (1975a), as "one further point we have still to explain", but remains unresolved in the literature. For example, Marler (1976) regarded the song of white-crowned sparrow isolates as being an acoustic correlate of the birds' "unimproved template". Although "an ornithologist would probably identify these isolated songs as coming from a member of the genus [*Zonotrichia*], but would be puzzled as to the species" (Marler, 1967), this "unimproved template" is nevertheless "sufficient to serve as a kind of filter that will pass white-crowned sparrow song and reject songs of other species" (Marler, 1975a).

The behavioural distinctions and definitions of vocal copying can help to clarify this issue.

Invariably, the template theory has been associated only with species-specific vocal learning. This was possible as long as all interspecific vocal copying was defined as something essentially distinct from intraspecific copying. According to the behavioural view of vocal learning developed here, however, some instances of <u>interspecific</u> vocal copying represent essentially the same complex behaviour pattern as intraspecific copying, and should be regarded as <u>imitation</u>. In the light of this

view, innate auditory mechanisms used for species-specific learning could also be used for inter-species learning.

Chapter 5, below, will seek to demonstrate that the existence of common acoustic cues in the vocal communication sounds of such diverse vertebrate groups as birds and man does indeed facilitate both the copying of human speech sounds by birds, and the human perception of these avian copies as speech-like sounds.

There is considerable experimental evidence, reviewed in Chapter 4, for the shared existence in several avian and mammalian species (including *Homo sapiens*) of perceptual mechanisms selectively responsive to acoustic cues which the vocal communication sounds of those species have in common.

The phyletic stability of these features probably reflects, in part, the relative stability, throughout the recent course of evolution, of environmental factors which have been shown to determine certain parameters of acoustic communication sounds. (See, e.g., Morton, 1975).

If, then, the acoustic feature detectors which birds and man share as the result of a partly-shared phyletic inheritance (and a shared environment/transmission-channel) form the bases of templates for audition-dependent vocal learning, the function of these templates - at least at this basic level - could well be <u>interspecific</u> rather than <u>species-specific</u>.

It has been argued above that avian "speech" is an instance of vocal <u>imitation</u>, meaning that avian "speech" development follows the same audio-neuro-motor patterns of vocal ontogeny as the development of species-specific sounds. The sharing of acoustic cues means that mechanisms comprising an avian song-learning template could function, too, in the learning of human speech sounds by birds.

It must not be forgotten, of course, that the idea of a "template" is inseparable from that of a "filter". The

argument here is that, at this most basic level, the filtering criteria are not related to species-specificity, but rather to communication sounds. The concept of a "bio-communication cue" will be developed in Chapter 4.

(The function of communication itself imposes some common constraints on bio-acoustic signals. For example, information transfer necessarily involves patterning, and Brackenbury, 1978a, has discussed the widespread use of a "pulse code" - the patterned repetition of identical transient units - in bio-communication signals.)

Interspecific - as well as intraspecific - vocal imitation can in this way be understood as a natural consequence of a "bio-communication" filter; it need not be regarded (as in much of the literature on avian song learning) as a rather puzzling anomaly. On the basis of this inherited sensitivity to interspecific communication cues, additional acoustic factors (see below) could direct vocal learning towards species-specific models.

Redundancy of "template" hypothesis. Marler (1970b) used the concept of an innate auditory template in order to explain the selective learning response of white-crowned sparrows to tutor tapes: i.e., their selection of species-specific rather than non-species-specific tape-recorded song as a model for vocal learning.

The template concept remains, however, in Marler's writings and in those of his followers, essentially a description of the model-selection phenomenon rather than an explanation of it. In other words, his template hypothesis says little more than that birds selectively learn species-specific song because they are predisposed (according to certain unspecified inherited constraints) to do so. His hypothesis itself is thus, in a sense, redundant.

Marler recognized that "the underlying mechanisms may be several and complex, and may also differ in their

nature, number, and mode of coupling from species to species" (Marler, 1976). Only when these underlying mechanisms are specified, can the above-mentioned redundancy of the template concept be replaced by something of real explanatory value. The present discussion has proposed a basis of feature detectors sensitive to acoustic cues reflecting the evolution of vocal communication in vertebrates. It may now be possible to identify some higher-level auditory factors which could contribute species-specificity to the template filter. One such factor is discussed in the following section.

A species-specific auditory factor

Thorpe (1958a) reported that call notes are particularly evident in the song of acoustically isolated chaffinches. He regarded this as indicating that "vocal imitation is probably a later development from inherited patterns". Thielcke (1966) pursued this evolutionary approach with convincing spectrographic evidence that the songs of *Certhia* treecreepers evolved from species-specific calls. He regarded song as a "ritualized" development of calls.

Now, if (learnt) songs have evolved from (innate) calls, it follows that a young songbird would be predisposed to learn songs containing acoustic elements which correspond in some way to inherited species-specific call patterns. In the chaffinch, calls are "certainly inherited" (Thorpe, 1958a). (Nottebohm, 1967, however, reported abnormal development of some calls in acoustically isolated chaffinches.) Dittus and Lemon (1970) found that calls developed normally in early-deafened cardinals (*Richmondena cardinalis*), indicating that calls are innate in this songbird species.

Thus, recognition of innate call patterns could be an effective strategy for the selection of appropriate vocal

models, and could explain the selective-learning response to tutor tapes in species such as the chaffinch (Thorpe, 1958a) and the white-crowned sparrow (Marler, 1970b).

Social factors

To recapitulate briefly on avian vocal-learning model appropriation: the argument thus far is that, at the most basic level, a filter of acoustic feature detectors passes sounds containing "bio-communication cues" - i.e., cues which reflect common features in the evolution of vocal communication sounds and are thus interspecific. The sounds passed need not, therefore, be species-specific. At a higher level, species-specific auditory factors - such as the recognition of innate call patterns in the model - can direct song-learning towards a species-specific model.

It follows that, in the absence of such higher-level auditory constraints, a bird hearing only allospecific communication sounds could appropriate them as a model. It has been pointed out above, for example, that some songbirds learn calls as well as song. These birds thus lack the model-appropriation constraint - discussed in the previous section - of inherited call patterns. In the context of avian "speech", it is significant that such songbirds - e.g., mynahs and crows - make excellent "talkers" (when exposed to human rather than avian communication sounds).

Male zebra finches - which readily learn non-species-specific song in cross-fostering experiments (see above) - when raised by females will accept and learn female call notes as the constituents of their song pattern (Eales, 1987b; Price, 1979). "The female-tutoring experiment shows that males can learn calls and incorporate them into the song-unit" (Price, *ibid.*).

The key concept here is that of communication. In the discussion of definitions and distinctions above, it was

pointed out that the learning of non-species-specific communication sounds (as in cross-fostered zebra finches) can follow the same complex ontogenetic behaviour pattern as species-specific vocal imitation. It was argued, subsequently, that this pattern includes the special processing of interspecific "bio-communication cues". Mimicry, on the other hand, does not follow the ontogenetic pattern typical of imitation (as pointed out above in reference to subsong mimicry).

Interactive communication is a social function, and the concept of "bio-communication cues" as the factor which creates the potential for inter-species vocal imitation is complemented by the concept of a "social bond" which helps to realize that potential. (Mimicry, on the other hand, is not associated with a social bond: no-one would claim, for example, that there is such a bond between a lyrebird and any of the individuals it mimics.)

Evidence has been accumulating, over the past 30 years, for the importance in avian vocal learning of this social factor - usually referred to in the literature as an "emotional" or "social bond".

Nicolai (1959) and Immelmann (1969) attributed interspecific song learning in cross-fostered bullfinches (with canary foster parents) and zebra finches (with Bengalese finch foster parents) to the formation of such an emotional bond. Slater (1983) considered that his discovery of interspecific song-learning potential in the chaffinch was "in line with the increasing evidence that one of the constraints on the song learning of young birds of many species is not so much a song template as a constraint making it more likely that learning will occur from individuals with which the bird has a social relationship". Payne *et al.* (1984), in reporting an example of inter-species vocal copying in the chestnut-sided warbler (*Dendroica pensylvanica*) considered their observations "consistent with the suggestion that song learning in birds may normally depend upon social

interaction".

Pepperberg (1985a, 1985b) advocated a comparative approach to avian vocal learning and human social modelling (Bandura 1969, 1971, 1977) based on the efficacy in both systems of live, interacting social tutors.

This relationship - at least in conspecific vocal learning - need not necessarily be one of parent-offspring. Baptista and Petrinovich (1984) argued for the importance of aggressive social interactions among males in song learning. Such aggressive interactions include the dominance-subdominance pattern of countersinging "duels" in marsh wrens, *Cistothorus palustris* (Kroodsma, 1979), and sedge wrens, *C. platensis* (Kroodsma and Verner, 1978). Competitive, interactive song matching is a conspicuous vocal behaviour in the indigo bunting, *Passerina cyanea* (Payne, 1981, 1982, 1983), and the saddleback, *Philesturnus crunculatus* (Jenkins, 1977).

Growing appreciation, during the past decade, of the role of social bonding in songbird vocal development has led to an awareness that some earlier reports on "selective" song learning - in, e.g., the white-crowned sparrow (Marler, 1970b; Marler and Tamura, 1964) - which were based on experiments using tape recordings rather than live tutors, have only limited applicability. (See Baptista and Petrinovich, 1984, for a more recent approach, and a critical appraisal of the earlier work.) Baptista and Morton (1981) showed that, in the white-crowned sparrow, auditory factors in vocal model selection can - in the presence of live allospecific tutors - be overriden by social factors.

Nevertheless, song-learning experiments with tape recordings can provide useful comparative data. For example, Marler *et al.* (1972) found that a red-winged blackbird (*Agelaius phoeniceus*) learnt - from tape recordings - species-specific song patterns *and* song patterns of a closely related species, the Baltimore

oriole (*Icterus galbula*). The marked contrast between this behaviour and the selective learning response to tutor tapes of the white-crowned sparrow is discussed by Marler (1975a).

At the opposite extreme from the red-winged blackbird are birds such as the indigo bunting (Payne, 1981) and the cardinal (Lemon and Scott, 1966). These birds will learn species-specific song only from a live, interacting tutor. Thus, unlike the white-crowned sparrow, they will not learn species-specific song from tutor tapes; nor will they learn from live conspecifics with whom they have no interactive relationship.

In songbirds, then, a complex interaction of acoustic and social factors determines vocal model appropriation in individuals. In interspecific model appropriation, social factors can be decisive. Social relationships thus provide an important behavioural context in which interspecies vocal copying must be considered.

Duetting

Mundinger (1970) reported both species-specific and interspecific copying of flight calls among cardueline finches. He observed that call learning could continue throughout a bird's life, and that it occurred both in the context of a pair bond (conspecific or heterospecific), in which both males and females could learn, and in that of a dominance-subdominance relationship between males. After considering the ability of the canary - a close relative of the cardueline finches - to copy human speech, he concluded that, "like the pine siskin's mimicry of a European siskin flight call, avian mimicry of the human voice may be a manifestation of interspecific social integration occurring in unnatural circumstances" (Mundinger, *ibid.*).

Mundinger's speculations on avian "speech" were influenced by the work of Thorpe and North (1965, 1966)

on the duetting vocal behaviour of some tropical birds. These authors believed that "extreme developments of the imitative ability occur where the main function of the song is to provide for social recognition and cohesion rather than for territorial defence"; and, further, that "one of the major functions of the imitative ability of birds is to establish and strengthen the individual pair bond" (Thorpe and North, 1965). Among the birds inhabiting thickly-wooded tropical environments they found species which appeared to rely on mutual vocalization to strengthen social/sexual bonds, and they pointed out that most "talking" birds derive from such tropical habitats. (See also Nottebohm, 1970.)

Thorpe and North (*ibid.*) compared the "purposive" use of speech sounds by pet parrots with the behaviour of birds which engage in duetting behaviour in the wild: e.g., when one of the partners is absent, the other sings the absent one's part of the duet apparently in an attempt to effect his or her return. They suggested that the speech-copying of pet birds could be derived from such behaviour. (See also Gwinner and Kneutgen, 1962; Mowrer, 1950; Todt, 1975a). This idea was corroborated by Nottebohm (1970) who reported duetting behaviour in wild orange-winged Amazon parrots (*Amazona amazonica*) - a well-known "talking" species - and by Farabaugh (1982) who listed eight Panamanian parrot species in which duetting occurs - including *Amazona ochrocephala*, one of the most renowned of New World "talkers".

Bertram (1970) made a careful study of social interaction in the calling behaviour of Indian hill mynahs, and outlined the complex patterns of relationship which locate particular calls within a network of individual repertoires, mated-pair repertoires, and repertoires of neighbouring pairs.

While this complex interdependence of call types and social relationships is highly developed in mynahs, there appears to be no aspect of acoustic patterning which is

species-specific. "Inherited song patterns are probably a primitive feature; in more advanced species, the role of song or calls in behaviour other than territorial defence (e.g. recognition and social stimulation of the mate and other individuals, pair bond development and maintenance, etc.) probably requires adaptability in vocalizing, one aspect of which is vocal imitation in the development of a bird's noises" (Bertram, *ibid.*).

The above discussion has concentrated on yet another variable factor in avian vocal learning: the relative importance of social factors and species-specific acoustic factors in vocal model appropriation.

It may now be possible to complete the model of the hypothetical "ideal talker" which has been taking shape throughout this chapter. In such a bird, learning would play a major role in vocal ontogeny, and would be largely concerned with copying the sound patterns made by other individuals rather than with improvisation. This copying would follow the complex audio-vocal behaviour patterns typical of imitation rather than those associated with mimicry. Within these behaviour patterns, social factors would predominate over acoustic factors in vocal model appropriation.

These specifications would appear to be met by mynahs and parrots.

PSYCHOLOGY OF "TALKING" BIRDS: LITERATURE SURVEY

Social bonding theories of avian "speech"

The psychology of "talking" birds has been studied from two main viewpoints: (1) social (i.e., as involving interaction with human speakers), and (2) comparative (i.e., as involving "homologies" or analogies with the neuro-physiology of human speech). The former category - those studies concerned primarily with social factors - will be reviewed in this section.

Mowrer's (1950) "autistic" theory of avian "speech", was based on his belief that it is rewarding for a bird to hear itself produce sounds it has heard its owner produce in rewarding circumstances. After this "autistic" stage, the bird learns to use the words to "ask for" the particular rewards associated with them. Finally, the bird becomes so attached to the owner that the most rewarding situation for it is simply that person's company. As a result of this, it may "talk" with particular clarity when it fears that the owner is about to leave the room.

Mowrer raised the question: "Who is training whom?", and suggested that "only by . . . letting the bird gain some control over the owner by means of vocalization does it seem possible to bring the bird's powers as a talker to their fullest development".

He believed that when a bird is caged and fed exclusively by hand a regression to infancy is induced, in which state it is susceptible to speech-learning.

Mowrer (1952) restated his "autism" concept in terms of learning theory. Thus "initially neutral sounds, by virtue of their occurrence in temporal contiguity with primary reinforcements, . . . [acquire] secondary reinforcing properties". He expanded on this (Mowrer,

1958) by suggesting that, in "speech" acquisition, the closer a bird comes to matching its acoustic output with its auditory memory the more satisfaction (secondary reinforcement) it will experience. It learns, thus, "through a sort of built-in, self-directed and self-rewarded rehearsal" (Mowrer, *ibid.*).

A comment by Skinner (1964), to the effect that a parrot's verbal behaviour is reinforced simply by hearing itself make sounds similar to those it has heard its owner make, is in accord with Mowrer's "secondary reinforcement".

Although Mowrer was writing from the behaviourist position, he apparently saw the investigation of avian "speech" as a means of relaxing some of the strictures of that position. In commenting on some negative results of Grosslight and Zaynor (1967) from mynah "speech"-training experiments, he (Mowrer, 1980) suggested that the social complexity of the birds' life might override stimulus response in their verbal behaviour. "Perhaps in an organism as social as the mynah bird the conditions for 'learning to talk' must approximate those implied by the concept of identification rather than by the more rudimentary phenomenon of 'imitation'" (Mowrer, *ibid.*). Here Mowrer's concept of "identification" is identical with that of "imitation" in the present study - i.e., the copying of sounds which the copier identifies with and thus perceives as species-specific.

The significance of the emotional context in avian speech-sound learning is highlighted in several anecdotes by Lorenz (1952). Because of its inordinate fear of the chimney-sweep, a blue-fronted Amazon (*Amazona aestiva*) learnt the phrase "The chimney-sweep is coming" - which it heard in the fearful context - after no more than a few exposures to it. A crow which had obviously had its foot caught in a trap learnt the phrase "Got him in the trap" on that one traumatic occasion. Bremond (1963) corroborated the association of such exceptional speech

learning with "an extraordinary emotion". "We had a jay that was able to repeat a bird call that it had only heard on the day that it was captured" (Bremond, *ibid.*).

Experiments on vocal copying by budgerigars led Gramza (1970) to believe that such copying could represent a form of "play", or flexible interaction with the environment as a whole.

An early observation by Lashley (1913) is important in the light of subsequent research. Lashley noted that a parrot's movements during speech-training "frequently suggest the courting activities of other birds". "It seems not improbable that the principal motive for the parrot's reproduction of sounds is to be sought in a perverted form of sex rivalry" (Lashley, *ibid.*).

At this point it is necesary to reintroduce the subject of duetting. Koehler (1972) reported that his parrot could complete a whistled tune, taking it up exactly from the point where Koehler had stopped, and continuing it in the same pitch and tempo in which Koehler had begun. He compared this behaviour of a captive bird with duetting behaviour in the wild. (Early this century Waite, 1903, and Henschel, 1903, reported the formation of duets from such human-taught tunes in pairs - both conspecific and interspecific - of captive birds.)

Todt (1975a) developed a "speech"-training procedure for an African grey parrot based on the rhythmic antiphonal structure, and the interactive social motivation, of duetting behaviour, and which probably included something of the "rivalry" motivation suggested by Lashley (1913). This procedure employed two human trainers: a "cooperative partner", and a "rival" for that partner's attention. The vocal patterns were in "X-Y" form - "X" being the cooperative partner's section of the duet, and "Y" the parrot's/rival's. Todt found that the parrot quickly learnt the human rival's antiphonal ("Y") responses to the cooperative partner's "X" sections,

apparently in an attempt to gain the approval which it saw the cooperative partner bestow upon the rival for completion of the duet pattern.

In this way, the parrot learnt from the rival in order to communicate with the cooperative partner. Furthermore, once the duetting pattern had been established, the parrot would respond only to "X" sections spoken by the cooperative partner in the parrot's presence, and neither to the same phrases spoken by someone else, nor to recordings of the cooperative partner. (Todt referred to the selective mechanism involved as the "partner filter", and suggested that such a filter could explain, for instance, why parrots do not copy other species in the wild.)

Pepperberg (1981, 1983) adapted Todt's model/rival training procedure in her studies of purposive, context-dependent manipulation of anomalous vocal signals (i.e. speech sounds) in an African grey parrot. Her adaptation of Todt's procedure involved a regular reversal of roles by the two human trainers, thus avoiding the "partner filter" effect described by Todt, and demonstrating the adaptability of psittacine social behaviour.

Pepperberg's results, which are of great importance, will be discussed in a subsequent section.

Psychological kinship theories of avian "speech"

In the "M Notebook" of Charles Darwin appears the following enigmatic entry: "It is known that birds learn to sing & do not acquire it instinctively. May not this be connected with their power of acquiring language" (Darwin, 1980).

Darwin is here obviously comparing "learn[ing] to sing" with "acquiring language". It would be natural for the modern reader to concentrate on the concept of "learning" in both these phrases, and thus interpret the entry as simply attributing avian "speech" to the

operation of species-specific behaviour patterns. Gruber (1974) has pointed out, however, that the key words in the comparison are probably "sing" and "language", referring to Darwin's theory - later developed in *The Descent of Man* (1871) - that language originated in the quasi-musical expression of emotions. In that case, the notebook entry is suggesting that birds are able to learn speech sounds, in particular, because of this evolutionary relationship between song and speech: birds learn to sing, and song is incipient speech.

This latter interpretation places Darwin's notebook entry in the context of other nineteenth-century writings on avian "speech" (such as those of Romanes, 1882; Russ, 1884; Thorndike, 1898), which typically concentrate on some aspect of human-avian psychological "kinship" (such as musical expression) rather than on more general questions of vocal learning. For example Romanes (1882), considering reports of parrots "talking" in their sleep, saw these as evidence of a "striking similarity of psychical processes in the operations of the [avian] memory with those which occur in ourselves".

Anecdotal evidence led Romanes to believe that, in "speech"-learning, birds slowly match their acoustic output, during a period of trial-and-error, with their auditory memory of the phrase they are learning. Even after the phrase is mastered, he observed, occasional aberrations in its acoustic-output pattern are recognized by the bird, and the phrase repeated correctly. Recordings made in the course of the present study support these hundred-year-old reports of self-correction by "talking" birds.

Morgan (1891) reacted against the prevalent view of avian "speech" as evidence of a special psychological kinship between birds and man, insisting that birds cannot understand "the form of words" (i.e., <u>syntactic</u> patterns) in their speech imitations. On the other hand, he presented anecdotal evidence of complex <u>acoustic</u>

patterning: for example, in a parrot which jumbled the lines from several poems, but always so that the rhyme scheme was retained. (See also Abe, 1969.) Todt (1975b) presented evidence of acoustic patterning in the recombination of speech-imitation syllables by grey parrots. More recently, Wallschlager (1981) reported in some detail on such sound patterning in the "speech" of a budgerigar whose repertoire consisted of more than 600 words. The fine detail of acoustic patterning in the verbal output of a budgerigar will be examined below in the Appendix to Chapter 6.

Bierens de Haan (1929), drawing on a collection of anecdotes, argued that avian "speech" can have some of the symbolic significance of human language. Thorpe (1951), too, drew similar inferences from anecdotal material. He quoted examples of apparent generalization by a parrot in the use of speech phrases which it had originally associated with particular actions. "If this is a correct interpretation it implies that the parrots are at the beginning of acquiring a real language" (Thorpe, *ibid.*).

Koehler (1972), although beginning his discussion of "talking" birds with the comment that their "intellectual abilities" had been rather exaggerated in the past, continued: "Today, however, scientists run the risk of making too little, if anything, of this phenomenon." He referred to some examples of avian "speech" as representing "the first preliminary steps of designating".

This emphasis on incipient language places Bierens de Haan, Thorpe, and Koehler firmly in the "psychological kinship" tradition in their thinking on avian "speech".

Related to Koehler's "designating" is the "propositional language" of Ginsburg (1960), which he distinguished from "emotional language". "In order to qualify as propositional behaviour, animal vocalization must first be shown to be instrumental, or operant,

rather than respondent behaviour" (Ginsburg, *ibid.*). He found that the budgerigar's "chrp" has "the properties of an operant because it can be brought under stimulus control and operates in a lawful manner under partial reinforcement" (Ginsburg, *ibid.*). The budgerigar's "chrp" is the central subject of Chapter 5, below, in which it is shown to be one of the major building-blocks of budgerigar "speech".

Ginsburg (1963) was able, similarly, to bring some verbal vocalizations of a mynah under stimulus control. He regarded such "instrumentality" as a characteristic of language, as "human language is instrumental in securing reinforcement from the social environment". Ginsburg showed, too, that the bird could respond discriminatively to two different stimuli with two different phrases of human speech. "How many stimulus-response connexions of this kind can be retained concurrently by a bird has yet to be determined" (Ginsburg, *ibid.*). He believed that such experiments as his might be able to "bridge the gap" between animal communication and human language.

Grosslight *et al.* (1962, 1964) reported similar experimental results with mynahs. On the other hand, Lashley (1913) and Foss (1964) reported a lack of positive correlation between food reinforcement and vocalization in parrots and mynahs respectively.

Turney (1982) investigated the relationship between visual concepts and imitative vocalizations in a trained mynah. The mynah was operantly conditioned to say "hello" when shown pictures of people, and to bark like a dog when shown pictures of trees. Turney reported that the visual concept appeared to form before the bird learnt to use the appropriate vocalization: to begin with, it would sometimes bark at pictures of people, but would remain silent when shown pictures neither of people nor of trees. The bird usually responded to pictures of people within 3-5 seconds; this period was longer for the less dominant "bark" response. Turney thought that "the

necessity of first inhibiting the dominant 'hello' response prior to the 'bark' might acount for the longer latency". He compared birds and man as being "both visually gifted and vocally plastic", and regarded the ability to associate visual concepts with imitative vocalizations as something which birds alone share with man. He believed that this association could be a key to understanding the evolution and neurophysiology of vocal communiction.

Turney's experiment-based speculations thus bear a strong general resemblance to nineteenth-century "kinship" notions, while his concept of the vision-vocalization relationship, and Darwin's song-speech hypothesis, form analogous links between human and avian vocal phylogeny.

Observations by Baldwin (1914) and Gregory and Hopkins (1974) are related to Turney's latency measurements. Baldwin observed the "speech"-learning behaviour of two grey parrots (*Psittacus erithacus*) and reported that, instead of going through a period of trial-and-error, these birds will often remain apparently oblivious to the phrase being taught until, after "ten days or a fortnight the teacher is rewarded by what seems to be a first-rate imitation, springing full-formed from the bill of the parrot" (Baldwin, 1914). He suggested that a process analogous to trial-and-error could be occurring in the bird's "receptive auditory centre". (Grosslight and Zaynor, 1967, reported that mynahs sometimes produce a word once - and then never again.)

Gregory and Hopkins (1974) noticed that a yellow-fronted Amazon (*Amazona ochrocephala*) contracted the pupils of its eyes when "talking". "The pupils would sometimes respond similarly while the parrot was listening to familiar words or sounds. Further, occasionally there would be a small shrinking of the pupils immediately prior to her speaking; as though, we thought, following some kind of internal 'rehearsal' for

speech production" (Gregory and Hopkins, *ibid.*).

HUMAN/AVIAN SIMILARITIES IN VOCAL ONTOGENY AND CONTROL

In addition to the psychological kinship theories, reviewed immediately above, which are directly concerned with avian "speech", there has been considerable discussion in the literature of parallels between the ontogeny and neural control of human speech on the one hand and avian vocalizations on the other. While the connexion with avian "speech" is made explicit only in some instances (e.g. Thorpe, 1967), this connexion is implicit throughout such discussions.

Vocal ontogeny in general

Several authors (e.g. Dooling and Searcy, 1980; Kroodsma, 1981, 1982; Marler, 1970a, 1970b, 1976; Marler and Peters, 1981; Nottebohm, 1970; Pepperberg, 1985a) have drawn attention to similarities between avian vocal ontogeny and speech development in human infants. These similarities - all centred on vocal learning - include: imitation; an early sensitive period; the developmental phase of babble/subsong; the role of memory; the importance of auditory feedback; an apparent predisposition for the learning of species-specific sounds. "Nowhere else in the animal kingdom, not even among the non-human primates, have such parallels with the development of human speech been uncovered" (Kroodsma, 1981). Marler and Peters (1981) suggested that "vocal perception in birds provides a useful ontogenetic model for furthering understanding of certain aspects of the development of speech perception in infants".

The formulation of transformational-generative linguistics in the 1950s and 1960s, with its hypothesis of innate, universal "deep structures" in human language

(Chomsky, 1965), suggested parallels to researchers on inherited constraints in avian vocal learning. In man, as well as in birds, such constraints have been referred to, collectively, as a "template" (see, e.g., Smith and Wilson, 1979) which selects the sounds to be learnt and processes them according to inherited acoustic and syntactic patterns.

"Between the evolution of the apes as we know them and the first men, two immense steps in mentality and the production of the associated neurological structures were achieved. The first was the power of vocal imitation; the second was the 'deep structures' underlying human language. Birds have the first and something strongly resembling the second; apes, it seems, have neither" (Thorpe, 1967).

The timing of vocal learning

Available data on chronological patterns in avian vocal learning reveal a complex interplay of social and neuro-physiological factors. (See Kroodsma, 1981, 1982 for reviews.)

In this connexion, it is important to keep in mind some of the distinctions between calls and song which were discussed in a previous section of this chapter. One of these distinctions is that social factors are, on the whole, more important in call motivation than in song motivation.

On the basis of chronological patterns, Nottebohm (1984) distinguished between "critical-period" learning and "open-ended" learning in avian vocal ontogeny.

Critical-period learning is typically associated with song - as, for example, in the chaffinch (Thorpe, 1958a), the white-crowned sparrow (Marler and Tamura, 1964), and the zebra finch (Immelmann, 1969); open-ended learning is typically associated with calls - as in the mynah (Bertram, 1970) and some cardueline finches (Mundinger,

1970).

Thus critical-period learning reflects the constraints of innate neuro-physiological factors in vocal development; open-ended learning reflects responsiveness to a changing social environment. However, while physiological factors necessarily influence call-learning behaviour, social factors can likewise modify chronological patterns of song-learning behaviour - as in the white-crowned sparrow (Baptista and Petrinovich, 1984), the marsh wren (Kroodsma and Pickert, 1984), and the zebra finch (Eales, 1987b).

In comparing avian vocal ontogeny with language development in humans, parallels have been drawn both with the early "critical period" in language development (e.g. Marler, 1970a), and with the "open-endedness" of language learning (e.g. Nottebohm, 1970).

The critical period in human and avian vocal learning is associated with the behaviour of babbling (in human infants) and subsong (in birds) - see Chapter 5. The discussion under "'Imitation' and 'mimicry'", above, has suggested that the avian subsong phase represents a period in which the facility for indiscriminate vocal copying (as employed in vocal mimicry) is coordinated with the audio-neuro-vocal patterns involved in true vocal imitation.

Such a process would seem to be necessary in all true vocal imitation (as defined above in terms of these underlying patterns).

With regard to "talking" birds, both parrots (Nottebohm, 1970, 1975) and mynahs (Bertram, 1970) are reported to have a developmental phase comparable to oscine subsong. In addition to the early establishment of audio-neuro-vocal imitation patterns, however, these "talking" birds have an open-ended vocal plasticity which allows them to use early-established patterns in order to supplement and modify their vocal repertoires later in life. Such continuing plasticity, after the end of the

critical period, is also characteristic of human speech. (Rausch and Scheich, 1982, however, interpreted their observations of dendritic spine loss and enlargement in HVc of one-year-old speech-trained mynahs as indicating stabilization of the repertoire in adulthood.)

Nottebohm's (1984) distinction between "critical-period" and "open-ended" learning was made in the context of a discussion of vocal development in the canary. In referring to the canary as an "open-ended" learner, Nottebohm associated it with call-learning birds such as the mynah. In fact, however, Nottebohm's research has revealed the canary as rather a critical-period learner with this difference: the critical period recurs in successive breeding seasons.

Nottebohm and Nottebohm (1978) showed that male canaries can alter their song repertoires from year to year. Nottebohm (1980b) reported that the telencephalic vocal-control nuclei *hyperstriatum ventralis, pars caudalis* (HVc) and *robustus archistriatalis* (RA), which are smaller in female canaries than in males, could be induced to grow in adult female canaries by testosterone treatment. "This system is unique in the extent to which gross neural plasticity normally associated with early development can be induced in adulthood" (Nottebohm, *ibid.*). DeVoogd *et al.* (1982) concluded that the greater volume of central vocal-control nuclei in male (and testosterone-treated female) canaries was due to "a substantially larger synaptic network" than in non-singing females, while Nottebohm *et al.* (1981) demonstrated a positive correlation between the size of song repertoires in adult male canaries and the volume of HVc and RA. Similarly, Canady *et al.* (1984) reported positive correlation of (1) HVc and RA size, and (2) vocal repertoire size, in two populations of marsh wrens (*Cistothorus palustris*) which differ considerably in both (1) and (2).

In the light of his discovery that, in male canaries,

HVc is twice as large (and RA almost twice as large) in spring as it is in autumn, Nottebohm (1981) argued that "the seasonal swelling and shrinking of forebrain nuclei involved in song control is related to the seasonal learning and forgetting of a song repertoire" (Nottebohm, *ibid.*). This hypothesis was supported by experimental results of Baker *et al.* (1984), which showed that there is no seasonal fluctuation in the volume of song control centres in the white-crowned sparrow. The song of this bird remains the same from year to year, while androgen levels and frequency of singing behaviour both fluctuate seasonally as in the canary. On the basis of this comparison, Baker *et al.* concluded that seasonal fluctuation in the volume of canary song-control centres probably does reflect patterns of learning and forgetting, as hypothesized by Nottebohm, rather than fluctuation in androgen levels or the frequency of singing behaviour.

On the other hand, regression of a canary's vocal output to plastic song in late summer (Nottebohm, 1984) may be at least partly related to androgen levels, as comparable vocal regression between breeding seasons has been recorded in songbirds with no apparent seasonal fluctuation in neuronal complexity. (See Lanyon, 1960, and Marler and Peters, 1982, for reviews.)

Nottebohm *et al.* (1986) drew attention to the well-documented projection of auditory fibres to HVc (see Chapter 4), and suggested a perceptual - as well as a motor - function for the seasonal contraction and recrudescence of this telencephalic nucleus in the canary. (For example, a canary has to adjust to seasonal alterations in the songs of its conspecifics.)

Possible implications for neurology

The possibility that increases in the volume of canary vocal-control centres might result not only from the

growth of new dendrites and new synapses, but also from the appearance of new neurons, was suggested by Nottebohm (1980b). Subsequent chemical labelling experiments (reviewed in Nottebohm, 1984) have supported the neurogenesis-in-adulthood idea. Evidence of such neurogenesis in HVc of adult zebra finches (Nottebohm, *ibid.*) as well in adult canaries, however, is one indication that new neurons need not be associated with an increase in the volume of the network: this does not fluctuate in adult zebra finches (Arnold, 1980). Even in canaries, Nottebohm (*ibid.*) considered that "the recruitment of new neurons must be accompanied by neuronal death".

These two continuing processes in canary HVc - dendritic growth and retraction (associated in the literature with vocal learning and forgetting), and neuronal birth and death (not necessarily associated with vocal learning) have no known counterparts in the adult human brain. Nottebohm (1984) believes, however, that they provide hope for a "new neurology" in which - on the basis of the avian model - genetic control both of neurogenesis, and of dendritic growth and retraction, could be induced in human patients, ultimately replacing many forms of neurosurgery.

Vocal control

Nottebohm's discovery of left hypoglossal dominance in the control of song production in the chaffinch, *Fringilla coelebs* (Nottebohm, 1970, 1971, 1972a), was "the first report of a consistent right-left asymmetry of neural function in a vertebrate other than man" (Arnold, 1982). (For a brief survey of syringeal innervation see "Innervation of the syringeal muscles" in Chapter 2, below.) This discovery (related in Nottebohm, 1980) was followed by comparable findings in several other songbirds - *Serinus canarius* and *Zonotrichia leucophrys*

(Nottebohm and Nottebohm, 1976), *Zonotrichia albicollis* (Lemon, 1973), and *Padda oryzivora* (Seller, 1979)'.

Left hypoglossal dominance in song production was discussed in connexion with left hemispheric dominance in human communication by Nottebohm himself (Nottebohm, 1970, 1977; Nottebohm *et al.*, 1976), as well by other ornithologists (e.g. Marler, 1970a), writers on human communication (e.g. Busnel and Classe, 1976), and comparative neurologists (e.g. Galaburda *et al.*, 1978).

This analogy, while suggestive in outline, apparently does not obtain in anatomical detail. Lateralization in songbird vocal control relates to left-side dominance in a bilaterally symmetrical pattern of efferent neural projections to the syrinx. All efferent vocal pathways in songbirds are ipsilateral (Arnold, 1982; Nottebohm, 1980), except for some crossed connexions in the midbrain which have been shown to be unimportant in vocal control (Nottebohm *et al.*, 1976). As the songbird syrinx is paired bilaterally (see Chapter 2), left hemispheric and hypoglossal dominance result in the majority of vocal sounds being produced by the left half of the syrinx.

No auditory asymmetries comparable to right-ear advantage in human speech perception (see Chapter 4) have yet been discovered in birds (Nottebohm, 1977).

Although he suggested that both "hemispheric dominance in man and hypoglossal dominance in songbirds are associated with vocal learning", Nottebohm (1977) was well aware that there is no necessary association between vocal learning and such functional neural asymmetries. The case in point is the parrots, in which avian group there is no hypoglossal dominance for vocal control (Nottebohm, 1976). The psittacine syrinx is a unitary sound source, both sides of which are innervated by each tracheosyringealis branch of the hypoglossus nerve (Manogue and Nottebohm, 1982; Nottebohm, 1976). "Clearly left or right hypoglossal dominance has not been an inevitable consequence of the evolution of vocal

learning" (Nottebohm, 1980).

Evidence (reviewed in Chapter 4) that HVc is involved in the perception as well as the production of song has prompted an analogy between HVc and Broca's area for speech control in the human frontal lobe (Nottebohm, 1984). A further analogy has been drawn between RA and layer five of the mammalian motor cortex (Nottebohm *et al.*, 1976).

However, Nottebohm *et al.* (1986) could discern no systematic bilateral asymmetry in either HVc or RA of canaries at any stage in the seasonal fluctuation of their volume. This finding may be contrasted with asymmetries in the human cerebral cortex (reviewed by Galaburda *et al.*, 1978).

Response of motor neurons during song perception

Williams and Nottebohm (1985) recorded auditory response from neurons within the hypoglossal (i.e. syrinx-innervating) motor nuclei in zebra finches. Individual units showed selectivity for particular zebra finch song syllables. Latency measurements indicated that this auditory response could be part of a perceptual loop in which auditory signals from telencephalic audio-vocal centres (see Chapter 4) are transmitted to the hypoglossal nucleus to be analysed there in terms of vocal gestures, and subsequently follow a recurrent pathway back to the forebrain.

The authors point out that this model represents an avian counterpart of the motor theory of speech perception in humans (Liberman *et al.*, 1957).

While the motor theory of human speech perception remains highly controversial (Lane, 1965; Rosen and Howell, 1987), the findings of Williams and Nottebohm have important implications for the study of avian "speech".

PSYCHOLOGY OF "TALKING" BIRDS: A UNIFIED VIEW

The work of Pepperberg (1981, 1983), mentioned under "Social bonding theories of avian 'speech'", above, is founded on the establishment of a social relationship between a bird and its human trainer(s), and uses the communicative potential of such a relationship in order to explore the bird's cognitive processes and abilities, and relate them to their human counterparts. In this way it represents a fruitful combination of the two historical approaches to avian "speech" - social bond and psychological kinship - reviewed above.

After 26 months of training, Pepperberg's parrot - Alex - had "acquired a verbal vocabulary consisting of three colour adjectives, two phrases for describing shape, nine nouns, and functional use of the word 'no'" (Pepperberg, 1981). Alex was able to use a noun in conjunction with an adjective of colour or shape in order to identify or request more than 30 objects, and to use the word 'no' in refusing to identify them or in rejecting them. Her carefully conducted tests showed, to Pepperberg's satisfaction, that "a certain level of communicatory behaviour" had been achieved.

A subsequent training program (Pepperberg, 1983) demonstrated that Alex was able to respond appropriately to the questions "What colour?" and "What shape?" when presented with an object whose colour and shape were both represented in his vocabulary. "This skill is not just the ability to recognize a particular instance of a category (e.g., "greenness"), or even to perform a conditional discrimination (e.g., to match shapes under one stimulus and colours under another) but rather the ability to extract the relevant category from the multi-variant item and respond with the label for the one, correct instance of this category" (Pepperberg, *ibid.*).

Even errors in the early stages of training with a new object showed evidence of categorization: for example, incorrect identification of "rawhide" as "paper" was far more common than its identification as "key" or "wood" (Pepperberg, 1981).

Another significant form of generalization in the parrot's use of speech sounds resulted from a training procedure involving several human models. Pepperberg (1981) reported that Alex "has acquired vocalization patterns from all of his trainers and interchanges these patterns [e.g. in repetitions of a single word] in what appears to be a random fashion". In other words the parrot, like the human listener, is able to distinguish between general acoustic patterns of denotative significance and individual variations of no denotative significance.

Pepperberg (1981) speculated that Alex's functional use of the word "no" could represent a "mapping" of human language onto an existing behaviour pattern - i.e., that a parrot is capable of generalizing about situations in which a negative vocal response is appropriate, and of using a human vocal negative in order to express such a response in human/psittacine interactions. She reported that "'no' appeared initially to replace the parrot warning/distress call, RAAAKKKK (which can be viewed as an emphatic, or 'strong no')", but was quickly extended to situations in which a less emphatic, "weak no" was probably intended. (For further discussion of such "mapping" of speech sounds onto animal behaviour patterns, see Pepperberg, 1985b.)

Lorenz (1952) relates that a pet raven used its own (speech-sound) name in "calling" to Lorenz in a behavioural context (enticement away from apparent danger) for which there is a particular species-specific call. Lorenz implies that the bird "translated" the species-specific sound into the human-speech sound in the context of "calling" (i.e. "naming") its human social

partner.

Nevertheless, Lorenz (*ibid.*) maintained that birds were incapable of "talking to achieve an end" - e.g. of asking for something. Pepperberg, however, has shown that a parrot can use speech sounds as requests.

So little is known about the vocal behaviour of *Psittacus erithacus* in the wild, that it is impossible to say to what extent "mapping" of communication sounds could involve a one-to-one "translation" of psittacine into human vocal signals (e.g., whether there is a parrot-call equivalent of generalized "no"), and to what extent the human sounds may be mapped onto behaviour patterns which are not necessarily asociated with particular psittacine vocalizations.

Conclusions arrived at in the course of this thesis, however, indicate that the possible significance, to a bird, of the "speech" sounds in its vocal repertoire could be much more complex and multi-dimensional than is implied by this simple "mapping" schema.

These conclusions are: (1) avian "speech" represents, physically, a transformation of species-specific sounds into speech-like sounds - see Chapter 5; (2) this transformation process is facilitated by the existence of underlying bio-communication cues common to both the human and the avian sounds involved - see Chapter 4; (3) avian "speech" is acquired not merely through mimetic facility, but by means of complex audio-vocal imitation patterns characteristic of species-specific vocal learning, and (4) there is an overall correlation between "talking" behaviour and species-specific singing or calling behaviour - see this chapter, above.

Thus, in its development, structure and function, avian "speech" bears the imprint of the species-specific communication systems on which it is based (as well as the underlying features which label it as essentially communicative.) In order to explore to the full its inter-specific communicative potential, therefore, some

understanding of those species-specific systems is essential. Such detailed investigation of psittacine vocal communication is the necessary next step in the study of human-parrot vocal communication potential so auspiciously begun by Pepperberg.

CONCLUSIONS

The study of avian "speech" can provide new perspectives on bird sounds, on human speech, and on relationships between the two.

Bird sounds. In copying the complex sounds of speech, birds reveal the possibilities and limitations of those sound production, perception and processing mechanisms which determine the structure of avian vocalizations in general.

Speech. Avian "speech" is a bioacoustic reflector, in which we can observe aspects of our own communication sounds (as filtered through a non-human perceptual system) which are inaccessible to the self-scrutiny of conventional linguistics and phonetics.

Comparative bioacoustics. In revealing significant acoustic elements which bird sounds and human speech have in common, avian "speech" provides a basis for investigating, in an evolutionary context, questions of "homology" and analogy in vocal communication sounds.

A new dimension. Central to the argument of this chapter has been the distinction drawn between imitation and mimicry in avian vocal learning. That distinction, as elucidated above, rests on re-definitions of these terms according to behavioural criteria rather than (as in all previous definitions) according to descriptive or functional criteria. Consequent categorization of avian "speech" as imitation has established a conceptual framework which reveals a profoundly communicative dimension of avian "speech". It is within such a framework that avian "speech" can most profitably be investigated.

NOTES TO CHAPTER 1

[1]'Avian "speech"' is a convenient label for the sounds under examination. Although the inverted commas - which imply "quasi-speech" - are not strictly necessary in the context of this thesis, in which "speech" refers simply to "speech sounds" and not to "language", they will be retained throughout in order to avoid confusion when referring to both (human) speech sounds and (avian) "speech" sounds. It must be re-emphasised that questions of language - its definition, and possible use in some form or other by non-human species - are beyond the scope of this thesis.

[2]Similarly, in human speech, laryngectomized patients can learn to produce, by an esophageal mechanism, words which, although differing spectrally from their larynx-generated counterparts, are nevertheless functionally the "same" words.

[3]There is some confusion in the literature concerning the use of the word "innate" in reference to avian vocalizations (e.g. Marler, 1975a). I restrict the term to refer to sounds which develop normally in the absence of all auditory information - whether from an external model or from auditory feedback. As explained in the text, differences between the (innate) vocalizations of early-deafened birds and those of acoustic isolates are the result of learning through auditory feedback.

[4]Having no first-hand experience of canary "speech", I rely on the descriptions of Maynard (1928) and the acoustic data of Greenewalt (1968).

[5]Baylis (1982) avoids consideration of subsong mimicry "because of ambiguity in terminology and the difficulty of evaluating the criteria used for establishing mimicry in such rambling and plastic vocalizations".

[6]There is a report on the copying of primate calls by wild mynahs. Tenaza (1976) observed copies of "a particular loud larm call of Kloss's gibbon (*Hylobates klossii*) and loud spacing call of male pig-tailed langurs (*Simias concolor*)".

[7]Youngren *et al.* (1974) discovered asymmetry in the innervation of tracheal muscles in *Gallus gallus*. They found that part of the left *cervicalis descendens superior* branch of the hypoglossal nerve - never the right - "crosses the trachea and supplies a major portion of the innervation for the right sternotrachealis muscle" (Youngren *et al.*, *ibid.*). This crossover does not apply to innervation of the tracheolateralis. (EMG recordings by these authors showed that, unlike the tracheolateralis muscles, the sternotrachealis muscles are used only for vocalization and not for breathing.) Inactivation of the right *c.d.s.* branch of the hypoglossal nerve reduced but did not eliminate EMG activity in the right sternotrachealis muscle, confirming that this muscle is partly innervated by the left *c.d.s.* branch. Lockner and Youngren (1976) discovered the same innervation assymetry in the mallard (*Anas platyrhynchos*).

CHAPTER 2:
ANATOMY AND PHYSIOLOGY

INTRODUCTION

The mechanism of vocal sound production has never been observed in a living bird. All theories of avian vocalization are either based on anatomical, acoustic, electromyographic or air-pressure data, or are extrapolations from the results of *in vitro* experiments.

The assumption basic to most of these theories - at least until recent years - is that the fundamental vibrations of vocal sound in birds are produced by the oscillation of membranous or cartilaginous elements in the respiratory tract.

Gaunt and Wells (1973) tentatively suggested, however, that some avian whistles might be produced by an aerodynamic rather than a mechanical mechanism - i.e., one in which the vibrating system is the air itself. Nottebohm (1976) believed that many sounds of parrots could be aerodynamic whistles, and Gaunt and Gaunt (1985) presented a hypothetical model for the production of such whistles in the psittacine syrinx. S. L. L. Gaunt and A. S. Gaunt (1980) invoked an aerodynamic mechanism in explaining the generation of a carrier frequency in calls by the ring dove (*Streptopelia risoria*), and the same authors (A. S. Gaunt and S. L. L. Gaunt, 1980) suggested that this mechanism could be "widely used among birds". In a subsequent, more detailed analysis of *Streptopelia* vocalizations (Gaunt *et al.*, 1982) they asserted that a sound-production model involving aerodynamic whistles "has major implications for the study of avian vocal mechanics in general". These theories are discussed below, this chapter, under "Theories of syrinx function".

The primary source of vocal sound in most birds is thought to be the syrinx, which was identified by Herissant (1753) and named by Huxley (1871). According to some theories, sound-producing and modulating mechanical vibrations occur only in the syrinx, whereas other

writers (Bock, 1978; Brackenbury, 1978b; Owen, 1866; Thorpe, 1959; White and Chubb, 1968; Zweers, 1985) have suggested that the larynx may also have some (unspecified) vibratory function.

The syrinx is a modification of cartilaginous rings, and their connecting tissue, at or near the junction of trachea and bronchi. (See Plates 2.1, 2.2 and 2.4.) Ellis (1973) deplored the convention of describing the syrinx in terms of trachea and bronchi, rather than according it the status of a distinct organ. (He was not concerned with evolutionary factors.) Certainly, there is much confusion in the literature as to which elements of the syrinx are "tracheal" and which are "bronchial". For instance Warner (1969) refers to a certain element in a parrot's syrinx as "the last tracheal ring", whereas Nottebohm (1976) refers to the analogous element in a parrot of a different species as "the first bronchial half-ring". (Gaunt and Gaunt, 1985, quote further examples of such discrepancies.) Miller (1934) advocated a system in which "the most posterior ring with which the pessulus merges" is considered "the last tracheal ring". (To add to the confusion, the element referred to as "the last tracheal ring" is usually numbered "1" in figures.) Ames (1971) avoided these difficulties by defining elements according to their shape and consistency rather than according to their supposed tracheal or bronchial affiliation.

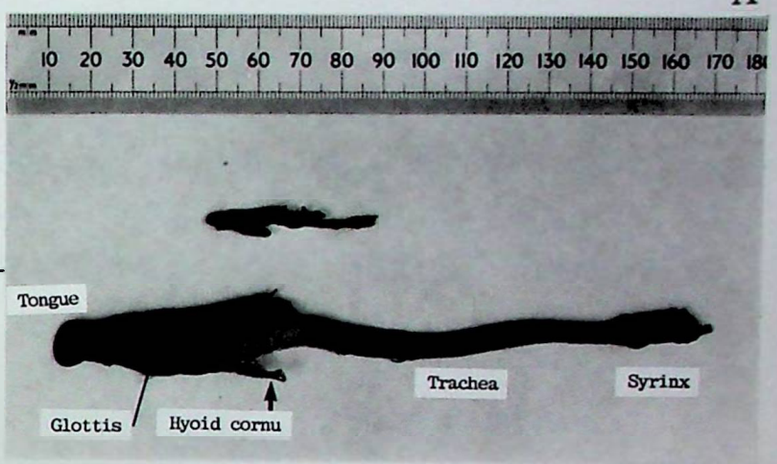
The adjectives "tracheal", "bronchial" and "tracheo-bronchial" have been applied to the syrinx itself, in order to define its position along the airway (Müller, 1847; Huxley, 1871; Beddard, 1898; Warner, 1972b; Van Tyne and Berger, 1976). This usage, however, involves an implicit definition of the anterior and posterior boundaries of the syrinx. Such a definition has never been explicitly agreed upon (King, 1979), although Häcker (1900) attempted a general definition of the boundary between trachea and syrinx (see King, *op. cit.*). Indeed,

PLATE 2.1

The avian vocal tract

A

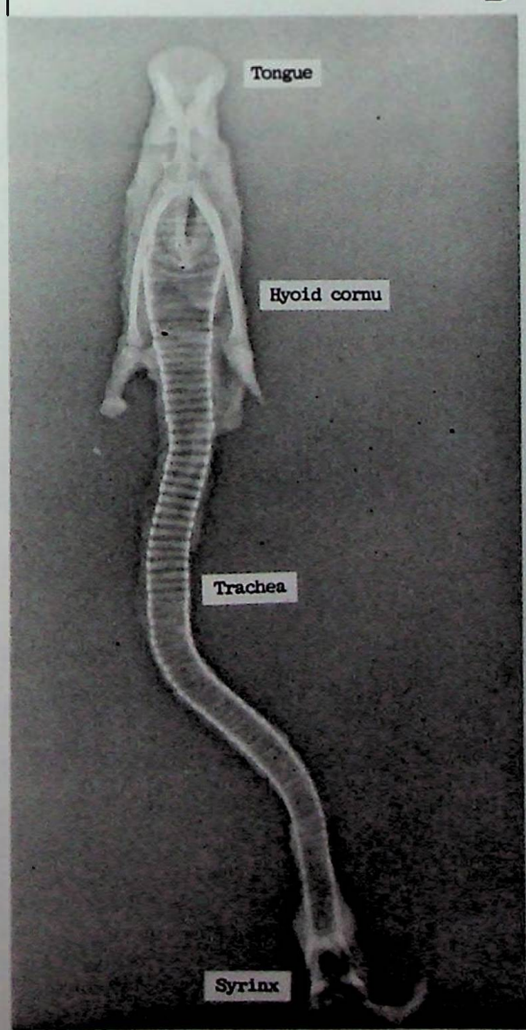
Melopsittacus undulatus



(DORSAL VIEW)

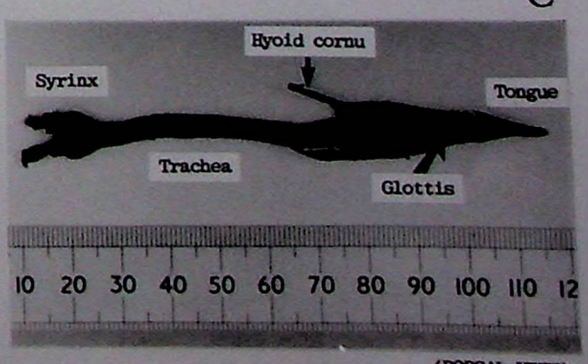
Amazona guildingii

B



Gracula religiosa

C



(DORSAL VIEW)

it would be difficult to form a general concept of such boundaries because of the great variability of syringeal structure.

It was this variability which attracted the attention of nineteenth-century taxonomists. Müller (1847) was able to subdivide the Passeriformes on the basis of syringeal anatomy. For Beddard (1898), "many of the variations [in the syrinx] are more of systematic than of morphological interest" (p 60).

The turn of the twentieth century saw a growing interest in syringeal function, which was perhaps stimulated by the detailed anatomical illustrations of Häcker (1900) and Setterwall (1901: see Greenewalt, 1968), and which produced the important work of Ruppell (1933).

It has long been known that pressure in the clavicular air sac, which surrounds the syrinx, is an important factor in the sound production of some species. Hérissant (1753) reported that by rupturing this sac, and consequently causing pressure external to the syrinx to drop, a bird could be rendered voiceless. Rüppell (1933) performed an *in vitro* experiment, enclosing an excised herring gull syrinx in a chamber in which pressure could be built up to simulate that in the clavicular sac. When air was blown through the bronchi towards the trachea, and the chamber pressure was maintained at a level above that of the atmosphere, membranes in the syrinx oscillated and produced a loud sound. These oscillations ceased when pressure in the chamber was reduced to the level of atmospheric pressure. The experiment was repeated, with the same result, by Gross (1964a) using the syrinx of a chicken, and by Wurdinger (1970) using that of a goose.

These findings suggest that the most important sound-producing elements in the syrinx are those membranes which form a barrier between clavicular-sac pressure on one side and intra-syringeal pressure on the other. (An

important exception is the syrinx of male Anatidae, in which the vibratile elements are not in contact with the clavicular sac. See below.)

VIBRATILE ELEMENTS

The paired tympaniform membranes occur either laterally (lateral or "external" tympaniform membranes[1]), or medially (medial or "internal" tympaniform membranes). In passerine species they typically occur medially (Ames, 1971; Warner, 1972b), as illustrated in Figure 2.3D. In non-passerine species they can occur medially, as in owls (Miller, 1934), ducks (Warner, 1971; Lockner and Youngren, 1976) and pigeons (Warner, 1972a), or laterally, as in parrots (Nottebohm, 1976), as depicted in Figure 2.2C.

This basically simple scheme is complicated by the occurrence along the syringeal walls of other membranous areas which have often been called "tympaniform". Thus some authors have described lateral or "external" tympaniform membranes in passerine syringes, and in non-passerine syringes having vibratile membranes medially (e.g. Rüppell on the herring gull). Conversely, Gross (1964a), while establishing that the lateral membranes are the primary sound source in the chicken (*Gallus gallus*), identifies "internal tympaniform membranes" in his figure of this syrinx.

"Tympaniform" is a morphological term (and its use to describe lateral membranes in the passerine syrinx was objected to by Ames, 1971, on morphological grounds), but its implication of a physiological function - that of vibration - has led to some confusion in the literature. A significant cause of this confusion was Häcker's (1900) influential drawing of a passerine syrinx (that of *Turdus merula*) in medial longitudinal section. This drawing, which showed lateral as well as medial "tympaniform" membranes, was reproduced and adapted to illustrate several subsequent discussions of passerine syringeal anatomy. Among those who have acknowledged Häcker as the

source of their illustrations are Stresemann (1927-1934, and hence Paulsen, 1967), Portmann (1950), Thorpe (1959, 1961), Greenewalt (1968) and King (1979). (Others who have used a similar sectional drawing have probably been influenced by Häcker.)

After this drawing became current, the idea of multiple syringeal vibrators was almost axiomatic in the study of bird sounds - to such an extent that Maynard (1928) could seriously postulate connexions between particular membranes and particular calls in a bird's repertoire. Thus in parrots he describes "tympaniform membranes" (medial tympaniform membranes, which most subsequent investigators have not found in Psittaciformes), and "outer bronchial membranes" (lateral tympaniform membranes). "Undoubtedly the loud notes that most of these birds uttered were produced by the vibrations of the ample tympaniforms", he wrote, "but other, more variable sounds, were probably due to the shorter vibrations of the outer bronchial membrane, the tensing of which could be nicely adjusted by the oscillations of the free syringeal semirings. Indeed the wonderful power of so accurately mimicing the human voice is almost certainly vested in this membrane and its control."

Secret development of the sound spectrograph during the war years culminated in the publication, in 1947, of *Visible Speech*, by Potter, Kopp and Green. This marked a turning point in the study of bird sounds, as spectrographic analysis provided detailed information on their complex structure. For instance, the authors of *Visible Speech* were the first to note the "double tone" phenomenon in bird song - i.e., the temporal overlapping of two non-harmonically related spectral components (p 411). The book also contained probably the first spectrogram ever published of speech imitation by a bird (p 323), demonstrating the potential importance of the spectrograph in the study of inter-specific differences

in sound production.

During the 1950s the use of the spectrograph in the study of bird sounds developed alongside its use in the study of speech. Borror and Reese (1956) produced spectrographic evidence of as many as four simultaneous, unrelated components in the song of the reed warbler. (D. B. Miller, 1977, published comparable data resulting from his analysis of calls made by the wood duck, *Aix sponsa*, a non-passerine species.) Thorpe (1959, 1961), in an attempt to correlate this newly-observable acoustic complexity of bird song with the number of membranes traditionally identified in the song bird's vocal tract, named what he considered to be four possible sound sources in songbirds: "the two sets of tympaniform membranes", the semilunar membrane, and elements in the larynx.

The semilunar membrane, occurring in some birds, is a cranial projection of the mucous membrane which covers the pessulus. Despite Setterwall's (1901: see Greenewalt, 1968) carefully-argued objections to the involvement of this element in sound production (summarized by Greenewalt, 1968), some writers (e.g. Miller, 1934) continued to assign it a role in vocalization. Miskimen (1951) effectively settled the issue by demonstrating that the winded syrinx of a dead starling retained its sound-production capacity after removal of the semilunar membrane.

Greenewalt (1968), mainly on the basis of Setterwall's observations concerning its area and thickness, discounted the lateral tympaniform membrane as a possible vibrator in the oscine syrinx. In Häcker's drawing of the thrush syrinx, and those derived from it, this membrane is depicted as forming the inner surface of the thick syringeal wall between what, in Häcker's terminology, are the second and third bronchial half-rings (the "A-2" and

"A-1" elements of Ames, 1971, respectively). Its position is shown as being cranial to that of the medial tympaniform membranes, and so it could not be opposed to the medial membrane in vibration. On the other hand, Chamberlain *et al.* (1968), in their report on syringeal anatomy in the crow (*Corvus brachyrinchos*), depict the lateral membranes as stretching between "A-2" and "A-1" (in their terminology, the third and fourth "bronchial bones" or "rings"), and opposed to the medial membranes. Their observations during *in vitro* aeration led them to believe that these lateral membranes were the primary vibrators in this species. (It is difficult to understand from their paper how these vibrators could have been accurately observed, as they describe the lateral membranes as lying beneath "auxiliary membranes" in the syringeal walls.)

Suthers and Hector (1982) postulated opposition of lateral and medial syringeal membranes in the production of echolocating clicks by the grey swiftlet (*Collocalia spodiopygia*). While the medial membranes are the main vibrators in the model these authors propose, they report that both lateral and medial membranes "are lined with a non-ciliated, flattened epithelium, while the remainder of the trachea, syrinx and primary bronchi have a ciliated epithelium". In their view, this indicates that both pairs of membranes could be involved in mechanical vibration during sound production.

Miskimen (1951) reported simultaneous vibration of medial and lateral membranes in the aerated syrinx of a freshly killed starling (*Sturnus vulgaris*), observing that "during the vibrations the external and internal membranes come in contact". Sectional figures of starling syringes in Warner (1972b) and Gaunt and Wells (1973), however, conform with the pattern established by Haecker in showing all tissue which could be associated with a lateral membrane as cranial to the medial membrane, making vibratory opposition impossible. Such anomalies

could reflect general problems in relating physiological function to structures observed in chemically fixed specimens.

Warner (1972b), commenting on the paper of Chamberlain *et al.*, reported that in his own wide-ranging examination of passeriform syringes (including some of the Corvidae) he found no evidence of vibratile membranes in the lateral walls of the syrinx. "The connective tissue between the first and third bronchial semirings is very loose, and it is easy in gross dissection to divide the mass, giving the appearance of an 'external tympaniform membrane' backed by an air space" (Warner, 1972b).

My own dissections of chemically preserved *Gracula religiosa* syringes revealed large, oval-shaped, translucent areas of thin connective tissue extending between the widely-spaced A-1 and A-2 elements of Ames (Warner's "second and third bronchial semirings"). While they are completely overlaid by the intrinsic syringeal muscles (Fig. 2.2A), there is only one small area of actual muscle insertion on each of these surfaces. Thus, although free vibration of the tissue would be restricted because of this muscle attachment and contiguity (Greenewalt, 1968), strictly myogenic vibration - serving to periodically alter the dimensions of the intrasyringeal airway - is an obvious possibility. (See "Theories of syrinx function", below.)

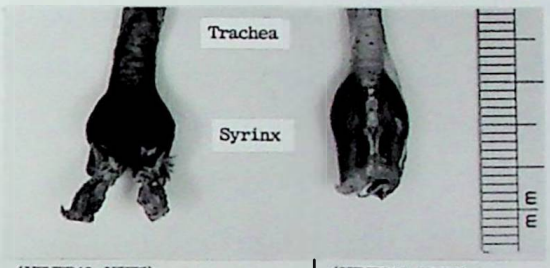
One of these membranes has been exposed in Figure 2.2B, which is a ventral view of the same *Gracula* syrinx shown in Figure 2.2A. (The A-1 and A-2 elements are labelled simply "1" and "2".) The muscle insertion can be seen on its ventrocaudal extremity (near the bottom of the photograph). While certainly thicker than the medial tympaniform membranes of *Gracula*, these lateral membranes seem little - if at all - thicker than the lateral tympaniform membranes of, for example, *Amazona guildingii* (Fig. 2.2C), which are generally regarded as vibrators. The photograph clearly shows a thickening of the tissue

PLATE 2.2

Comparative syringeal anatomy

Gracula religiosa *Psittacus erithacus*

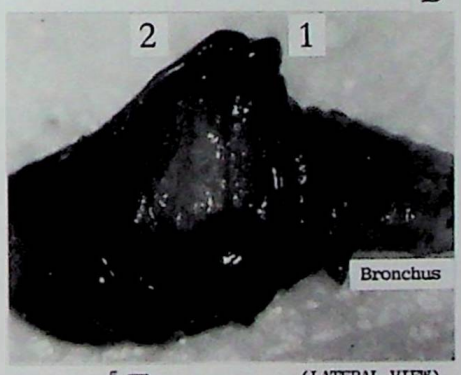
A



(VENTRAL VIEW) (VENTRAL VIEW)

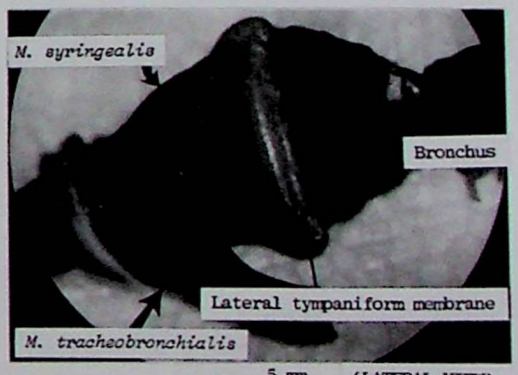
Amazona guildingii

B



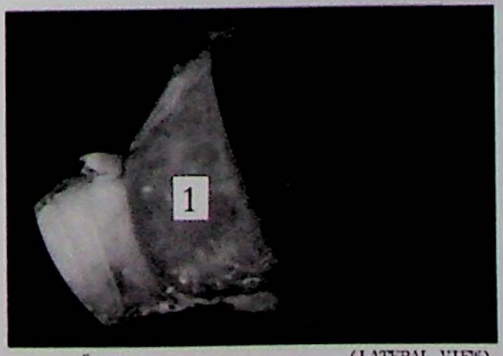
5 mm (LATERAL VIEW)

C



5 mm (LATERAL VIEW)

D



5 mm (LATERAL VIEW)

E

(VENTRAL VIEW)

towards its caudal edge, in that it becomes, caudally, increasingly opaque. This thickening represents the craniad extent of the lateral labium. Stein (1968) suggested that these lateral membranes in songbirds may act in conjunction with the lateral labia (see the following paragraph) "by allowing for additional flexibility in the movement of the latter".

The lateral labia are pads of loose connective tissue continuous with the caudal extremity of each lateral membrane in songbirds. Greenewalt (1968), as well as Stein, attributed important functions in oscine vocalization to these labia, and subsequent writers (e.g. Gaunt, Stein and Gaunt, 1973; Gaunt and Wells, 1973) have endorsed this view. The lateral labia also play a vital role in the mechanism of mynah speech imitation hypothesized by Klatt and Stefanski (1974). These hypothetical functions of the lateral labia will be discussed below under "Theories of syrinx function".

A pair of medial labia are also usually identified in songbirds. These are continuous with the cranial extremity of each medial tympaniform membrane, and are "effectively the tracheal portions of these membranes" (Warner, 1972b). At their cranial ends they "are attached to the pessulus or are histologically continuous with it" (Ames, 1971). Their size and shape vary greatly among species (Warner, 1972b). Although these medial labia have "no established function" (Gaunt and Wells, 1973), Haecker's comparative drawings of male and female syringes, showing several elements, including the medial labia, absent or rudimentary in the female, suggest that they may play a part in male song production.

A vibratile function has been attributed to "labial" folds (Lockner and Youngren, 1976) in the syringes of male Anatidae (figured in Warner, 1971). As mentioned above, these "labia" (which are not homologous to the

labia in the oscine syrinx) are not in contact with the clavicular sac. Lockner and Youngren (*op. cit.*) have suggested that they vibrate "in a manner similar to the mammalian vocal cords".

AIR FLOW AND PRESSURE

The clavicular air sac - *Saccus clavicularis* (King, 1979) - which encloses the syrinx is medial and unpaired. Most of the other air sacs which, together with the lungs, trachea and primary bronchi, comprise the avian respiratory system, are paired bilaterally. (The system is figured in Brackenbury, 1980, 1982; Evans, 1969; Gaunt *et al.*, 1973.) Radiographic observations of gas movements within the respiratory system (James *et al.*, 1976) indicated that the air sacs serve to move air into and out of the lungs. The lungs are relatively rigid, and expiration is effected by compression of the air sacs by the intercostal and abdominal muscles. Miskimen's (1951) *in vitro* air-flow experiments using the syringes of both passerine and non-passerine species showed that sound is produced during this exhalant, or active, phase of respiration. This finding was subsequently corroborated by electromyographic data on the synchronicity of expiratory muscular activity and vocalization (Gaunt *et al.*, 1976; Lockner and Youngren, 1976).

Gaunt and Gaunt (1977) recorded greatly increased electromyographic activity in the abdominal muscles of chickens just before the onset of vocalization. In crowing, this high level of activity continued throughout the call, whereas during some other calls it subsided from an initial peak. Gaunt *et al.* (1982) found that during vocalization in ring doves both the pattern of activity in the abdominal muscles and variations in air sac pressure corresponded closely with the complex pattern of the sound pressure waveform. Muscular activity began 10-30 msec before the onset of the sound.

Increases in air sac pressure and air flow rate as a result of expiratory muscular action during both quiet breathing and vocalization have been directly measured in

a number of species. (See Brackenbury, 1982, for review.) In geese, air sac pressure during honking is 40-100 times greater than during quiet expiration, while air flow rate is only 17-20 times greater, indicating an increase in airway resistance of two to six times (Brackenbury, 1978b). This resistance during vocalization appears to be much geater in some passerine birds than in non-passerine species (Gaunt *et al.*, 1973, 1976). (See below, this chapter, for a discussion of mechanisms which could impose such a resistance.)

Gaunt *et al.* (1973) found that tracheal pressures during distress calls by starlings were, despite increased air-sac pressures, usually lower than during quiet respiration and often scarcely higher than atmospheric pressure. This showed, they said, "that the syringeal valve effectively separates the air sacs from the trachea, thus providing a sharp gradient for the conversion of mechanical energy (pressure) to sound energy". Tracheal pressures during vocalization in chickens (Gaunt *et al.*, 1976; Brackenbury, 1978b) and geese (Brackenbury, *op. cit.*) showed considerable variation both within and between calls, reflecting changes in syringeal resistance independent of air-sac pressure. One call of the chicken - referred to as a "wail" by Gaunt et al. (1976) and Gaunt and Gaunt (1977) - involves stable tracheal pressures as low as any of those recorded in the starling, while pressures during other calls, such as crowing, fluctuate between this low level and 12cm H_2O (Brackenbury, *op. cit.*). (These data will be considered in relation to physiological models later in this chapter, and in relation to acoustic data in Chapter 3.)

In their investigation of echolocating clicks by grey swiftlets, Suthers and Hector (1982) recorded a sudden drop in tracheal pressure in the "intraclick interval", while air sac pressure remained high. They interpreted this as indicating complete closure of the syringeal

valve during this 25 msec interval of silence.

An ability to impose a variable resistance on expiratory air flow implies the presence of Bernoulli forces in the syrinx as air particles move at increased velocity past these obstructions. If pressure in the clavicular air sac were equal to that in an unobstructed syrinx, the imposition of such a resistance would create a pressure difference across the syringeal wall separating the clavicular sac from the intra-syringeal cavity at the point of obstruction. If the wall at this point were membranous, vibrations of such a membrane could occur.

Even if the syringeal lumen were not actively obstructed, the configuration of the lumen could cause a significant increase in air-particle velocity through the syrinx and thus a drop in syringeal pressure resulting in membrane vibration. The *in vitro* experiments of Rüppell (1933), Gross (1964a) and Wurdinger (1970), described in the Introduction to this chapter, indicated that clavicular-sac pressure is important in such passive[2] sound production as well as in vocalization which is actively initiated. Rüppell's results were questioned by Gottlieb and Vandenbergh (1968), who produced sounds from duckling syringes in the atmosphere by gentle aeration. Several other authors have produced sounds from atmosphere-exposed syringes, but most of these (e.g. Miskimen, 1951) had to pass air through them at unnaturally high pressures. Gaunt *et al.* (1973) reported that, in *in vivo* experiments, higher flow rates were required to force vocalizing vibrations in the syringes of conscious starlings than in unconscious birds, "indicating some ability to suppress sound production".

Gaunt *et al.* (1973) argued that the silencing of starlings by rupturing the clavicular air sac was due to the re-routing of air flow through the cannula and away from the syringeal valve, rather than to decreased clavicular-sac pressure. Youngren *et al.* (1974) showed

that this is not so, at least in chickens, by blocking the connexions between lungs and clavicular sac[3], thus forcing air through the syrinx at a normal vocalizing rate even though the clavicular sac was ruptured. No sound was produced. When the rupture was resealed, vocalization was restored even though rates of flow through the syrinx were low. Lockner and Murrish (1975) threw further doubt on the "alternative pathway" theory by showing that mallard drakes, whose vocalizing ability is not dependent on clavicular-sac pressure (see above), vocalize normally when such an alternative pathway is provided by rupturing the sac. (Gottlieb and Vandenbergh, 1968, found that even mallard ducklings, in which the sound-producing mechanism independent of air-sac pressure had not yet developed, were unaffected vocally by rupturing the clavicular sac. They used these observations in arguing that clavicular-sac pressure is unimportant in avian sound production generally.)

SYRINGEAL MUSCULATURE

The muscles of the upper airway in birds have been described according to several systems of nomenclature. Tables of synonymy can be found in Ames (1971), Warner (1972) and King (1979). One of the main conceptual disagreements underlying these differences has been whether the muscles should be named on the basis of function or of location. Ames (*op. cit.*) has cogently argued that "until more is known about the operation of the syrinx it seems unwise to name the muscles on a functional basis". The muscles are usually divided into "tracheal" and "syringeal" muscles but, as King (1979) points out, this distinction is not really clear. "A muscle which has extensive contact with the trachea could reasonably be called a tracheal muscle; yet by pulling on the trachea such a muscle will almost certainly act indirectly on the syrinx, and could therefore be regarded as a syringeal muscle" (King, *op. cit.*).

Most non-passerine species have only a single pair of indisputably syringeal muscles - the tracheolateralis muscles, which originate on the cricoid cartilage of the larynx and extend down the sides of the trachea to insert on the syrinx. Some ratite birds, storks, and American vultures lack even this (Beddard, 1898; Van Tyne and Berger, 1976). The adjective "extrinsic" is often applied to this syringeal muscle because one of its attachments lies beyond the syrinx. In evolutionary terms, *M. tracheolateralis* can be seen as the archetypal syringeal muscle, and the "intrinsic" syringeal muscles[4] typical of the oscines as subdivisions of its caudal extremity (Maynard, 1928; Ames, 1971).

As well as many passerine species, the other group of birds to have developed a notable intrinsic syringeal musculature is the parrots (Beddard, 1898). This

musculature is less complex in Psittaciformes, however, than in the typical songbird syrinx, and Maynard (1928) remarked on the anomaly that "the parrots, in spite of their unrivalled ability to mimic human speech, possess comparatively simple vocal organs". It is significant that the Passeriformes and the Psittaciformes include the best-known "talking" species.

The various systems of nomenclature have differed in the number of intrinsic syringeal muscles described - discrepancies arising from disagreement as to "whether a single mass of muscle fibres attached to several cartilaginous elements constitute one muscle or several" (Ames, 1971). As many as eight pairs of intrinsic syringeal muscles have been named in Passeriformes (see tables referred to above) whereas, among nineteenth-century writers, Garrod (1874) distinguished only "four or five" and, among modern writers, Ames has named four and Warner (1969, 1972) five.

INNERVATION OF THE SYRINGEAL MUSCLES

The syringeal muscles (both extrinsic and intrinsic) are innervated mainly through branches of *N. hypoglossocervicalis* (Breazile and Yasuda, 1979). The major tracheal/syringeal branch of *N. hypoglossocervicalis*, the *ramus cervicalis descendens* of Breazile and Yasuda, is more commonly termed *ramus cervicalis descendens superior*. This is in recognition of an *inferior* branch which, according to Nottebohm and Nottebohm (1976) lies "in close apposition to the vagus nerve" and may contribute to syringeal innervation in some species. According to these authors, "the syrinx is also supposed to receive a very slender branch of the recurrens Xth (vagus) nerve".

Nottebohm and Nottebohm expressed a belief that most of the nerve fibres distributed to the syrinx through branches of *N. hypoglossocervicalis* are of hypoglossal origin. This belief is implicit in the writings of several other authors (e.g. Evans, 1969; Lemon, 1973; Lockner and Youngren, 1976; Seller, 1979; Youngren *et al.*, 1974). Breazile and Yasuda, however, emphasize the confusion which prevails concerning the distribution of fibres from the cranial nerves through these branches, and suggest a cervical and/or vagal origin for many of them.

There are simple (perhaps over-simplified) figures in Nottebohm (1976) and Manogue and Nottebohm (1982) illustrating syringeal innervation in psittacine species, and similar figures relating to passerine species in Nottebohm (1971) and Nottebohm and Nottebohm (1976). More detailed figures relating to other species include an illustration of tracheal innervation in the mallard (Lockner and Youngren, 1976), and one of syringeal innervation in the chicken (Youngren *et al.*, 1974).

FUNCTION OF THE SYRINGEAL MUSCLES

Extrinsic syringeal muscles

Much of the early work on muscle function during avian sound production concentrated on the chicken (*Gallus gallus*). Gross (1964), mainly on the basis of *in vitro* experiments, propounded a theory of vocalization for this species in which the sternotrachealis muscles play a major role.

Mm. sternotrachealis (Owen, 1866, and most subsequent authors) originate on the sternum and insert on either side of the trachea. In *Gallus* "these muscles reach the trachea at the level of the tenth tracheal ring (above the syrinx) where they are bound to it by a fascial sheath. Their fibres begin insertion at the thirteenth ring and finish at the twenty-first" (Youngren *et al.*, 1974).

Gross proposed that contraction of *Mm. sternotrachealis* stretches the trachea in such a way that its posterior end - the tympanum - is drawn caudad. This movement shortens the distance between the tympanum and the pessulus, thus relaxing tension on the lateral tympaniform membranes[6]. "Normally the membranes are held apart by tension of the trachea. In this position air resistance is minimized and sound cannot be produced. Since only the tympaniform membranes attach the tympanum to the pessulus (a distance of about 1.5 mm), the syrinx can easily be shortened by a posterior movement of the tympanum. This is ordinarily accomplished by contracting the sternotrachealis muscles" (Gross, 1964). Relaxation of tension allows the membranes to partly occlude the airway, thus increasing resistance to air flow and initiating vibration. (For a discussion of air-pressure

forces acting on the membranes, see above under "Air flow and pressure".)

This concept of a caudad movement of the tympanum which places the syrinx in a "sound-producing configuration" (Brackenbury, 1980a) has been the foundation of most subsequent theories of vocalization in *Gallus*.

The development of electromyographic techniques allowed these theories to be tested by recording activity in the muscles of a living bird. Electromyographic data reported by Youngren *et al.* (1974) showed activity in the sternotrachealis muscles of *Gallus* only during vocalization - an indication, at least, that these muscles play a significant part in sound production. On the other hand, the tracheolateralis muscles were active during quiet breathing as well as vocalization. (Youngren *et al.* elicited vocalizations from anaesthetized chickens by electrical stimulation of the brain. In the *ad libitum* studies of Gaunt and Gaunt, 1977 - summarized below - the tracheolateralis muscles were inactive during most periods of quiet breathing. In the opinion of these authors, and of Gross, 1964a, the elasticity of the trachea is sufficient to keep the membranes out of the airway during quiet respiration in conscious birds.)

Mm. tracheolateralis (Owen, 1866; Ames, 1971; Warner, 1972b; King, 1979) extend along either side of the trachea for almost its entire length. Youngren *et al.* found that, in *Gallus*, these muscles both originate and insert on the tracheal rings, the insertion being "just rostral of the tympanum". They suggested that their function during quiet respiration is to maintain enough tension on the lateral tympaniform membranes to keep them out of the airway. During vocalization, when contraction of the sternotrachealis muscles, according to the "classic model"[6] described above, relaxes tension on the membranes and allows them to vibrate, contraction of *Mm.*

tracheolateralis could adjust membrane tension, thus controlling the frequency of vibration. "The opposed actions of the two pairs of muscles provides for very precise control of membrane tension" (Youngren *et al.*, 1974).

(Miskimen, 1951, on the basis of *in vitro* experiments on starling syringes, had proposed that such antagonistic activity of *Mm. sternotrachealis* and *tracheolateralis* was fundamental to vocalization in songbirds. See below.)

Mm. cleidotrachealis. There are two major difficulties with a theory which assigns to *Mm. sternotrachealis* the task of preparing the syrinx of the chicken for vocalization. The first is that these muscles are relatively weak in chickens (Gaunt and Gaunt, 1977) as they are in songbirds (Warner, 1972b), and the second is that electromyograms show the onset of their activity is coincident with that of vocalization (Gaunt and Gaunt, 1977; Youngren *et al.*, 1974). An anticipatory phase of activity might have been expected if these muscles were indeed drawing the syrinx into a vocalizing position. Gaunt and Gaunt (1977) proposed a solution to these problems by nominating the cleidotrachealis muscles as the main agents of syringeal preparation.

M. cleidotrachealis (Beddard, 1898; see Gaunt and Gaunt, 1977, and King, 1979, for synonymy) is a relatively large muscle which originates on the clavicle. In *Gallus* it inserts on the ventral surface of the cricoid cartilage of the larynx. "Its caudal third is closely applied to the overlying skin, while its cranial two-thirds form a sling around the trachea. Except in the vicinity of its insertion, connection to the trachea is not direct, but by a loose fascia" (Gaunt and Gaunt, 1977). Contraction of the cleidotrachealis muscles, according to these authors, "shifts the trachea caudad as a unit". They support their theory with electromyographic data, which show a short anticipatory phase of activity

in these muscles. Then, once the syrinx has been set in its vocalizing configuration, the antagonistic action of *Mm. sternotrachealis* and *tracheolateralis*, as in previous theories, could control the tension of the vibrating membranes.

Brackenbury (1980a), by electrically stimulating the tracheal muscles of anaesthetized chickens, confirmed the established model of antagonism in the actions of *Mm. sternotrachealis* and *tracheolateralis*, but did not confirm the function of *M. cleidotrachealis* as proposed by Gaunt and Gaunt.

Muscle inactivation experiments by Youngren *et al.* (1974) also supported the established theory, in showing that vocalizations of *Gallus* deteriorated less after inactivation of both *Mm. sternotrachealis* and *tracheolateralis* than after removal of *M. sternotrachealis* alone. After the latter operation, all low-intensity calls were eliminated. (Miskimen, 1951, found that domestic cocks were able to crow loudly after removal of *Mm. sternotrachealis*. Brackenbury, 1978b, reported on experiments in which clucking as well as crowing was virtually unaffected by sectioning these muscles, and Gross, 1964b, reported similar results.)

Youngren *et al.* interpreted their own results as reflecting the disruptive effect of inactivating only the muscles which release membrane tension, while the tensor muscles remain in operation. If both pairs of muscles are inactivated, air-pressure forces alone might be able to initiate and maintain membrane vibration. In a subsequent paper (Phillips and Youngren, 1981) these authors reported that inactivating *M. cleidotrachealis* had no noticeable effect on calls.

Contrary to its effect on *Gallus*, Miskimen (1951) found that removal of *Mm. sternotrachealis* in the cardinal - a songbird - resulted in complete and permanent muteness. (Smith, 1977, reported virtually

unchanged vocalizations after section of these muscles in three other species of songbird. Here, as in the experiments on *Gallus* summarized above, the discrepancies may result from differing effects of the mere section of the muscle, and its excision.) Miskimen believed that, in songbirds, fine control of vocalization was achieved by the intrinsic syringeal muscles once the extrinsic muscles had produced the basic sound-producing configuration. The role of the extrinsic syringeal muscles during vocalization in songbirds remains unclear however, for it will be shown below that the intrinsic muscles have functions analogous to those of the extrinsic muscles in *Gallus*. Warner (1972b) believed that the sternotrachealis and tracheolateralis muscles "play a minor part in vocal control" in passerine birds.

Intrinsic syringeal muscles

Ames (1971) rationalized earlier accounts of intrinsic syringeal muscles (see above, under "Syringeal musculature") by recognizing only four pairs of intrinsic muscles in passerines. These are *Mm. tracheobronchialis dorsalis* and *ventralis* (Warner, 1972; King, 1979), and *Mm. syringealis dorsalis* and *ventralis* (King, 1979; see King for synonymy of passerine intrinsic muscles).

Mm. tracheobronchialis dorsalis and *ventralis* originate in the bifurcation of the tracheolateralis muscles (Müller, 1847; Miskimen, 1951; Warner, 1972b) and insert on the dorsal and ventral ends of those syringeal elements which would be called "A-1" according to Ames's system of nomenclature. (See Ames, 1971, for synonymy of syringeal cartilages. The photograph in Figure 2.2B illustrates the position of these elements in a mynah's syrinx. Compare this figure with the medial longitudinal section of a mynah syrinx illustrated in Warner, 1972.)

Müller (1847) recognized the functional significance

of the tracheobronchialis muscles' attachments. He wrote that contraction of these muscles results in the "rotation" of the syringeal elements on which they insert, thus altering the configuration of the syringeal airway. Owen (1866) wrote that the "third bronchial half-ring" ("A-1" of Ames) "is susceptible of a rotary movement on its axis, and is an important agent in the modulation of the voice".

Miskimen's experimental results support the deductions of Mueller and Owen. She, too, uses the word "rotation" in describing the movement of the syringeal cartilage on which the tracheobronchialis muscles insert. Contraction of *Mm. tracheobronchialis posticus* and *anticus*, she observes, draws the cartilage cephalad and rotates its cephalad edge inward. The caudal edge of the lateral membrane, which is attached to this cartilage (see Fig. 2.2B), is thus moved towards the medial membrane in the opposite bronchial wall, thereby partly occluding the bronchial lumen.

Unfortunately, the illustrations in Miskimen's paper tend to confuse the issue; much better illustrations, showing this "rotation" of the cartilage and consequent displacement of the lateral tympaniform membrane, are presented by Chamberlain *et al.* (1968). Although their accompanying text is badly edited (there is confusion in the names of muscles and the functions attributed to them), it is clear that they are in agreement with Miskimen on the function of the tracheobronchialis muscles as outlined above.

Mm. syringealis dorsalis and *ventralis* originate on the tympanum and insert on the dorsal and ventral ends of Ames's syringeal element A-2. *M. syringealis dorsalis* is absent in some species (Miskimen, 1951). According to Chamberlain *et al.* its probable function is to apply tension dorsally to the lateral membrane. Miskimen noted that contraction of *M. syringealis ventralis* is

accompanied by a rise in the pitch of the sound produced. The most likely explanation of this is that the muscle controls tension in the medial membrane. The influence of *M. syringealis ventralis* on the medial membrane is supported by Chamberlain *et al.*, who observed that contraction of this muscle "turns the internal tympaniform membrane close to the external tympaniform membrane".

On the available evidence, then, the function of the intrinsic syringeal muscles in oscines can be summarized as follows: contraction of *Mm. tracheobronchialis dorsalis* and *ventralis* produces a gross movement of the lateral tympaniform membranes into the bronchial lumina towards the medial tympaniform membranes, thus preparing the syrinx for vocalization in a manner analogous to that of the sternotrachealis or cleidotrachealis muscles in chickens; contraction of *Mm. syringealis dorsalis* and *ventralis* controls tension and/or fine positioning of the lateral and medial membranes.

Psittaciformes

The two pairs of extrinsic syringeal muscles in parrots are homologous with extrinsic syringeal muscles in other groups of birds. Thus one pair of extrinsic muscles is termed *sternotrachealis* (Evans, 1969; Warner, 1969), even though these muscles insert "on the oblique septum near the entrance of the bronchi into the lungs, not on to the sternum as in other birds" (Gaunt and Gaunt, 1985; see also Evans, 1969). The tracheal portions of *Mm. sternotrachealis* are relatively thick, but, from the beginning of their passage over the surface of the longer pair of intrinsic muscles to their insertion points, they are no more than slender tendons. (See figures in Evans, 1969, and Gaunt and Gaunt, 1985.)

The tympanum of the psittacine syrinx, like that of the chicken, is attached to those cartilaginous elements

caudal to it only by soft tissue, consisting mainly of the lateral tympaniform membranes. As in the chicken, contraction of *Mm. sternotrachealis* "can shorten the tracheo-bronchial junction and thus relax the syringeal membranes" (Evans, 1969). There is a detailed description of this process under "Observations on syrinx function in parrots", below.

Electromyographic studies reported by Gaunt and Gaunt (1985) showed that, as in other birds, the sternotrachealis muscles are involved more in the basic activation of the sound-production mechanism than in the precise modulation of the signal.

Mm. tracheolateralis (Warner, 1969) are poorly developed and are absent in some species (Gaunt and Gaunt, 1985).

Beddard (1898) linked the oscines and the parrots in their both exhibiting "at its extreme . . . an hypertrophy of the intrinsic [syringeal] muscles" (p. 66). The comparably powerful intrinsic syringeal musculature of both these groups of birds is illustrated in Figure 2.2A, which shows the syringes of a mynah (*Gracula religiosa*) and a grey parrot (*Psittacus erithacus*). While the muscles of both these birds are comparable in thickness, the musculature of the mynah is obviously the more complex. It conforms to the description of Warner (1972b), who wrote that "in smaller passerine birds, the syringeal muscles completely envelop the tracheobronchial junction in a globose muscle mass, and identification of the individual muscles is difficult".

The major difference between the intrinsic musculature of the passerine and the psittacine syrinx can be generalized simply by saying that the passerine syrinx has four pairs of intrinsic muscles, while the psittacine syrinx has only two. This ratio of 4:2 (2:1) reflects the basic structural difference between a "double" syrinx

(mynah) and a "single" syrinx (parrot).

Thus the parrot syrinx has one pair of muscles analogous to the two pairs of tracheobronchialis muscles in songbirds, and likewise one pair of muscles analogous to the songbirds' two pairs of syringealis muscles. (According to Gaunt and Gaunt, 1985, neither of these pairs of muscles appears to be homologous with their analogues in the songbird syrinx.)

Mm. tracheobronchialis (Gaunt and Gaunt, 1985; Nottebohm, 1976; Warner, 1969; synonymous with *Mm. bronchotrachealis* of Evans, 1969, and Manogue and Nottebohm, 1982), in *Psittacus erithacus*, originate on the sixth and seventh cartilaginous elements of the trachea cranial to the lateral tympaniform membranes, and insert on each bronchus on the second to the sixth elements caudal to the membranes (Fig. 2.2A). (There are slight variations on this pattern in other species in the present study - e.g. *Amazona guildingii* in Figure 2.2E - and species studied by Warner, 1969, Nottebohm, 1976, and Gaunt and Gaunt, 1985.)

It has been found convenient in this work to adapt the system of Ames (1971) in naming the bony and cartilaginous elements of the psittacine syrinx. For reasons discussed below under "Observations of syrinx function in parrots", the series extending craniad from the lateral membrane is labelled A-1, A-2, etc., the element immediately caudal to the membrane is called the "syringeal cartilage", and the series extending caudad from the syringeal cartilage is labelled B-1, B-2, etc. See Warner (1969) and Nottebohm (1976) for useful illustrations of the psittacine syrinx in medial longitudinal section.

Nottebohm (1976) propounded a theory of psittacine vocal physiology according to which the movement of the syringeal cartilages and the attached membranes in the "single" syrinx of parrots (see Nottebohm's figure) is

remarkably similar to that in the "double" syrinx of songbirds as described and figured by Chamberlain *et al.* (1968). Nottebohm proposed that contraction of the tracheobronchialis muscles draws the caudal ends of the syringeal cartilages craniad. This would result in the adduction of their cranial ends, which oppose each other on either side of the tracheal lumen. The lateral tympaniform membranes, which are attached to the cranial edges of these elements, would thus be brought closer together - close enough to allow turbulence or Bernoulli forces to operate in the transformation of mechanical energy into acoustic energy.

Gaunt and Gaunt (1985), however, pointed out that "the bronchi are fixed in position and cannot be drawn craniad". They proposed, instead, that contraction of *Mm. tracheobronchialis* produces a similar effect to contraction of *Mm. sternotrachialis*: a caudad movement of the tympanum. (The tympanum consists of the fused elements A-2, A-3 and A-4. A-1, which articulates with the latero-caudal edges of the tympanum, is, of course, drawn caudad along with it. See the detailed description below, under "Observations on syrinx function in parrots".)

The Gaunts' electromyograms revealed that, in contrast to the activity of the sternotrachealis muscles, contraction of the tracheobronchialis muscles is pulsatile, the pulses generally coinciding with periodic peaks in the acoustic waveform. Thus the action of *Mm. tracheobronchialis* can be seen as supplementing the gross effects of *Mm. sternotrachealis*: as providing, that is, the fine control over the basic caudad movement of the tympanum necessary for complex acoustic signalling.

(The results of my own experiments with excised psittacine syringes - see Plate 2.3 and the discussion, below - indicate that the mechanism proposed by Gaunt and Gaunt is nearer the truth than that of Nottebohm.)

Mm. syringealis' (*Mm. bronchialis* of Evans, 1969; *Mm. trachealis* of Warner, 1969; *Mm. syringeus* of Nottebohm, 1976, Manogue and Nottebohm, 1982, and Gaunt and Gaunt, 1985) are broad muscles extending over the lateral surfaces of the tympanum and A-1 elements (Fig 2.2C; in Figure 2.2D the muscle has been removed from element A-1). In *Psittacus erithacus* these muscles originate on the A-4 and A-5 elements. (There are slight interspecific variations on this pattern, as reported above for *Mm. tracheobronchialis*. See Figures 2.2A and 2.2E for the relative positions of these two pairs of muscles.) The syringealis muscles insert on the caudal processes of the A-1 elements. Histological preparations by Gaunt and Gaunt showed that their insertion "blends into" the lateral tympaniform membranes, so that the membranes can act as aponeuroses which insert on the syringeal cartileges.

Contraction of *M. syringealis* exerts a craniad force on the caudal edge of A-1 and, as a result, a laterad force on the cranial edge of the attached membrane, tending to withdraw the membrane from the tracheal lumen. Thus *Mm. tracheobronchialis* and *syringealis* function antagonistically in controlling the aperture of the syringeal valve.

THEORIES OF SYRINX FUNCTION

The foregoing survey has revealed patterns of muscular activity which can control, in a complex but complementary manner, both the overall resistance to air flow through the syrinx, and the tension of vibratile membranes near the place of constriction. It remains to discuss the interaction of this variable airway resistance and membrane tension with variable air flow and pressure in the generation of acoustic signals.

Mechanical vibrations involved in the production of periodic acoustic signals will be discussed as either "source vibrations" or "modulating vibrations". A source vibration is one whose fundamental frequency can correspond to a component frequency of the amplitude spectrum of the sound. A modulating vibration is one whose fundamental frequency typically does not correspond to a component frequency of the amplitude spectrum.

Source vibrations

Aerodynamic. In those species, particularly, in which syringeal constriction of the airway involves adduction of opposing skeletal elements, the possibility that destabilized air flow through this rigid constriction could produce aerodynamic whistles must be considered. It has been explained, above, how contraction of *Mm. tracheobronchialis* can cause such a constriction in the psittacine syrinx; Nottebohm (1976) suggested, but did not elaborate on, the possibility of syrinx-generated aerodynamic whistles in parrots. He also suggested, in passing, a possible interaction between those acoustic vibrations caused by periodic vortices in the destabilized airstream and those caused by membrane oscillation. (Gaunt and Wells, 1973, explicitly mentioned

the possibility of syrinx-generated aerodynamic whistles in songbirds, but doubted the overall importance of such a technique. Implicitly, however, their treatment of membrane vibration amplitude as a function of the "frequency of driving pressure perturbation" divided by the resonant frequency of the membrane foreshadowed the later work of Gaunt and his team outlined in the following paragraphs.)

Gaunt and Gaunt (1985), who advocated a somewhat different mechanism of psittacine syrinx constriction from that proposed by Nottebohm (see "Function of syringeal muscles", above), nevertheless agreed in principle with his suggestion on aerodynamic whistles. The Gaunts' most detailed treatment of this subject, however, is in connexion with their work on sound production in the ring dove (*Streptopelia risoria*), in which they developed Nottebohm's idea of an interaction between aerodynamically and mechanically-generated vibrations.

S. L. L. Gaunt and A. S. Gaunt (1980) explained the carrier frequency of the ring dove's coo as "a periodic turbulence in a flowing column of air (a whistle) that couples to and is amplified, but not generated, by the internal tympaniform membranes". Gaunt *et al.* (1982) expanded on this theme and suggested a variety of means by which the whistled carrier could be modulated. They conceived of the dove's syrinx as being forced into the configuration of a two-hole tone whistle (Wilson *et al.*, 1971) by contraction of the sternotrachealis and tracheolateralis muscles - the airway constrictions being (1) at the cranial end of each bronchus and (2) between the infolded lateral tympaniform membranes. They interpreted quasi-periodic amplitude peaks in the waveform envelope of vocalizations as reflecting pulsatile activity of abdominal or syringeal muscles, amplitude modulation imposed on the carrier by vibration of the lateral tympaniform membranes (which are cranial

to the medial membranes, near which they believed the carrier to be generated), or beats caused by slight frequency differences in the carrier wave produced at each tracheo-bronchial junction.

Gaunt *et al.* (1982) arrived at the aerodynamic-whistle model of avian vocalization partly as a result of asking a fundamental question: How can a vibrating membrane produce a pure (sinusoidal) tone? Pure-tone phrases are a common component of passerine song - see Greenewalt's (1968) "Portfolio of whistled song" - as well as occurring in the call repertoires of many non-passerine species (the grey parrot being a good example - Forshaw, 1973). Earlier models of avian vocalization (e.g. those of Greenewalt, 1968; Stein, 1968; Klatt and Stefanski, 1974) had assumed that vibrating syringeal membranes (and, in some cases, labia) were the source of all periodic acoustic vibrations - simple as well as complex. Gaunt and his colleagues reported, however, that "nothing we have seen in the analysis of . . . membrane vibrations suggests that overtones could be absent", thus calling into question the assumption that simple (i.e., sinusoidal) acoustic vibrations can be produced by an oscillating membrane.

An even more basic assumption in such sound-production models as those of Greenewalt and Stein was that syringeal membranes *must* vibrate in an exhalatory air stream, once a "vocalizing" configuration of the syrinx has been effected. These theories were primarily concerned with explaining the means of bringing about such a configuration; the physics and physiology of membrane vibration itself were dealt with cursorily, if at all, by their authors. It was assumed that passive elastic restoring forces would necessarily interact with the Bernoulli forces (resulting from rapid air flow through the syrinx) acting on a "vocalizing" membrane, thus establishing and maintaining membrane oscillation. Casey (1981), however, suggested that, in some cases, the

elasticity of the membrane could be actively compensated for by muscular contraction, the function of the membrane then being simply to respond to the Bernoulli forces in maintaining partial occlusion of the airway during production of aerodynamic whistles.

Mechanical. Nevertheless, the vibratory capacity of syringeal membranes, and the direct relationship between such mechanical vibrations and the acoustic vibrations associated with at least some avian vocalizations, has been convincingly demonstrated *in vitro* by Ruppell (1933), Miskimen (1951), Gross (1964a and 1964b), Hersh (1966), Paulsen (1967), Chamberlain *et al.* (1968), Harris *et al.* (1968), and Wurdinger (1970). There is still no adequate physical analysis of these observed vibrations.

Greenewalt (1968) apparently believed that, when subject to more than a critical level of myogenic tension, the medial tympaniform membranes in a vocalizing songbird's syrinx would oscillate - normally to the airstream and simply in the fundamental mode - to produce sinusoidal acoustic vibrations. In fact, while such a single-mode membrane vibration is theoretically conceivable, the necessary conditions "are extremely rigid and would only rarely be met by any biological system" (Gaunt *et al.*, 1982).

Greenewalt's theory goes on to state that, below the critical level of myogenic tension, the relaxed membrane is capable of making contact with the opposite wall of the bronchus - thus, incidentally, interrupting bronchial airflow - during its periodic excursions into the lumen. When this occurs, the membrane is forced to depart from what Greenewalt considered its erstwhile sinusoidal motion, with the result that one or more harmonics become apparent in the acoustic spectrum. According to his theory, spectral envelope shape as well as fundamental frequency is controlled simply by myogenic tension applied to the membrane. Greenewalt offered no

experimental or theoretical evidence for such generation and control of a purely harmonic spectrum.

Casey's (1981) theoretical analysis of vibration in rectangular, circular and elliptical "syringeal" membranes showed that partials (i.e., spectral components at non-integer multiples of the fundamental) are always present; in rectangular membranes harmonics are generated as well as partials, but in the vibratory patterns of circular and elliptical membranes only partials are to be found. It is unlikely, then, that any membrane, even under the external constraints postulated by Greenewalt, could vibrate in such a way as to produce a purely harmonic spectrum - unless it were forced to do so by another vibration (see below). Such purely harmonic spectra, however, are common in bird sounds - see Greenewalt's "portfolio".

Phonation. The periodic complete closure of the bronchial lumen which is a necessary - but incidental - part of Greenewalt's theory of harmonic-spectra generation is probably, in fact, much more directly involved in the production of such spectra. An harmonically-patterned complex acoustic vibration can be considered as not primarily related to vibratory modes of the membrane itself, but rather to a pulse train resulting from periodic occlusion of the airway by the membrane. The mechanism would then be similar in effect - although not in many of the functional details - to that of mammalian phonation. This will be referred to as the "phonatory" model of songbird vocalization, as distinct from Greenewalt's "tympanic" model.

Even if the "phonatory" model is accepted, however, there remains the mystery of the vanishing partials. Although all spectral components generated simply by the pulse train would be harmonically related there must be other components, generated by the complex vibratory patterns of the membrane itself, which are not so

related. In many instances these partials are nowhere to be seen.

A possible answer to this problem can be found in recent research on the vibration of kettledrum membranes (Rossing, 1982a and 1982b). It has been observed that "although the modes of vibration of an ideal membrane are not harmonic, a carefully tuned kettledrum will sound a strong principal note plus two or more harmonic overtones" (Rossing, 1982a). Rossing explains this phenomenon as being the result of several factors - among them the interaction of drum-cavity acoustic resonances with the vibratory modes of the membrane.

With regard to a syringeal membrane functioning in the "phonatory" manner described above, harmonic components in the complex periodic acoustic vibration associated with a pulse train could interact with the vibratory modes of the membrane itself, forcing conformity to the harmonic pattern. The relative power of such a pulse train - in comparison with acoustic resonances within a kettledrum - suggests that the forcing effect would be much greater in the case of the syringeal membrane than in that of the drum membrane.

Another factor affecting the vibration of kettledrum membranes, according to Rossing, is the surrounding "sea of air", disturbances within which tend to lower the lowest-frequency components of the spectrum so that a more harmonic-like distribution of components is achieved. If a drum membrane vibrates in a "sea of air", then a syringeal memrane could be said to be the vibratile barrier between a fast-flowing (bronchial) current and a tidal pool of (clavicular-sac) air. The complex patterns of air movement thus impinging on the membrane (see "Air flow and pressure", above), could contribute to the shifting of partials into harmonic frequencies.

Gaunt and Wells (1973) pointed out that, in any model which involves periodic complete closure of the syringeal

lumen during vocalization, the development of a pressure head caudal to the closure during the closed phase must contribute to initiation of the open phase. They suggested a model in which the intra-bronchial configuration of the membrane is maintained (by some unspecified means) throughout vocalization. Once the membrane has effected bronchial closure, it acts no longer as a membrane but as the moving part of a "flutter valve", being periodically forced away from the opposing bronchial wall by the pressure head caudal to the closure, and then returning to the closed position as a result of Bernoulli and elastic forces.

The acoustic effect of this variant of the "phonatory" model would be, once again, a pulse train having an harmonic spectrum. The proposed mechanism is very similar to that controlling vibration of the mammalian vocal folds. Gaunt and Wells apparently regarded the membrane, as employed in this model, as being indeed a "vocal fold" - i.e., as being folded into the bronchial lumen in such a way as to assume the mechanical characteristics of a much more massive soft-tissue element. Thus they avoided the missing-partials problem which must be encountered in any model involving true membrane vibration.

Klatt and Stefanski (1974) considered the lateral labium to be a more likely candidate for such a vocal-fold-like function. Their paper is concerned only with mynah speech-imitation sounds, and they proposed labial vibration as the most likely source of the low-frequency fundamental in such sounds. Contraction of *Mm. tracheobronchialis*, resulting in "rotation of skeletal element A-1", they explain, would insert the labium into the bronchial lumen, where it would interact with the exhalatory airstream according to the "flutter valve" mechanism invoked by Gaunt and Wells. (See above.)

Their theory goes on to state that, during the open phase of the labium's vibration, "air rushes past the internal tympaniform membrane, inducing wave-like

vibrations in this structure". They offer no real physical explanation as to the generation of these "wave-like vibrations" in the medial membrane, however.

Casey (1981), in his exposition of the Klatt and Stefanski theory, argues that Bernoulli forces need not necessarily act normally to the membrane. "When acting acutely or obliquely to the membrane's surface, Bernoulli forces could be easily transmitted longitudinally along its surface. They would set up, in effect, a 'ripple' or wave down the length of the membrane" (Casey, *ibid.*). Casey does not propose a mechanical model for the generation and propagation of these "Bernoulli waves".

Assuming, for the moment, that labial vibration of the "flutter valve" variety does occur, the generation of travelling waves on the medial membrane can probably be understood in terms of Bernoulli forces. As the tip of the labium approaches a certain, relatively small, area of the membrane's surface, these forces could produce an inward bulge at that part of the membrane. This bulge could then be transmitted along the membrane as a "ripple" by one - or a combination - of several factors, including: (1) continued movement of the labium in close proximity to the membrane, (2) bronchial airflow, and (3) a vector parallel to the membrane surface (as suggested by Casey) resulting, in part, from the angle of the applied Bernoulli force. Resonant frequencies of the membrane would in most cases determine the period of the travelling wave.

Membrane resonances. According to Klatt and Stefanski, "membrane oscillations can occur at several resonant frequencies simultaneously, although the lowest frequency resonance often predominates in the calls of songbirds. We believe that these membrane resonances produce some of the formant-like resonance patterns that are seen in the [speech-imitation] spectrograms."

These authors go on to suggest that, as the labium strikes the membrane at the peak of each of its vibratory cycles, "it resets the state of the membrane and thus maintains a periodic vibration pattern".

Amplitude-frequency coupling. Stein (1968) proposed an actual mechanical connexion between the lateral labium and the medial membrane in explaining certain spectral patterns in songbird vocalizations. These patterns included a correlation between periodic changes in amplitude (amplitude modulation) and periodic changes in frequency (frequency modulation). Stein regarded vibration of the labium as an important means of amplitude modulating the membrane-generated source signal. He believed AM to be the result of "periodic changing of the cross-sectional area of the lumen" by this modulating vibration (see definition above) of the labium. If FM is the result of periodic changes in membrane tension, as Stein believed, a mechanical connexion between membrane and labium could possibly explain the observed coupling of amplitude and frequency within a modulating period. Stein went no further in his explanation of this.

Greenewalt (1968), too, observed and explained coupling of amplitude and frequency - both within a modulating period and also in non-periodic amplitude-frequency changes. Greenewalt observed both direct and inverse coupling - i.e., amplitude varying directly with frequency and amplitude varying inversely with frequency. He found that, for any particular species, there is a

critical frequency below which coupling is direct and above which it is inverse. His explanation is, basically, that the amplitude of the medial membrane's vibration is a function of membrane tension. Up to a certain point, increases in tension withdraw the vibrating membrane further from the opposing bronchial wall, allowing greater vibratory excursion. (This is a misconception - see the following paragraph.) Beyond the critical point, tension begins to restrict vibratory excursion. As the fundamental frequency of vibration is also a function of membrane tension, changes in tension would affect, simultaneously, both the amplitude and the frequency of the source vibration according to the observed patterns. (Greenewalt, working simultaneously with - and independently from - Stein, proposed an identical labial mechanism to Stein's in explaining those instances of amplitude change in which there is no corresponding change of frequency.)

Greenewalt's theory is here marred by a fundamental confusion between distension of the membrane into the lumen under air pressure forces, and wholesale mechanical movement of the membrane (complete with skeletal attachments) towards the opposing bronchial wall. In the first case, the greater the distension the greater must be the amplitude of vibration; in the second case, progressive narrowing of the lumen could result in a corresponding fall in vibration amplitude. While Greenewalt's whole theory is founded on the axiom of the first concept (distension), he unwittingly introduces the consequences of the second. This flaw has remained undetected by his critics.

Stein, more simply and more realistically, maintained that the range of tension-regulated resonant frequencies at which a syringeal membrane can vibrate in the exhalatory airstream has both upper-frequency and lower-frequency limits, and that amplitude will necessarily be attenuated in the vicinity of both these limiting

frequencies.

Membrane response curve. Gaunt and Wells (1973) added another variable to Stein's implicit equation in providing an alternative explantion for Greenewalt's amplitude-frequency coupling data. They called that additional variable "driving pressure perturbation".

Gaunt and Wells addressed the problem of explaining the critical frequency observed by Greenewalt, at which the amplitude of an FM signal reaches a maximum, and above and below which the amplitude diminishes. They proposed that this critical frequency corresponded to a periodic pressure fluctuation in the exhalant airstream. If this "driving pressure perturbation" were to remain constant, a gradual increase in the tension-regulated resonant frequency of the membrane, starting near the lower limit, would eventually result in a correspondence of membrane resonance and perturbation frequency, at which point the amplitude of membrane vibration would reach a maximum. Further increase in the tension-regulated resonant frequency of the membrane would result in a decrease in vibration amplitude. A plotted curve showing vibration amplitude as a function of membrane resonant frequency would thus be, in effect, the response curve of a resonator, the peak response frequency corresponding with the frequency of "driving pressure perturbation".

(A comparable curve could presumably be obtained by maintaining a constant membrane resonant frequency, varying the airflow perturbation, and plotting vibration amplitude as a function of driving perturbation frequency. In this case, the peak response frequency would correspond with the resonant frequency of the membrane.)

As Greenewalt's amplitude-frequency data can be expressed by just such a curve, the Gaunt and Wells model must be considered as a possible alternative to

Greenewalt's. The Gaunt and Wells theory of 1973 was deficient, however, in that it made no attempt to explain the nature or origin of the essential "driving pressure perturbation". It remained for Gaunt and his colleagues (1982) to associate this periodic perturbation with the vortex series of an aerodynamic whistle.

Gaunt *et al.* (1982) regarded the medial tympaniform membranes as resonators which can respond to the input energy of the aerodynamic perturbation. They went on to propose that membrane-generated vibrations could become part of a feedback loop controlling the frequency of the driving perturbation itself. In this way, membrane tension could actually determine the fundamental frequency of the acoustic signal. Correspondences between membrane tension and fundamental frequency (as observed, for example, by Harris *et al.*, 1968), which had formed the basis of earlier sound-production theories (such as that of Greenewalt), could thus be explained in a new light.

Modulating vibrations

Stein (1968) reported modulating frequencies between 90 Hz and 300 Hz among 12 songbird species. He quoted data on vertebrate neuro-muscular activity which indicated that active, cycle-by-cycle neuro-muscular control of tissue vibration was unlikely to occur at frequencies above 50 - 60 Hz.

Electromyographic studies by Gaunt and Gaunt (1985) of vocal control in parrots showed a pulse-for-pulse correspondence between amplitude modulation and activity of the syringeal muscles at frequencies no higher than about 50 Hz.

Greenewalt (1968) presented data on modulating frequencies in the recorded songs of 15 passerine species. The range of these modulating frequencies is about 150-350 Hz. He noted that, in the song sparrow

(*Zonotrichia melodia*), for example, modulating frequencies appeared to cluster around an upper limit of 300 Hz. As "it is most unlikely that frequencies of this order can be produced by neural pulses operating one or more times per modulating period" (Greenewalt, *ibid.*), he regarded the prevalent modulating frequency of 300 Hz as "the natural frequency of an oscillating system comprising one or more pairs of syringeal muscles, with the mass of the system, the muscular tension, and the elasticity of the muscle tissue controlling the frequency" (Greenewalt, *ibid.*).

A model of psittacine sound production, which will be presented in Chapter 3, involves such resonant oscillation of a syringeal system.

OBSERVATIONS ON SYRINX STRUCTURE AND FUNCTION IN PARROTS

Structure

The following anatomical descriptions are based on the examination of two *Psittacus* syringes (one fresh and one preserved in alcohol) and two *Amazona* syringes (likewise, one fresh and one preserved in alcohol). They cannot, therefore, be regarded as definitive in terms of species, genus or family, but are intended rather as a necessary introduction to the experimental observations on syrinx function which it is the aim of this section to communicate.

Amazona guildingii. Ames (1971) remarked that "the borderline between cartilage and bone in the passerine syrinx is not a sharp one", and described all the supporting elements as "cartilaginous". This ambiguity does not apply in the psittacine syrinx. Examination of an Amazona guildingii syrinx (Figs 2.2C, 2.2D and 2.2E), subsequent to staining with toluidine blue, revealed a clear demarcation between elements cranial to and caudal to the lateral tympaniform membranes: elements cranial to the membranes are osseous; elements caudal to the membranes are cartilaginous. An identical pattern was reported by Gaunt and Gaunt (1985) in the five parrot species they investigated, and wider studies in the future may well confirm this as a general feature of the psittacine syrinx. (This basic pattern does not, of course, take into account the varying degrees - among species and among conspecific individuals - to which some calcification of elements caudal to the membranes may occur.)

This clear demarcation of bone from cartilage provides a convenient basis for the naming of elements in a

parrot's syrinx. Gaunt and Gaunt (1985) refer to the bony elements as "tracheal" and the cartilaginous elements as "bronchial" but, as has been pointed out above, these terms, although hallowed by tradition, are anatomically imprecise and can lead to confusion. In the present work the nomenclature of Ames, which was designed to reflect a distinction between two basic kinds of cartilaginous element in the passerine syrinx, has been adapted - with one important exception - to reflect the bone/cartilage distinction in the psittacine syrinx.

The exception is the pair of elements immediately caudal to the lateral tympaniform membranes. ("Caudal to" really is an inadequate description of the relationship of cartilage and membrane here, as this thin cartilage is so intimately involved with the very fabric of the membrane - see Figure 2.2C. Nottebohm, 1976, described how, in *Amazona amazonica*, the cranial edge of the cartilage merges "very smoothly with the caudal edge of the external tympaniform membrane". "This blending between cartilage and membrane is so subtle," he wrote, "that the presence of the first bronchial half-ring[a] was overlooked when the ventral view of the syrinx was first drawn." This blending of tissue makes it difficult to distinguish the outline of the cartilage in Figure 2.2C, but Evans's 1969 drawing of the lateral aspect of a *Melopsittacus* syrinx clearly shows the shape and position of this cartilage in the budgerigar to be similar to that in the *A. guildingii* specimen pictured here. Examination of *Psittacus erithacus* specimens revealed variations in the shape and position of the cartilage, discussed below, which may be of significance in speech imitation.)

Ames based his classificatory distinction on the cross-sectional shape of cartilages as well as on their consistency, and the thin, dagger-shaped cross section of this most cranial cartilage in the parrot's syrinx sets it apart both from the bony elements cranial to it and from the cartilaginous elements caudal to it. This,

together with the unique features mentioned in the preceding paragraph, has prevented it from being allied, for purposes of nomenclature, with either of these series of elements. It will be referred to here by the provisional label of "syringeal cartilage".

The pair of elements immediately caudal to the "syringeal cartilages" are thus "B-1" elements, while the pair of elements immediately cranial to the lateral tympaniform membranes are "A-1" elements. (The A-1 elements are the only paired syringeal elements cranial to the lateral tympaniform membranes. Gaunt and Gaunt, 1985, while labelling these elements "tracheal", considered the possibility that they may have developed from "enlarged, or fused, ossified bronchial semirings". Evans, 1969, indeed, regarded the A-1 elements as "bronchial", and referred to them as "free semirings" because of their freedom of movement - see next paragraph.)

The A-1 elements are massively developed. (See Figure 2.2D, where this element is labelled simply "1".) They are unrounded dorso-ventrally. (This can be clearly observed in Figure 2.1B.) Their caudal edges are semi-oval in outline, their dorsal and ventral extremities extending caudad as tapering cornua. These extremities, viewed laterally (Figs 2.2C and 2.2D), extend dorsally and ventrally beyond all other syringeal tissue. Viewed ventrally - Figs 2.2A (*Psittacus erithacus*) and 2.2E (*Amazona guildingii*) - and dorsally, the cornua of the two A-1 elements lie adjacent to each other, but not contiguous, and are capable of some lateral movement (Nottebohm, 1976). The cranial edges of the two A-1 elements articulate with the unpaired A-2 element so that their caudal edges can be drawn craniad.

The lateral tympaniform membranes are suspended from the caudal edges of the A-1 elements. Caudally, these membranes are bounded by the syringeal cartilages, while ventrally and dorsally, they are bounded by ligaments

which extend from the cornua of the A-1 elements to the ventral and dorsal extremities of the syringeal cartilages.

As mentioned above, the shape of the syringeal cartilage in *Amazona guildingii* (Fig. 2.2C) resembles that in *Melopsittacus undulatus* as figured by Evans (1969). Its lateral aspect is that of a shallow U, with serifs directed towards the tips of the A-1 cornua. (These cornua extend dorsally and ventrally beyond the corresponding extremities of the syringeal cartilages.) Thus the lateral tympaniform membrane occupies a "roughly elliptical" area, as described by Gaunt and Gaunt (1985).

Contraction of *Mm. tracheobronchialis* (see above, under "Function of syringeal muscles") draws the tympanum, and thus the A-1 elements, caudad, so that the caudal edges of these elements approach the dorsal and ventral extremities of the syringeal cartilages. This decreases the syringeal surface area occupied by the lateral tympaniform membranes which, as a result, fold into the tracheal lumen along straight lines connecting the cranial edges of the syringeal cartilages' dorsal and ventral extremities. Mediad folding of the membranes is ensured by the means outlined in the following paragraph.

Caudad movement of the A-1 elements exerts a caudad force on the dorsal and ventral extremities of the syringeal cartilages through the ligaments which extend from these extremities to the A-1 cornua. The rostrocaudal orientation of the syringeal cartilages - their cranial edges directed mediad at the angle of the bronchi's junction with the trachea (see Figure 2.2E) - means that this caudad force results in mediad rotation of the cartilages' cranial edges into the tracheal lumen. Another factor in bringing about this configuration is the craniad force which contraction of *Mm. tracheobronchialis* exerts upon the "B" elements on which they insert. (See above, under "Function of syringeal muscles"). This results in a craniad force on the caudal

edges of the syringeal cartilages (Nottebohm, 1976), which reinforces the mediad rotation of their cranial edges. Thus the mediad folding of the lateral tympaniform membranes, which are so intimately connected to the cranial edges of the syringeal cartilages, is ensured.

Clavicular sac pressure no doubt also reinforces this mediad folding of the membranes. (See "Air flow and pressure", above.)

Psittacus erithacus. The shape and position of the syringeal cartilages in the grey parrot (*Psittacus erithacus*) differ from those in other psittacine species either examined here or described in the literature, and may be a significant factor in the celebrated speech-imitation capacity of this species. (It is surprising, considering the renown of grey parrots as "talkers", that the literature contains no detailed study of their sound production mechanism.)

In *Psittacus erithacus* the dorso-ventrally oriented cranial edges of the syringeal cartilages are straight, and not semi-oval, in outline. (The caudal edges of the A-1 elements, medial to the cornua which extend caudad at their dorsal and ventral ends, are, correspondingly, straighter than those exemplified in Figure 2.2D.) Also, the syringeal cartilages are positioned further craniad than in all other parrot species studied, so that, even in the non-vocalizing syrinx, they lie largely within the area outlined by the caudal edges of the A-1 elements, the tips of the A-1 cornua extending caudad to the level of the cartilages' caudal edges. As a result the lateral tympaniform membranes are rostrocaudally shorter, relative to the overall size of the syrinx, than in the *Amazona* species studied, and the syringeal surface area occupied by the membranes is correspondingly smaller. These modifications to the shape and position of the syringeal cartilages in *Psittacus erithacus* could - for the two reasons outlined below - facilitate control of

the intra-syringeal aperture.

Firstly, the straight cranial edges of the syringeal cartilages ensure that, when these edges rotate into the tracheal lumen towards each other, the folds in the lateral tympaniform membranes of the vocalizing syrinx oppose each other with parallel straight edges oriented dorso-ventrally. Contact between the two opposing membranes is thus precisely along the midsagittal plane of the lumen. This contact can be of the edges only (as in human falsetto phonation: see below) or of more extensive membrane surfaces (as in human modal-register phonation) folded together under increased muscular tension. (See Gaunt *et al.*, 1976, for a description of an analogous syringeal-membrane configuration in chickens.) Secondly, the more cranial position of the syringeal cartilages in the syrinx of *Psittacus erithacus*, and the correspondingly shorter tympaniform membranes (rostrocaudally), may allow movements of cartilage and membrane to be more precisely dependent on each other.

Function

Phonation mechanism. Basic syringeal function during vocalization in *Psittacus erithacus* can be demonstrated using an excised syrinx. (The experiment is illustrated in Figures 2.3A, 2.3B and 2.3C.) The trachea of the *Psittacus* specimen illustrated in Figure 2.2A was severed about 1 mm cranial to the origin of the tracheobronchial muscles. The syrinx was then propped upright with pins on a horizontal surface, so that the cranial aspect of the preparation revealed the internal surface of the bronchial junction caudal to the open syringeal lumen (Fig. 2.3A). This surface appears as a brightly-illuminated area occupying most of the intra-syringeal part of the image. (Ruppell, 1933, presented a drawing of this surface in the syrinx of a swan, *Cygnus cygnus*, viewed cranially after section of the syrinx through the

PLATE 2.3

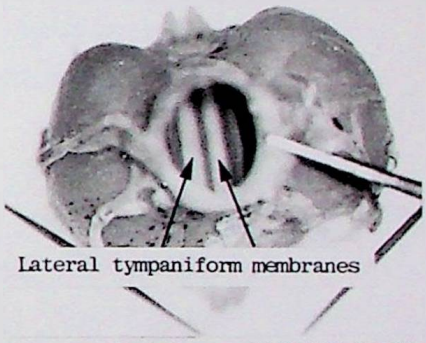
"Single" and "double" syrinx

Psittacus erithacus

Open A

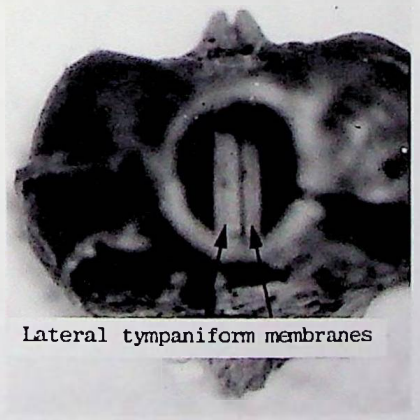
5 mm (CRANIAL VIEW)

Partly closed B



5 mm (CRANIAL VIEW)

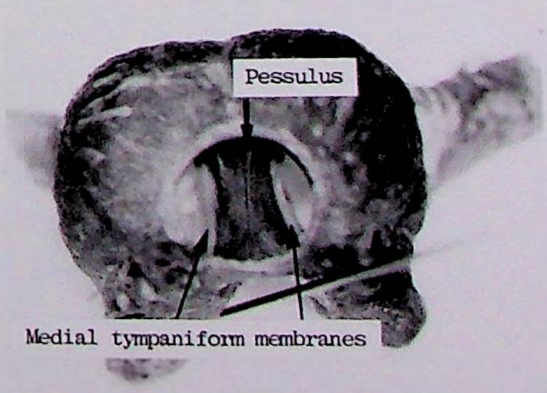
Closed C



5 mm (CRANIAL VIEW)

Gracula religiosa

D



5 mm (CRANIAL VIEW)

dorso-ventral midline of the lateral tympaniform membranes°.) The internal surfaces of the A-1 elements and, medial and caudal to these, the cranial edges of the retracted membranes, can be seen on either side of the lumen.

A small caudad force applied to the cranial end of the preparation effects such a caudad movement of the tympanum and A-1 elements as, in the living bird, is produced by contraction of *Mm. sternotrachealis* and *tracheobronchialis*. (The resistance of the surface beneath the preparation produces the effect, simultaneously, of a craniad force on the bronchi similar to that postulated by Nottebohm, 1976. See above, under "Function of syringeal muscles".) The consequent mediad folding of the lateral tympaniform membranes can be clearly observed (Fig. 2.3B).

When adduction of the membranes continues under the application of a slightly increased force, the folds in the membranes, sharply defined dorso-ventrally by the straight cranial edges of the syringeal cartilages, meet along the midsagittal plane of the lumen (Fig. 2.3C) as postulated above on anatomical grounds.

This function of the lateral tympaniform membranes in the psittacine syrinx was noted briefly by Gaunt and Gaunt (1985); a similar mechanism in *Gallus gallus* had received similarly brief treatment by Gaunt *et al.* (1976) and Gaunt and Gaunt (1977).

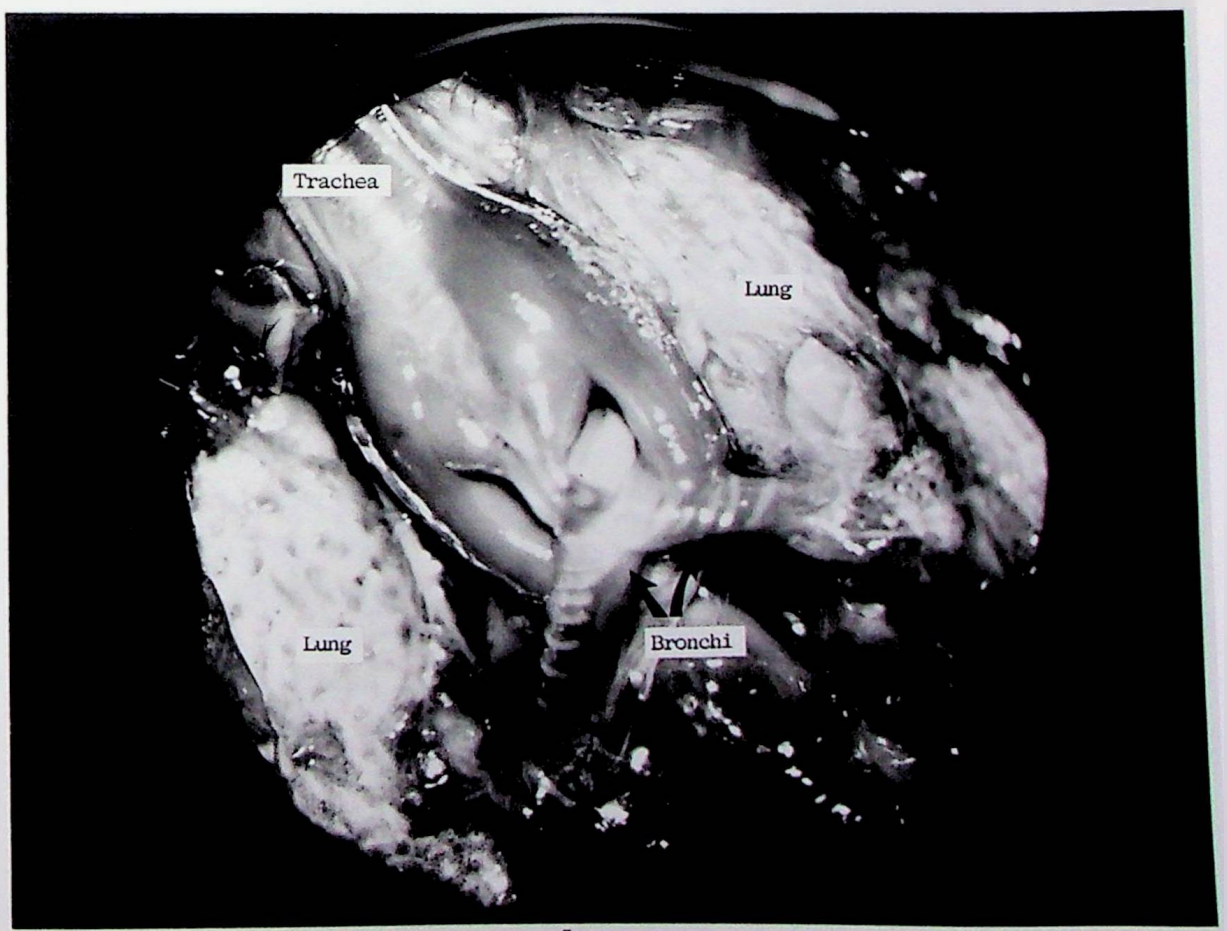
The three relevant figures on Plate 2.3 represent the first attempt to illustrate this function.

(Gaunt and Gaunt, 1977, were mistaken in believing that the figures of Beebe, 1925, were illustrations of such a function. Beebe performed a similar experiment to the one described above with the excised syrinx of a tinamiform bird, *Crypturus variegatus variegatus* [Gmelin], a species which, according to Beebe, lacks all intrinsic syringeal muscles. Although Beebe's description is unclear, his series of figures indicates that each

PLATE 2.4

Location of the syrinx

Psittacus erithacus



(VENTRAL VIEW)

bronchus of the excised syrinx was, when examined, completely occluded by opposing elements which met along the midline of the bronchial lumen. These elements could be abducted by stretching the syrinx rostrocaudally, in imitation of the action of *Mm. tracheolateralis* - see "Function of syringeal muscles", above - indicating a mechanism of vocalization in this species very different from that in parrots being investigated here. Paulsen, 1967, published a series of frames from a high-speed cinefilm showing vibratory movements of one of the medial tympaniform membranes in the artificially aerated syrinx of a goose, *Anser anser domesticus*. The excised syrinx was filmed in cranial aspect, the membrane in question occupying a similar position in the picture to the right medial tympaniform membrane in Figure 2.3D[10].)

The corresponding aspect of a songbird syrinx (that of a mynah, *Gracula religiosa*) is presented (Fig. 2.3D) for comparison with the parrot syrinx. It is obvious here that midsagittal adduction of the (medial) tympaniform membranes is impossible because of the intervening bony pessulus.

Comparison with human vocal mechanism. There are striking similarities in appearance and function between the psittacine lateral tympaniform membranes, as illustrated on Plate 2.3, and the human vocal folds, as illustrated, for example, by Lieberman (1977). Compare the following passage, describing the cranial aspect of the vocal folds, with Figures 2.3A, 2.3B and 2.3C and the description of the psittacine syrinx above. "The vocal folds appear as two white ridges just below and medial to the vestibular [ventricular] folds. Actually, they are the prominent free margins of the cricovocal membrane, with the overlying mucosa tightly bound down" (Singh, 1980).

Owen (1866) and Warner (1969), on the basis of such anatomical similarities, suggested the possibility of

analogous functions in the psittacine syrinx and the human larynx. Gaunt and Gaunt (1985) made the same suggestion in relation to their comments on membrane function in parrots (see above), following their similar conclusions about a phonatory mechanism in chickens (Gaunt *et al.*, 1976; Gaunt and Gaunt, 1977). These suggestions can be graphically supported by comparing Figures 2.3A, 2.3B and 2.3C with the series of three photographs in Lieberman (1977, p 12) illustrating open, semi-closed and closed phases in a phonatory cycle of the human vocal folds.

The human vocal folds, adducted for phonation by contraction of the interarytenoid muscles (supplemented by contraction of the lateral cricoarytenoid muscles and the lateral parts of the thyroarytenoid muscles - see van den Berg, 1970), meet along the midsagittal plane of the larynx. Phonatory vibration, resulting from a passive interaction of mechanical and aerodynamic forces (see "Theories of syrinx function", above) follows this active (i.e., neuromuscularly effected) adduction of the vocal folds. Mechanical forces in the larynx are dependent upon the myoelastic properties of the vocal folds - i.e., their mass and stiffness as controlled by the cricothyroid muscles. The mass and stiffness of the vocal folds are functions of their variable length (van den Berg, 1970).

Smith (1954) was the first to propose a detailed model for the closure pattern of the vocal folds' medial surfaces along the midsagittal plane during phonation in the "modal" (i.e. normal speaking voice) register. This pattern, later esentially confirmed by stroboscopic photography (Baer, 1981a and 1981b), stroboscopic radiography (Saito *et al.*, 1981; Noscoe *et al.*, 1983) and ultrasonic observations (Kaneko *et al.*, 1981), can be characterized as an upward-propagating wave. (The <u>lateral</u> movement of a vibrating vocal fold was defined by Saito *et al.*, 1981, in terms of "a travelling wave which

consists of rotatory movements of individual points on the vocal fold with differentially shifted phases along the vibratory line of the fold". See also Fukuda *et al.*, 1983.) "The mechanical movements of the glottal walls take the form of a surface wave that propagates in an upward direction over a vertical length of 5-10 mm and with a velocity of the order of 1 m/sec" (Stevens, 1977). This closure pattern is reflected in the typical, non-sinusoidal waveform of modal-register speech.

In human falsetto phonation, on the other hand, in which the vocal folds are maximally extended dorso-ventrally (van den Berg, 1970), only their thin medial edges vibrate (Sonesson, 1970). This configuration of the vocal folds for falsetto voicing is the one which most resembles the configuration of the adducted psittacine syringeal membranes in Figure 2.3C.

To the human listener, all phonation by many of the larger parrots - including species of the genus *Amazona* - resembles the human falsetto register. (See the following paragraph for a cautionary note on "phonation".) This is so even if the trainer is a man with a relatively low range of fundamental voicing frequencies. Objective evidence confirms this impression: see, for example, the falsetto-like spectrum in Figure 6.5E, with its high-frequency fundamental and narrow peaks at harmonically-spaced intervals, representing the "speech" of a green-winged macaw, one of the world's largest parrots.

(It must be kept in mind that phonation - i.e., the use of a pulse train, produced by a quasi-periodic obstruction at the caudal end of the vocal tract, as the fine structure of an acoustic signal - is only one of the speech-imitation strategies available to parrots. The others, outlined under "Theories of syrinx function", above, will be dealt with in more detail in Chapters 3 and 6.)

The African grey parrot, however, which can produce phonated "speech" in the falsetto-like register typical

of other parrot species, more commonly uses a register characterized by a much lower perceived pitch and a less sinusoidal waveform[11]. It is probable that the straight cranial edges of the syringeal cartilages allow greater control of contiguous membrane surfaces during syrinx closure in *Psittacus erithacus* than is possible in other species. This would permit a vibration pattern resembling that in human modal-register voicing, as described above.

INVOLVEMENT OF THE LARYNX IN VOCALIZATION

Detailed studies of avian laryngeal morphology have been restricted to the Gallinae (White, 1975), the Corvidae (Bock, 1978), and the Columbidae (Zweers *et al.*, 1981). (See also Homberger, 1979, regarding laryngeal morphology in *Psittacus erithacus*.)

Published observations

In regard to laryngeal function, Bock (1978) wrote: "Opening and closing of the glottis may play a role during calling and singing of birds by regulating the rate of air flow through the trachea. Questions such as the speed of opening or closing of the glottis may be critical in elucidating the production of avian sounds."

Brackenbury (1978b) observed that "the ability of the glottis to act as an air valve appears to be even more pronounced than that of the syrinx". Brackenbury followed Owen (1866) in believing that the variable glottal aperture could play an important part in modulating syrinx-generated acoustic signals.

Experiments by Ruppell (1933) indicated that constriction of the glottis could produce a significant fall in tracheal resonant frequency. (See "Acoustic resonances of the vocal tract", below.) Such an effect was invoked by Klatt and Stefanski (1974) and Nottebohm (1976) in their discussions of formant transitions in avian "speech".

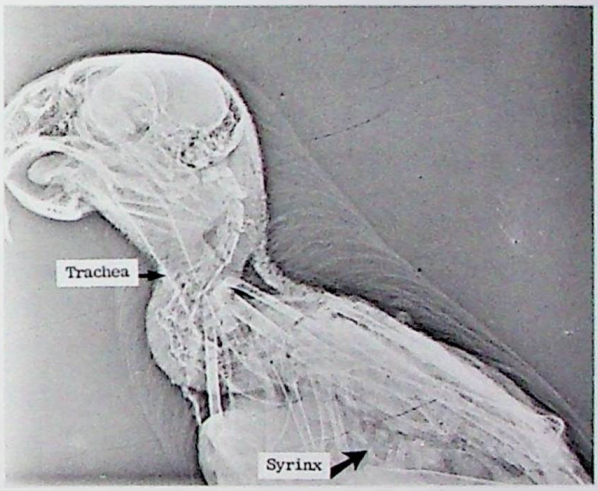
White (1968b) and White and Chubb (1968) used cineradiography to show how caudad raking movements of the larynx against the roof of the pharynx during deglutition in chickens help to propel food into the esophagus. These authors (White, 1968a, and White and Chubb, 1968) used the same technique to investigate the

PLATE 2.5

Location of the syrinx

Amazona ochrocephala

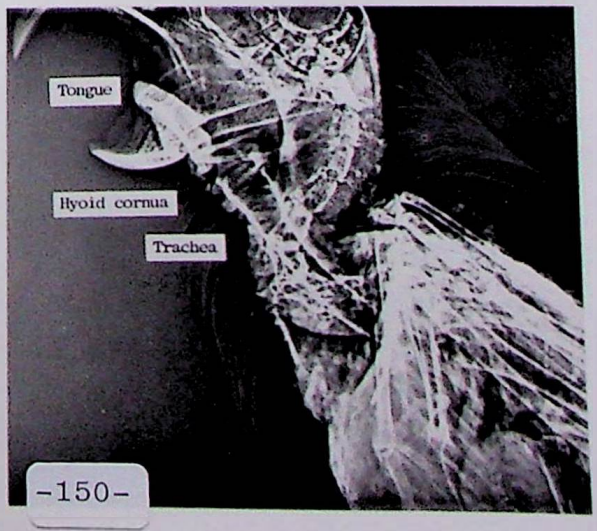
A



Suprasyringeal vocal tract

Amazona ochrocephala

B



movement of the larynx during crowing in roosters. They found that at the onset of crowing the larynx moves "several centimetres" caudad (White, 1975) to the base of the neck, where it "oscillates rostro-caudally in accordance with the phases of voice production" (White and Chubb, 1968), and inferred that larynx movement "modifies voice production by adjusting the relative lengths of the pharynx and trachea" (Ibid.). Nottebohm (1976) also believed that rostrocaudal movements of the larynx could significantly alter the acoustic resonating properties of the nasopharyngeal and buccal cavities as well as those of the tracheal tube.

White and Chubb (1968) explained rostrocaudal movements of the larynx as resulting from contraction of the cranial or caudal extrinsic laryngeal muscles[13], which both insert on to the cricoid cartilage. White (1968a) suggested that vibration of the cranial extrinsic muscle - which she calls *M. basibranchialis laryngeus*, the *M. cricohyoideus* of Vanden Berge (1979) - "could cause vibration of the adital edges of the arytenoid cartilages, and so conceivably produce, as well as modify, sound".

Although Zweers *et al.* presented a wide range of data on the movements of the pigeon's larynx derived from cinefilms, endoscopic filming, electromyography and temperature recording, they did not examine the possible involvement of such movements in vocalization. (Subsequently, Zweers, 1982a and 1982b, used cineradiography to study movements of the pigeon's larynx during the swallowing of solid food and liquids respectively.)

Zweers (1985), turning his attention to the possibility of laryngeal sound production, argues that evolutionary development of more efficient feeding mechanisms in songbirds has resulted, as a side effect, in a passerine larynx particularly well suited to the production of aerodynamic whistles. His studies have

shown that the cartilages of the passerine larynx have greater freedom of independent movement than laryngeal cartilages in other avian orders. This, he says, would allow the continuous relative adjustments necessary in such laryngeal production of glissando whistles - a common feature of passerine song. He briefly mentions an experiment, in support of his theory, which seems to have shown a correlation between changing glottis dimensions and changing frequency in a zebra finch's glissando whistle. While such a correlation need not necessarily lead to the conclusion of a glottal source for glissando whistles, the present study has produced X-ray data which add weight to Zweers's view. Some species of Psittacidae, which share the "talking" honours with their passerine cousins, are also adept at glissando whistles. The following section will present evidence of a relationship between glottis dimensions and whistle frequency in the vocalizations of a grey parrot.

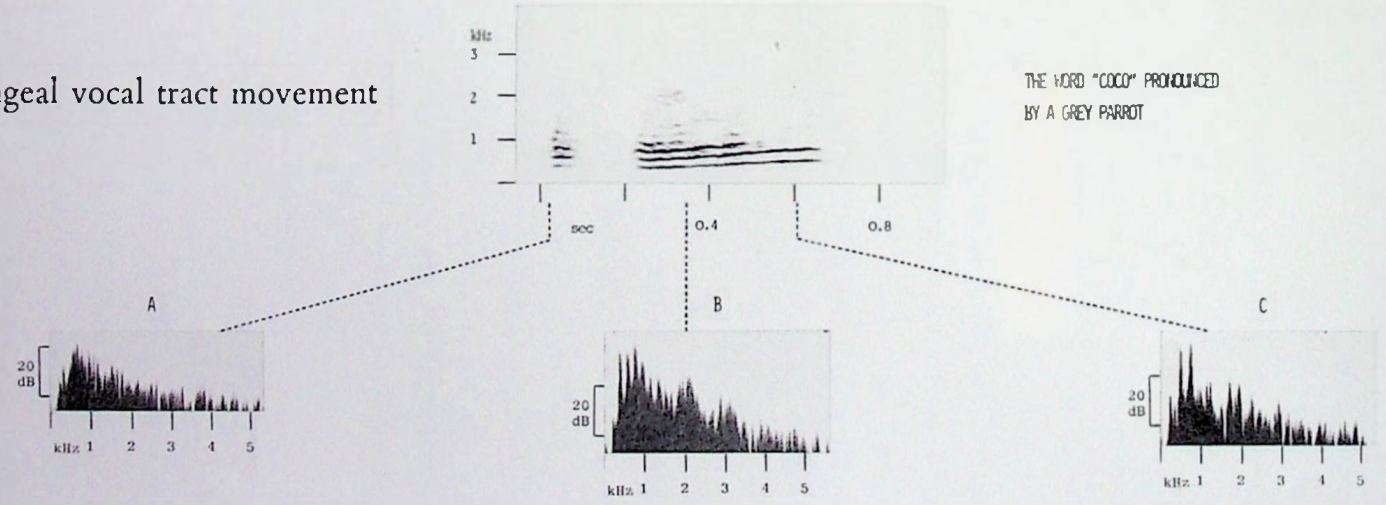
Evidence from X-ray films

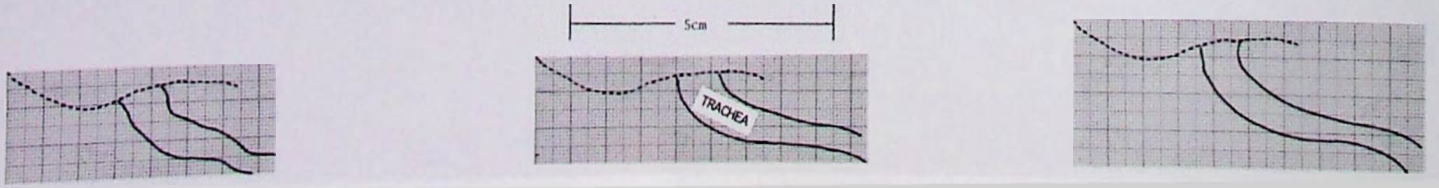
Both the films to be discussed have psittacine subjects - one an African grey parrot and the other a cross-bred Amazon. The film of the grey parrot - made by R.-G. Busnel about 20 years ago and analysed here for the first time - has an optical sound track, from which the spectral analyses on Plate 2.6 were derived. In making the film of the Amazon parrot, the screen of an oscilloscope, which was connected to a microphone (to be seen just above the bird's head in the images on Plate 2.7) was placed adjacent to the X-ray monitor screen, so that both the X-ray image of the vocalizing bird and the corresponding sound pressure waveform appear on each frame of the film. The microphone above the bird's head was also connected to a cassette recorder.

Both films concentrate on speech-like sound sequences produced by these "talking" birds.

PLATE 2.6

Suprasyringeal vocal tract movement





Cineradiography has been found to provide useful information on movements in only the upper part of the avian vocal tract - i.e., in parrots, that part cranial to the crop. Caudal to this, the image is obscured by intervening structures - including, in the lateral aspect as filmed here, the skeleton of the wing. (Location of the syrinx on the X-ray image, although possible on the Xerogram of Figure 2.5A, is not possible on cinefilm.) Thus the observations reported here are restricted to movements of the tongue, the larynx, and the upper part of the trachea.

A useful means of orientation in examining the cinefilm frames presented on Plate 2.6 is to study them in conjunction with the Xerogram of Figure 2.5B. In this figure, the hyobranchial junction of the hyoid apparatus is clearly delineated just caudal to the lower mandible at the ventro-cranial extremity of the trachea. (The structure of the hyoid apparatus can be closely examined on Figure 2.1B.) Turning now to the image under 'A' on Plate 2.6, a small dark patch can be seen in exactly the same position - i.e., just caudal to the lower mandible at the ventro-cranial extremity of the trachea. The similarly retracted position of the tongue in both these images would entail similar positioning of the hyobranchial junction. For these reasons, then, this ventro-cranial shadow on the image from the X-ray film can be regarded as representing the hyobranchial junction. (The correspondence is even more obvious when viewing the film itself.)

Immediately dorsal to the hyobranchial junction in the image under 'A' on Plate 2.6 is another small dark patch, which has no immediately obvious counterpart on Figure 2.5B. Its position in relation to the hyobranchial junction is precisely that of the cricoid cartilage of the larynx. (For a clear diagrammatic representation of the relative positions of these structures in the ventral aspect of the psittacine lingual/tracheal skeleton, see

Evans, 1969.) For this reason of anatomical orientation, and from observations of its movements (discused below), this more dorsal shadow at the cranial end of the trachea is identified with the cricoid cartilage.

Both these dark patches - but more spectacularly, the one representing the cricoid cartilage of the larynx - are in constant motion during vocalization. (In regard to *Psittacus erithacus* Homberger, 1979, reported that "the extrinsic musculture - *M. cricohyoideus dors.* and *ventr.* and *M. tracheolateralis* - moves the larynx with respect to the hyoid skeleton".) Viewing the film, it is impossible not to believe that the larynx plays an important part in sound production, as the position of the cricoid seems to be adjusted with each significant change in the acoustic spectrum.

Laryngeal movements. Gross movements of the larynx in birds are necessarily associated with movements of the tongue. (For the relationship of larynx and tongue, see Figures 2.1A and 2.1C. The psittacine tongue is uniquely mobile among avian orders, and is correspondingly unique in its possession of a truly intrinsic lingual musculature (King and McLelland, 1984; Homberger, 1986)[14]. Mivart (1895) initiated detailed study of psittacine lingual morphology with his paper on the hyoid apparatus of several parrot species, including *Psittacus erithacus*. He wrote: "So far as I have been able to ascertain, the whole order *Psittaci* is distinguished from every other order of birds by the shape of its hyoid", and continued: "That the parrots should have a tongue-bone of exceptional form is, of course, only what was to be expected from the exceptional form of their tongue as a whole". Mivart's work inspired Mudge (1903) to produce his visually beautiful and scientifically definitive work on psittacine lingual myology. More recently, Burton (1974) has remarked: "The tongue, its skeleton and its musculature are all highly modified in parrots, and

several features (e.g. the parahyal process on the basihyal to which *M. stylohyoideus* is attached) probably have no counterpart in other orders". Homberger (1986), in her work on the lingual apparatus of the grey parrot, wrote: "The total structural complexity of the psittacine tongue can be interpreted as having evolved in adaptation to the specialized diet of seeded shells".

This unique structure and musculature of the psittacine tongue allows particularly flexible laryngeal movement within the oro-pharyngeal cavity.

There are two main kinds of laryngeal movements associated with vocalization observable on the film - preparatory movements and synchronic movements. Preparatory (or pre-vocalizing) movements - involving gross movement of the lingual apparatus - transport the larynx to a vocalizing position; synchronic movements occur during vocalization itself.

It will be argued in the following paragraphs that all preparatory movements of the larynx are associated with the creation of a pressure differential across the laryngeal valve which can contribute to the rapid onset of acoustic vibration.

Ventro-caudad movement of the larynx. By far the most common preparatory movement is a ventro-caudad retraction of the tongue and larynx so that the larynx descends with the caudad extension of the pharyngeal floor. This preparatory "descent" of the larynx is similar to that observed in roosters by White (1968a) and White and Chubb (1968). Viewed laterally after this movement, as on Plate 2.6, both the larynx and the hyobranchial junction can be seen caudal to the lower mandible.

As a result of this lingual-laryngeal retraction, the cervical part of the trachea is forced ventrad into a bow-shaped configuration. The position of the trachea during quiet breathing is shown in Figure 2.5A, while its position at the end of this preparatory phase of

vocalization can be seen under 'A' on Plate 2.6. The latter figure is highly typical of the configuration of the upper vocal tract near the onset of vocalization. (Figure 2.5B does not represent a vocalizing gesture. Although the tongue is retracted, ventrad movement of the trachea is prevented by raising the head.)

Plate 2.6 presents views of the upper vocal tract at three instants during production of the word "Coco" (the name of this Gallic Polly). That under 'A' is at the very beginning of the sound; the associated amplitude spectrum, above the X-ray image, is typical of the short segment of aspiration noise immediately succeeding an unvoiced plosive. The cervical part of the trachea appears thicker and more compressed rostrocaudally (as outlined beneath the photograph) in this initial image than in the images under 'B' and 'C'. The rostral end of the retracted tongue is depressed against the floor of the buccal cavity. As described above, the cricoid cartilage and the hyobranchial junction can be seen as two similar small dark patches apparently on opposite sides (dorsal and ventral) of the cranial end of the trachea.

The images under 'B' and 'C', which represent the vocal tract configuration at two instants (separated by about a quarter of a second) during the second syllable of the word, show important changes in the position of the tongue, the relative positions of cricoid cartilage and hyobranchial junction, and the shape of the trachea.

Although the tongue is still withdrawn into the caudal end of the buccal cavity, it has moved dorsally - towards the roof of the pharynx - from its position in 'A'. This movement has been accompanied by a change in the relative positions, and the apparent size and shape, of the cricoid cartilage and the hyobranchial junction. The cricoid appears more massive than in 'A', and elongated rostrocaudally, while the position of the hyobranchial junction, in relation to the cricoid, is more cranial

than it was in 'A'. The cervical part of the trachea appears more elongated than in 'A' - possibly at least partly a result of the bird's having slightly raised its head.

(Ventrad movement of the trachea caudal to that part of it which appears in the images on Plate 2.6 is prevented, of course, by its entry into the thoracic cavity. The crop, in parrots, extends ventrally just cranial to the thoracic cavity, and could contribute to the rather darkened X-ray image of the trachea in this region: see the illustration of Evans, 1969. Cranial to the crop, the esophagus runs adjacent to the trachea. The esophagus is "thin walled and distensible", its "internal surface . . . longitudinally folded, thus increasing the distensibility of the tube" - King and McLelland, 1984. Ventrad movements of the trachea can thus stretch the esophagus dorsoventrally, accounting for the large "empty" space dorsal to the trachea in the images on Plate 2.6.)

An examination of the complete series of cinefilm frames (more than 20 in all) representing this utterance showed that the appearance of the cricoid and hyobranchial junction in 'A' - i.e., at about the time of the initial plosive - was duplicated at the beginning of the second syllable. In both instances, this appearance was retained for only two frames - about 0.08 sec - before moving towards the configuration of 'B'. It is thus reasonable to suppose that the relative positions of cricoid and hyobranchial junction in 'A', together with the gross preparatory movement of tongue and larynx, are associated with the onset of acoustic vibration. Furthermore, it is argued below that the cricoid-hyoid configuration in 'A' can be interpreted as glottis closure.

Suction-drinking analogy. Zweers (1982a) reported on, and explained in some detail, the mechanism of suction drinking in pigeons. Using cineradiographic techniques, he observed a caudad movement of the larynx just before the entry of liquid into the mouth. He explained this larynx movement, with its consequent caudal and ventral extension of the buccal air space, as functioning to reduce pressure within the buccal cavity and thus force liquid into the mouth under atmospheric pressure[15].

The caudad larynx movement described by Zweers is remarkably similar to that under examination here. This similarity prompted the question: Could the pressure-differential mechanism employed in suction drinking have an analogue in vocalization?

As the glottis is necessarily closed during drinking, its caudad movement in preparation for water-suction must be associated with glottis closure. With the glottis closed, a reduction of supraglottal pressure could produce another significant pressure difference besides the intra/extra buccal difference described by Zweers. This is a pressure difference across the laryngeal valve: a pressure difference which - considered in the context of vocalization rather than in that of drinking - would facilitate a sudden explosive release of subglottal air. Such a mechanism could well be the generator of many of the transients in bird vocalizations and, in avian "speech", of plosive consonants such as the repeated [a] in "Coco".

Descent of the larynx, together with closure of the glottis, would produce not only a decrease in supralaryngeal pressure but an increase in sublaryngeal pressure. The trachea can be significantly shortened by a caudad movement of its cranial end and, if both the syringeal and laryngeal valves were closed, the reduction of air space within the trachea would cause an increase in pressure. (This rostrocaudal compression - and thus overlapping - of the tracheal rings would explain the

uneven and somewhat distended shape of the cervical part of the trachea in the X-ray image of 'A', Plate 2.6, and as outlined below that image. As pointed out above, the tracheal outline in 'A' is peculiarly associated, in the X-ray data, with sound onset.) Thus, with the glottis closed, a reduction of intratracheal air space could contribute to an overall pressure difference across the laryngeal valve immediately prior to an explosive, sound-generating release.

Glottis closure. Zweers *et al.* (1981) found that closure of the glottis by mediolaterad movement of the arytenoids is accompanied by "rostrad" tilting of the procricoid and thus ventrocaudad tilting of the rostral end of the cricoid. ("The arytenoids, the cricoid, and the procricoid show one fixed pattern of combined movements during any opening and closing of the glottis", Zweers *et al.*, *ibid.*) This downward tilting of the cricoid "was recognized from the strongly decreasing longitudinal length of the visible part of the rostral side of the cricoid" (Zweers *et al.*, *ibid.*). The last-quoted sentence could appropriately be applied to the appearance of the cricoid in 'A' on Plate 2.6, when compared with its appearance in 'B' and 'C'.

Conversely, Zweers *et al.* observed that opening of the glottis is accompanied by an "upward tilting" of the rostral end of the cricoid. This produces, on an X-ray image, a rostrocaudal lengthening of the shadow associated with the cricoid, as it assumes a more rostrocaudal orientation. Such a lengthening of the shadow can be observed in 'B' on Plate 2.6, when compared with the corresponding shadow in 'A'. (Accompanying this upward tilting of the larynx, the visible part of the tongue has risen towards the palate.) These observations of Zweers *et al.* support the interpretation of the present X-ray data as indicating glottis closure just prior to vocalization, and rapid glottal opening

coinciding with rapid onset of sound.

Analysis of cricoid movements on the X-ray film led to the realization that glottal closure occurs not only prior to transient sounds such as clicks and "plosive consonants", but also immediately before purely periodic sounds. This means that the significance of rapid glottis opening is not limited to the production of avian plosives (attending bursts of subglottal air into the pharynx), but extends to the onset of periodic acoustic vibrations. In sounds with an undoubtedly syringeal source, such as vowel imitations lacking initial plosives, the drop in intratracheal pressure immediately following rapid opening of the glottis could play an important part in the onset of syringeal membrane vibration. The postulated mechanism at the syringeal level is, of course, identical to that at the larynx, involving a sudden drop in pressure on the cranial side of a closed valve in the airway and a consequent explosion of subglottal/subsyringeal air. In the case of the syrinx, as the valve is closed by the vibratile membranes themselves, such an explosion would cause rapid onset of membrane vibration. (Plate 2.7, to be discussed below, presents clear evidence of very rapid sound onset in vowel imitations.) Such an explanation of laryngeal involvement in the onset of complex periodic vibrations originating in the syrinx is an elaboration of Thorpe's (1959) "start-stop modulator" theory of larynx function.

Synchronized larynx movements during glissando whistles. Zweers (1985 - see "The larynx", above) believes that the larynx of songbirds is involved in the production of glissando whistles, and cites evidence of varying glottal aperture during such whistles. In the X-ray film of the grey parrot under discussion here, the slope of the shadow representing the cricoid cartilage can be seen systematically changing with changing frequency during the production of glissando whistles.

Even clearer are those sequences consisting of a series of sustained whistled notes each of a different frequency. Changes in the slope of the cricoid can be seen to coincide with the progression from one whistled note to the next. These are the first observations of the involvement of the psittacine larynx in the production of whistles.

Changes in the slope of the cricoid during whistled sequences are consistent throughout the film, the basic pattern being that an increase in slope corresponds to a fall in whistle frequency. In the light of the foregoing argument that increasing slope of the cricoid is associated with increasing glottis dimensions, it may be possible to explain whistle production in purely aerodynamic terms, in which frequency varies inversely with the size of the glottal opening. On the other hand, glottis dimensions may have more to do, here, with resonance than with sound-source.

These observations of systematic cricoid cartilage movements during whistle production in a grey parrot, and their theoretical consequences, are the subject of a forthcoming study (Scanlan, in preparation). Meanwhile, they provide another perspective on the argument that a maximally dorsoventral orientation of the cricoid cartilage - as observed in 'A' on Plate 2.6 - can be associated with larynx closure, and a maximally rostrocaudal orientation of the cartilage with maximum glottal opening.

(These simple movements of the larynx during purely whistled sequences are, of course, only part of the complex array of synchronic larynx movements involved in a more varied sequence of psittacine vocalizations.)

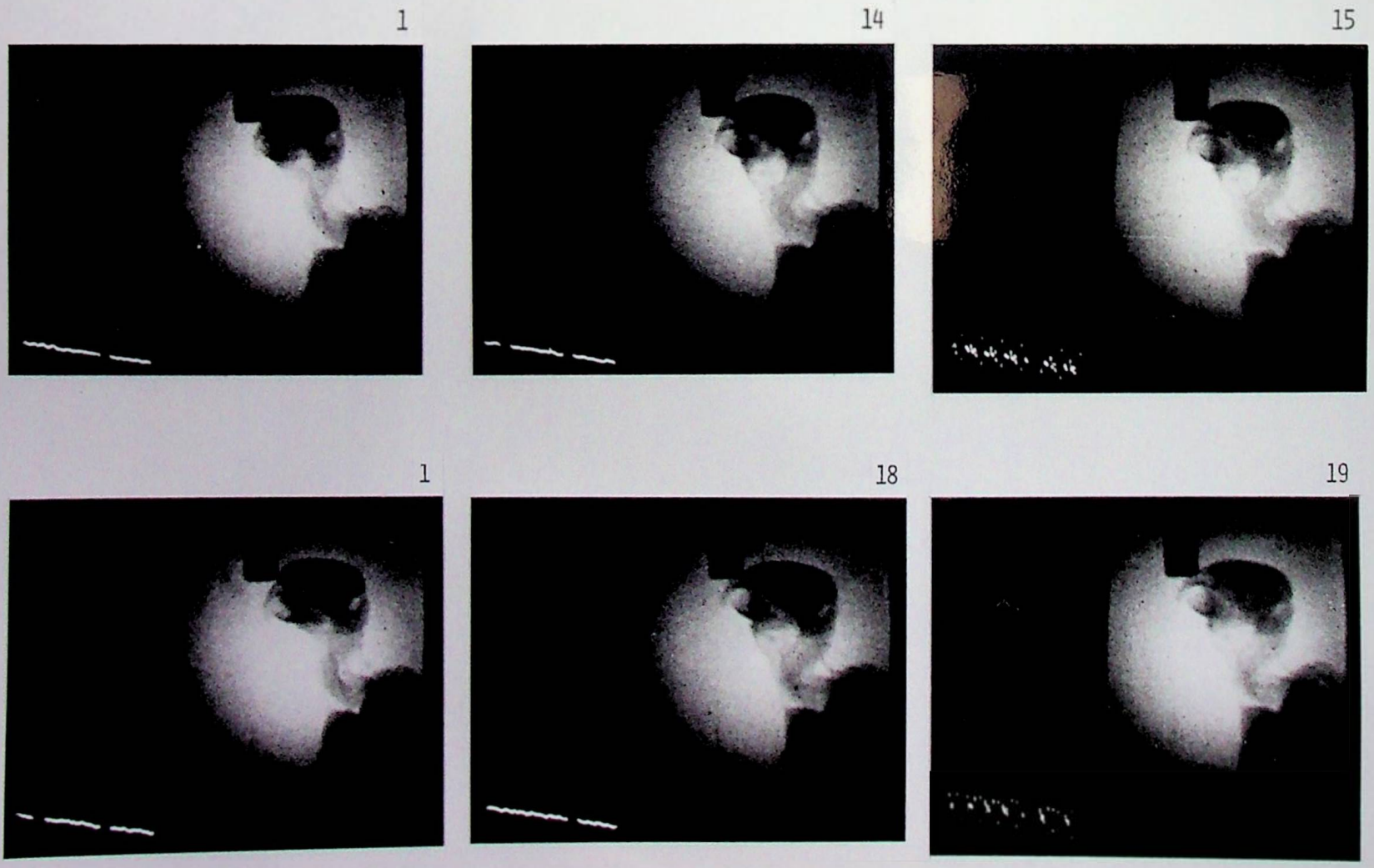
Preparatory phase in vocalization. That the primary function of ventrocaudad larynx (and ventrad trachea) movement is related to the *onset* of acoustic vibration, and that, secondarily, it assists but is not entirely necessary for the *maintenance* of vibration, is demonstrated by an analysis of the X-ray film with synchronous oscilloscope trace. Four instances of the vowel sound [a], pronounced by the bird in isolation, were examined, and it was found that, in each instance, the preparatory movement began more than half a second before the onset of sound.

Two of these instances are illustrated on Plate 2.7. In sequence A, Frame 1 marks the beginning of ventrad tracheal movement and Frame 15 marks the onset of sound. As the filming rate was 25 frames per second, the preparatory movement is shown to have begun 0.56 sec before sound onset. Correspondingly, Frames 1 and 19 in sequence B mark the beginning and end of a preparatory phase of 0.72 sec duration. (The preparatory phase durations of the other two instances examined were 0.64 sec and - by far the longest example - 1.28 sec.) In each instance, the movement sequence apparent on the film is completed in, at most, half the preparatory-phase interval, the remaining half probably being devoted to establishment of the required level of sublaryngeal air pressure by adjusting the intra-thoracic part of the vocal tract - including the syrinx - which is indiscernible here.

It is significant that, on this film, the preparatory phase is longer before the [a]s, and the larynx and trachea movements more pronounced, than before all other sounds (both avian and humanoid). This may be related to fundamental frequency, which is much lower in the [a] sounds (about 350 Hz) than in any other sound on the film. In producing a sound with such an unusually low fundamental for the bird, the greatest possible pressure difference is required to initiate the vibration of

PLATE 2.7

Suprasyringeal vocal tract movement



THE VOWEL [ɑ] PRONOUNCED BY AN AMAZON PARROT

syringeal membrane folds which have been rendered extremely massive and flaccid - by contraction of *Mm. tracheobronchialis* - in preparation for relatively low-frequency vibration. (Although this is a relatively low sound for the bird, it still has the falsetto-like quality discussed above.)

Frames 14 and 18 - of sequences A and B on Plate 2.7 respectively - are those immediately before the onset of sound (shown in Frames 15 and 19), with the observable part of the respiratory tract in full vocalizing position, but still with no sound registering on the oscilloscope trace. (The low-amplitude, aperiodic vibration apparent in the oscilloscope trace on the "silent" frames represents unavoidable background noise - mainly that of the X-ray generator and the camera. Short gaps in the trace represent "beating" of the camera and oscilloscope rates.) It can be seen that sound onset is rapid, the periodic, steady-amplitude waveforms of the oscilloscope trace on Frames 15 and 19 having been established in less than 1/25 (0.04) sec. This supports the hypothesis of rapid glottal opening at the onset of sound.

In contrast, at the end of each sound (not illustrated here), amplitude diminishes over several frames. Coincident with diminishing amplitude, the tongue, larynx and trachea return to their quiet breathing position, so that, on the last frame showing the periodic (although by now low-amplitude) oscilloscope trace of [a], the respiratory tract has all but returned to the quiet breathing position. This is so for all sounds recorded on film: the vocalizing position of larynx and trachea is never maintained - even to the slightest extent - after sound has ended.

Dorso-caudad larynx movements. Zweers (1982b) investigated the role of dorso-caudad larynx movements in the feeding behaviour of pigeons. He found that these movements can be either synchronized with a caudad moving constriction of the pharynx ("properistalsis"), or associated with the scraping action of erectable spines caudal to the larynx. (Both of these movement sequences are designed to transport food particles through the pharynx and into the esophagus.)

The former kind of dorso-caudad larynx retraction - that associated with properistalsis - as described and figured by Zweers (1982b), bears a striking similarity to that observed on the X-ray film of *Psittacus erithacus* and recognized as the second of the two main movements preparatory to vocalization. Rostrad movement of the larynx following such a retraction is accompanied by a sound - like an exaggerated human kiss - beloved of parrot owners.

In terms of the physics of its production, too - if the connexion with properistalsis is accepted - the sound must resemble a kiss. Both involve the noisy ingress of air into a small cavity - in the human kiss between the rounded lips, and in the avian "kiss" behind the dorso-caudally retracted larynx - in which a semi-vacuum has been established[15]. After retraction of the closed larynx into the dorso-caudal extremity of the pharynx, which has been constricted by properistalsis, rostrad movement of the larynx initially - before contact with the palate is entirely broken - creates a small airless space within that constriction.

(Both the human bilabial kiss and the avian pharyngeal "kiss" are closely connected with feeding gestures. The human kiss, as an expression of affection and desire, is probably derived from infant sucking gestures, in which the pressure changes involved are designed to force liquid into the mouth. See Stein, 1949.)

Dorso-caudad retraction of the larynx, on the evidence

of the X-ray film, is a much less common preparatory movement than its ventrocaudad retraction, the latter movement being fundamental to the great majority of grey parrot sounds - at least, those in the repertoire of this captive bird. On the film, dorso-caudad retraction is associated only with "kissing" sounds.

ACOUSTIC RESONANCES OF THE VOCAL TRACT

Wind-instrument and human-voice analogies

Early discussions of acoustic resonance in bird sounds were based on analogies with either the human voice or wind instruments, sometimes confusing the two. Thus Myers (1917) noted that cock crow became "noticeably higher" in perceived pitch after the trachea had been severed. He concluded from this that the trachea and larynx "serve to modulate voice, just as the pharynx and cavities in the mouth serve the same function in mammals". There is a confusion here between the effect of tube length on pitch (as in wind instruments) and on timbre (as in the human voice).

Rüppell (1933) kept this distinction clearly in mind. While he saw the function of the trachea as primarily one of modifying the shape of the source signal's spectral envelope, he recognized that in some species the frequency of the source vibration might be tuned by tracheal resonance. He demonstrated the latter effect in *in vitro* experiments using the vocal tracts of both cranes (which have very long tracheae) and chickens.

Thorpe (1959, 1961) inherited the old wind instrument/human voice models of avian vocalization (as well as the more sophisticated analysis of Rüppell), and proposed that some birds produce sounds according to the wind instrument model, while the remaining species employ a mechanism more akin to that of human speech. His criterion for voice-like sound production was the possession of intrinsic syringeal muscles.

Thorpe thought that, in birds which lack intrinsic syringeal muscles, membrane vibration must be a forced resonance, like that of an oboe reed, the frequency of vibration depending on the dimensions of the tracheal

tube. These dimensions could be modified by contraction of the extrinsic syringeal muscles. On the other hand, in birds possessing intrinsic syringeal muscles - notably the songbirds - membrane vibration, and therefore the fundamental frequency of the acoustic signal, was under direct muscular control. In these species, the adjustable dimensions of the suprasyringeal vocal tract would, like the human vocal tract, determine the sound's spectral envelope shape (associated, perceptually, with timbre) rather than its fundamental frequency (associated with pitch).

Relationship of fundamental and formant frequencies

Miller (1934), in his study of owls, demonstrated that the dimensions of the vibrating membranes are an important factor in determining fundamental vocal frequencies. The female of some owl species is larger than the male (with a proportionally longer and wider trachea), but the syrinx itself, including the vibratile membranes, is smaller. The higher fundamental frequencies in the calls of females can thus be explained in terms of membrane size, but not in terms of tracheal length. Miller found that when "the trachea was broken shortly in front of the syrinx much of the quality of the note was lost owing, I presume, to loss of resonance for the overtones as well as for the fundamental, but the pitch of the fundamental remained unchanged". His measurements led him to believe that the fundamental frequecy of an owl's "normal hoot" is inversely proportional to the length of its medial tympaniform membranes.

The results of "winding" experiments by Harris *et al.* (1968) on the vocal tracts of dead roosters are in line with Miller's findings in owls. These authors reported that changing tracheal length affected formant frequencies but not the fundamental. (They also noted that the first-formant frequency was higher when the beak

was open than when it was closed, indicating that resonance of the beak cavity, as well as of the trachea, can modify the shape of the spectral envelope.) On the other hand, changes in simulated muscle tension applied to the syrinx effected large changes in fundamental frequency. "The fundamental frequency . . . is determined exclusively by the syrinx tension", they concluded.

Hersh (1966) compared bird calls made in air with the same calls made in helium air (air in which the nitrogen has been replaced with helium). The increased velocity of sound in helium air (about one-and-a-half times greater than in air) has the same effect on resonance as a uniform reduction in the size of the resonating cavities. So, if the fundamental frequency of a call were determined by tracheal resonance, this frequency should be higher when the call is produced in helium air. Hersh, who experimented with both passerine and non-passerine species, found that fundamental frequencies were identical under the two conditions[12]. Analysis of all wide-bandwidth calls, however, revealed differences in spectral envelope shape between those calls made in air and those made in helium air.

Hersh concluded that bird sounds, like human speech, can contain formants, the frequency ranges of which are determined by acoustic resonances. There is no necessary connexion, however, between such resonances and the fundamental frequency of an utterance.

The experimental work of Nowicki (1987) confirmed Hersh's general conclusions but indicated that, while there is no <u>necessary</u> connexion between fundamental and resonant frequencies, there is often a <u>contrived</u> connexion. Nowicki recorded the songs and calls of nine species of songbird in both normal and helium air. He found that bird sounds which appeared, in amplitude/frequency and frequency/time analyses, to consist of only one component (i.e., to be quasi-

sinusoidal) exhibited a second harmonic when recorded in helium air.

He interpreted this as indicating that normally the vocal-tract resonator is adjusted so as to respond maximally to the fundamental, and that because the fundamental frequency in birdsong is typically high (2,000 - 5,000 Hz), the second harmonic lies outside the pass band of the vocal-tract filter and is thus strongly attenuated. "In helium, this filter is shifted upward, revealing the harmonic content. Since neither the fundamental frequency nor the second harmonic is now squarely centred on the filter's best frequency, both have reduced amplitudes compared to the fundamental in ordinary atmosphere" (Nowicki, *ibid.*).

Nowicki regarded these findings as evidence of "tracking" of the output fundamental of the syrinx by the vocal-tract acoustic resonator. "A bird could adjust the vocal filter in several ways: for example by varying tracheal length, by constriction of the larynx, or by flaring its throat and beak" (Nowicki, *ibid.*).

Detailed acoustic analysis by Greenewalt (1968) revealed no evidence of acoustic resonance in the whistled song of passerine birds. Greenewalt examined whistled notes at frequencies which - according to his calculations - could conceivably be reinforced by acoustic resonance, found that their amplitudes were no greater than those of notes he considered outside the influence of vocal-tract resonances, and concluded that no such resonant reinforcement occurs. Conversely, however, if considered in the light of Nowicki's (1987) "tracking" hypothesis, Greenewalt's results could be interpreted as confirming that virtually <u>all</u> notes of whistled song are reinforced by acoustic resonance.

In his chapter on "talking" birds, too, Greenewalt concentrated on passerine vocalization, studying several vowel sounds produced by a mynah and comparing them with the corresponding human sounds. He found that while the

sound pressure waveform of the human vocalization showed considerable decay in amplitude within the fundamental period - a typical effect of damped resonance - the mynah's waveform did not. This same contrast was evident in the waveforms of isolated formants, demonstrating to Greenewalt's satisfaction that the mynah "formants" studied were not the results of resonant reinforcement of source-generated spectral components.

Which avian "speech formants" are true formants?

Klatt and Stefanski (1974) believed that the first formant of mynah "speech" is, indeed, an acoustic resonance, and that the second and third formants represent mechanical resonances of one of the medial tympaniform membranes. (If such mechanical resonances were excited by the periodic striking of the membrane by the lateral labium, the amplitude of the resonant vibrations should decay within the fundamental period of the labium's oscillation in a manner analogous to that of acoustic resonance. If, however, the membrane resonances are excited by exhalant air flow, oscillation of the labium could modulate the resulting acoustic signal in a manner consistent with the data presented by Greenewalt. Klatt and Stefanski, while their argument remains unclear, seem to be postulating some sort of combination of these two mechanisms. See above, under "Theories of syrinx function".)

Klingholz (1979) argued against the proposal of Klatt and Stefanski that the first formant of mynah speech imitation represents an acoustic resonance of the trachea. His argument, which is patently fallacious, is based on the fact that a budgerigar, with its much shorter trachea, can produce a "formant" in imitated speech within the range of the mynah's first formant. On considerations of tracheal length, this "formant" of the budgerigar cannot be associated with an acoustic tube

resonance; therefore, Klingholz argues, the corresponding formant in mynah "speech" must not be associated with such a resonance. (In relation to this argument, see the comparison of a budgerigar vocal tract with that of a larger parrot in Figure 2.1A.)

This argument is based on the totally unwarranted assumption that the mechanisms of sound production in general, and speech imitation in particular, are the same for all birds. One of the conclusions to be drawn in this thesis is that sound-production mechanisms vary greatly not only among species, but even in a single individual.

Several authors (Hersh, 1966; Nottebohm, 1976; Owen 1866) have considered that the large fleshy tongue in parrots, which is unique among birds, is capable of a function similar to that of the human tongue in vocalization. Hersh included the shape of the parrot's beak in this oral anomaly, writing that these birds' "large oral cavities and eleborately muscled tongues are far removed from the usual avian pattern". Nottebohm, in his study of vocalization in *Amazona amazonica*, estimated the length of the nasopharynx, from glottis to nares, in adults of this species at 2.5 cm. "Acting as a tube closed or open at one end (glottis) and closed at the other (nares), the lowest frequencies that the nasopharynx could emphasize would be in the 3.5 to 7.0 kHz range, well within the phonatory range of *A. amazonica*" (Nottebohm, 1976). Craniad and caudad movements of the larynx, as discussed earlier in this chapter under "Evidence from X-ray films", could modify the resonating properties of both the nasopharynx and the trachea.

Calculations of possible vocal-tract resonances based on simple organ pipe models, while perhaps providing a rough guide, are open to several objections on grounds of both the physics of tube resonance, and the anatomical extent and variability of the cavities involved. Hersh (1966), in expounding Ruppell's analogy of the avian

vocal tract with a reed pipe, pointed out "the tendency [of some authors] to confuse the model system of the reed pipe with the classical organ pipes of elementary physics, and thus to lose sight of the number of variables available for manipulation". Nottebohm (1976) noted that the resonant frequencies of irregularly shaped tubes "are not simply related to the dimensions, as in a uniform pipe".

Harris *et al.* (1968) proposed a simple model of the avian vocal tract in which the trachea could be represented by a cylindrical tube, the pharynx and buccal cavities by an expansible chamber, and the laryngeal and rostral openings by variable slits. They pointed out that acoustic impedance associated with the slits could be an important factor in determining the resonant properties of such a system. Considerations of acoustic impedance, they said, explained a rise in first-formant frequency from 1.4 kHz to 1.8 kHz when a rooster's beak, during *in vitro* aeration, was opened. (They regarded this formant as the resonance of a tracheal tube open at the syrinx and closed at the glottis; this 0.4 kHz rise in frequency, then, would be equivalent to the effect of shortening the trachea by 22 per cent.)

Stevens and House (1955; see also Fant, 1960, and Lindblom and Sundberg, 1971), investigating the shape and size of the labial opening in human vowel production, stated that acoustic impedance at the lips is inversely proportional to "A/l", where "A" is the cross-sectional area of the tube-like cavity formed by the lips and the front of the mouth, and "l" is the length of that cavity. "A" is a function of labial opening, and "l" is a function of labial rounding. "A/l" is large (and therefore acoustic impedance is small) when the lips are unrounded and open wide; "A/l" is small (and impedance large) when the lips are rounded and the opening small. Stevens and House found that lip rounding and narrowing (representing an increase in acoustic impedance) resulted

in a fall in the frequencies of the first three vocalic formants.

Harris *et al.* used an analogous "A/l" ratio - that of cross-sectional area of glottis to tracheal length - in investigating the effect of variable glottal opening on acoustic resonances in bird sounds. Because the dimensions of the glottis are so small in relation to the length of the trachea, they concluded that the size of the glottal opening has little or no effect on formant frequencies. The spectrographic evidence they present in support of this conclusion, however, is ambiguous to say the least. Ruppell (1933), on the contrary, found in a similar but much better documented experiment that laryngeal constriction in the vocal tract of a crane (*Grus grus*) lowered the tracheal resonance as much as would a simple tubular extension several centimetres long. Hersh (1966), Klatt and Stefanski (1974) and Nottebohm (1976) followed Ruppell in believing that constriction of the larynx can cause a significant fall in formant frequency.

Several authors have presented arguments for or against the presence of vocal-tract resonances in bird sounds based on a comparison of audio-spectral peaks with hypothetical resonance peaks calculated from measurements of tracheal length. While such measurements can be useful in assessing the possibility of acoustic resonance within a range of frequencies, that range must be great enough to allow not only for complex nonlinearities in the bio-acoustic resonating system, but also for possible differences between the length of the trachea when measured and its length during vocalization. Hersh (1966) makes this latter point in discussing the study by Sutherland and McChesney (1965) of flight calls in two species of geese. Hersh writes: "There is . . . no reason to believe that the pickled tracheae from which the measurements were taken were preserved at exactly the length used by the bird in flight".

Among those authors who have ventured to use simple organ pipe models in investigating tracheal resonance, there has been disagreement about both the effective extent of the resonating tube, and whether its ends are to be considered "open" or "closed". Harris *et al.* (1968) found that the resonant frequency of a rooster's subglottal vocal tract (as revealed on spectrograms of sounds produced during *in vitro* aeration) corresponded to calculations from vocal tract measurements only if bronchial length was added to the length of the trachea. Although these authors are the only ones to have reported such a finding, acoustic coupling of subglottal and supraglottal cavities during the open phase of voicing is a factor known to affect resonant frequencies in human speech (Fant, 1982). It is possible that artificial "winding" of the avian vocal tract could induce a membrane vibratory cycle with an unnaturally long open phase, thus coupling subsyringeal cavities to the tracheal tube and producing a correspondingly low formant frequency.

Nottebohm (1976) wrote that "the average cross-sectional area of the syringeal passage is much less than the cross-sectional area of the trachea, and to this extent the tracheal tube is essentially closed at the syringeal end". On the other hand, Sutherland and McChesney (1966) and Harris *et al.* (1968) believed that the syringeal end of the tube could be considered open. As for the laryngeal end of the tube, Greenewalt (1968) as well as Sutherland and McChesney regarded it as "open", while Harris *et al.* (for reasons they do not give) regarded it as "closed". Nottebohm, much more reasonably, thought of the larynx as being able to assume configurations during sound production which could render it effectively either open or closed. Thus all four possible states of a tube with an open/closed valve at either end have been variously advocated in the literature as useful models of the suprasyringeal vocal

tract.

Sutherland and McChesney (1966) studied spectrograms of goose calls recorded in the wild, and concluded that the spectral components, which, they said, "may be treated as if they were harmonically related", represented source-generated harmonic spectra. They then compared the frequencies of those components which appeared to occur near peaks in the transfer function of the vocal tract with hypothetical resonant frequencies calculated from measurements of preserved tracheae. They found that these apparent formants in the spectral pattern approximated more closely to resonances calculated for an "open" tube (i.e., a tube open at both ends) than for a "closed" tube (i.e., closed at one end). Resorting to the time-honoured expedient of musical-instrument analogy, they suggested that the goose's vocal tract functions more like a trumpet than like the reed organ pipe of Ruppell.

Greenewalt (1968) calculated possible resonant frequencies of a mynah's tracheal tube (under both "open" and "closed" conditions) using a measurement of tracheal length (6.4 cm) supplied to him by an avian anatomist (Ames). He compared these possible resonances with the "formant" frequencies of a mynah's vowel reproduction, and concluded that the vocalic "formants" were unrelated to tracheal resonance.

Greenewalt arrived at this conclusion in spite of the fact that the second-formant frequency he quotes for the mynah's vowel [a], and the first resonance of the "closed" tube which he adduces as one of his model resonators, coincide exactly - both being 1,400 Hz. Greenewalt seems here to succumb to the fallacy of believing that, as acoustic resonance cannot explain the entire formant pattern of avian "speech", it therefore cannot provide an explanation for any part of that pattern.

Klatt and Stefanski (1974) attributed the first

formant of mynah "speech" to tracheal resonance. This can be challenged on several grounds, the first being simply that the tracheal length they quote - 9.5 cm - is too long. I have determined the tracheal length of an adult specimen of *Gracula religiosa* at approximately 7 cm, measured from the glottis to the cranial edge of the pessulus. (See Fig. 2.1C.) Even if the resonating tube is considered to extend caudally as far as the medial tympaniform membranes it is no longer than about 7.5 cm. Thus, even allowing for discrepancies due to preservation techniques, and the bird's ability to extend the length of the trachea during vocalization, the 9.5 cm trachea reported by Klatt and Stefanski seems too long. The length quoted by Greenewalt - 6.4 cm - is even further from their figure than is my own.

Assuming their figure to be correct, however, the lowest possible resonance of a 9.5 cm tube (about 870 Hz for the "closed" condition) is above the first-formant range (380-840 Hz) they quote in their analysis of mynah "speech". This, then, is the second objection to their association of F1 with a tracheal resonance. On the other hand, although this 870 Hz tube resonance is precisely at the bottom of the second-formant range quoted by Klatt and Stefanski (870-2,660 Hz), shortening the tube to 7 cm would shift the resonance to well within this range.

(These calculations of resonant frequency involve simply dividing the figure for the velocity of sound in air - 332 m/sec according to Helmholtz, 1885, p. 90 - by four times the length of a "closed" tube or twice the length of an "open" tube. The lowest resonance of an "open" tube is therefore twice as high as the lowest resonance of a "closed" tube.)

Thus, despite their own protestations to the contrary, data supplied by both Greenewalt and Klatt and Stefanski indicate the possibility that the second formant of mynah "speech" could be associated with the first tube resonance of a trachea having the acoustic properties of

a tube open at one end and closed at the other.

Chapters 3 and 6 will indicate that the second formant of avian "speech" is typically associated with a resonant membrane vibration.

This is in line with the findings of Nowicki (1987), mentioned above, in that the frequency range of human F2 (700 Hz - 3,500 Hz) roughly coincides with the range of fundamental frequencies in much whistled birdsong. (This correspondence will be discussed in detail in Chapter 4.) A resonance system designed to "track" these relatively high-frequency fundamentals could therefore be employed - in speech imitation - in manipulating a spectral peak in the region of human F2.

In the generation of whistled song, as discussed above under "Theories of syrinx function", interaction between an aerodynamic source perturbation and resonant vibration of the medial tympaniform membranes has been proposed (Gaunt *et al.*, 1982). Reinforcement by acoustic resonance in the suprasyringeal vocal tract would thus result in additional amplification of a source vibration already amplified by mechanical resonance within the syrinx. Such double reinforcement could help to explain the impressive loudness of much birdsong.

Chapter 6 will indicate, however, that these mechanisms, which operate "in series" in the production of whistle-like sounds with a single spectral-envelope peak, could also operate "in parallel" in the production of sounds with multiple spectral-envelope peaks - such as imitations of human speech.

SUMMARY AND CONCLUSIONS

The foregoing description - and theoretical and experimental interpretation - of avian vocal anatomy was undertaken as an essential preliminary to studying the acoustic phenomena of avian "speech". It has revealed - above all else - the flexibility of the avian vocal system (at least in songbirds and parrots).

The main sound source - the syrinx - is a valve which, under complex muscular control, is capable of assuming (in a single individual) a variety of configurations in constricting the tracheo-bronchial airway for the transformation of mechanical energy into acoustic energy. One of these syringeal constriction mechanisms in parrots - illustrated here for the first time - resembles that involved in human phonation.

Muscle-adjusted syringeal membranes are capable of sound-generating vibratory response to pressure events near the airway constriction (and/or sound-amplifying response to aerodynamic perturbation near the constriction). New evidence from X-ray films of "talking" birds indicates that air-pressure events at the laryngeal valve may also play an important part in avian vocalization. The hypothesized laryngeal mechanism could be the source of plosive-consonant sounds in avian speech imitations.

New X-ray data indicating laryngeal control of whistle frequencies (possibly involving the actual generation of aerodynamic whistles within the larynx) is discussed.

Syringeal (and, perhaps, laryngeal) elements can be involved in modulating vibrations as well as source vibrations. (The "double" syrinx of songbirds allows for independent source/modulation vibrations in each syringeal half.) Modulation can contribute a human-voice-like timbre to bird sounds.

Phonatory pulse trains, multiple source vibrations, and amplitude and frequency modulation all result in complex acoustic patterns. "Talking" birds are able to adjust these patterns, to a certain exent, to resemble those of human speech.

These complex patterns may be rendered even more speech-like by the imposition of a variable acoustic-resonance transfer function.

NOTES TO CHAPTER 2

[1]In an attempt to achieve a consistent terminology throughout this work, the nomenclature advocated by King (1979) has been adopted in an anglicized form. Thus "lateral tympaniform membrane" (*Membrana tympaniformis lateralis*) has been preferred to the more common "external tympaniform membrane". This use of "lateral" and "medial" is in line with the practice of Gaunt (1983), Gaunt and Gaunt (1985) and Gaunt *et al.* (1982). "Tympaniform" has been preferred to these authors' "tympanic", however, partly in sympathy with Ames's (1971) argument that this adjective is derived from considerations of morphology rather than of function. (See text.)

[2]Gaunt et al (1973) were the first to use the terms "active closure model" and "passive closure model" in referring to theories in which vocalization is initiated by the excursion of a syringeal element into the lumen under active muscular control, or in which such occlusion occurs passively as the result of Bernoulli forces within the syrinx. See also Brackenbury (1978a).

[3]"In order to fulfill the role of pressure equilibration, the interclavicular air sac must be able to relay internal pressure changes instantly to the syrinx; this short circuit is provided by the direct connexion of the interclavicular air sac to one of the ventrobronchi" (Brackenbury, 1982).

[4]"An intrinsic syringeal muscle should have all its attachments on the syrinx. Unequivocal intrinsic syringeal muscles appear to be quite rare, even in

Passeriformes . . ." (King, 1979). Miskimen (1951), in her inter-species comparisons of syringeal musculature, described one muscle (which she called *M. broncho- trachealis brevis*, after Owen, 1866) as "internal" in some species and "external" in others. In birds with a complex syringeal musculature, *M. tracheolateralis* is not, strictly speaking, a syringeal muscle because it divides into the "intrinsic" muscles cranial to the syrinx. Such difficulties and anomalies could be simply eliminated by using the membrane forming the clavicular sac as a demarcation line. Thus muscles having all attachments on or within the encircling sac could be referred to as "intrinsic syringeal muscles", while those having an attachment outside the clavicular sac would be "extrinsic syringeal muscles". (See Garrod, 1874, for a beautiful description of the way in which the sac membrane is actually involved in the origination of some of the "intrinsic" muscles.) A possible alternative to this descriptive (anatomical) definition is the functional (physiological) definition of Gaunt (1983). "I consider a muscle to be intrinsic if it can affect the syringeal configuration by some means other than adjusting the position of the caudal end of the trachea" (Gaunt, *ibid.*).

[5]Myers (1917), contrary to all subsequent accounts, reported that contraction of *Mm. sternotrachealis* induced by electrical stimulation of the muscles in an anaesthetized chicken shortened the trachea "from one-fourth to one-third of its original length" and increased tension in the lateral tympaniform membranes.

[6]So called by Gaunt *et al.* (1976).

[7]This name for the muscle is derived by analogy from the

system of nomenclature advocated by King (1979).

[9]Warner (1969) refers to this element as "the last tracheal ring". This is the example of confusion over "tracheal" and "bronchial" labels cited in the Introduction to this chapter. Gaunt and Gaunt (1985), while recognizing the uniqueness of this element, opted for a bronchial affiliation. I have avoided this unnecessary dilemma.

[9]Ruppell also presented a caudal view of the other half of the syrinx - i.e., the tracheal lumen, as seen through the retracted cranial halves of the lateral tympaniform membranes.

[10]Paulsen also published frames from a film showing the dorsal aspect of the same syrinx during aeration. The vibrating medial tympaniform membranes can be seen periodically moving towards each other in the inter-bronchial space.

[11]A good example of this low-frequency voicing ability in grey parrots compared with phonation in the Amazons can be heard on a commercial recording available in France, *La belle histoire de Jacotte et Ito*. Jacotte, an African grey, and Ito, a blue-fronted Amazon, were trained by the same man, M. Raoul Ours. Ito's "speech" is falsetto-like, while Jacotte's is characterized by low fundamental frequencies and a rather harsh, pulsatile quality. A photograph of the birds on the record sleeve shows that they are about the same size.

[12]Hersh did report "a very weak coupling of source and filter" in the California quail (*Lophortyx californicus*).

[13]"During descent the rostral extrinsic muscles must

relax, while the aryteno-glossal ligament which runs from the tips of the rostral arytenoid prongs to the glossal bone must be drawn taut. This elastic ligament, which is better developed in the male than in the female, may be homologous with the vocal cords of mammals" (White, 1968a).

[14]For a detailed description of lingual myology in parrots, see Mudge (1902). Burton (1974) drew attention to a unique range of similarities between the tongue musculature of parrots and that of the tooth-billed pigeon (*Didunculus strigirostris*). He presented these similarities in support of the view that the Columbiformes are the order most closely related to the Psittaciformes.

[15]Homberger (1983) outlined four different drinking methods which have evolved in parrots "as byproducts of evolutionary changes due to feeding specializations involving the tongue, bill, and palate". Some parrots employ a suction mechanism (Homberger, 1980). The drinking method employed by the parrot under scrutiny here is probably one in which "water is ladled with the spoon-shaped tip of the tongue and swallowed by pushing the tongue against the palate" (Homberger, 1986).

[16]A much closer analogy to the avian "kiss", in fact, is the human palatal click. See Figure 5.3H for a psittacine version of this sound. It is possible that such a sound could be produced by the bird in a manner similar to the human palatal click - i.e., simply by release of contact between tongue/larynx and palate, without accompanying properistalsis of the pharynx.

CHAPTER 3:
VOCAL REDUNDANCY

INTRODUCTION

Redundancy and variability are important features of the vocal communication systems of birds and man, occurring at each of the levels of production, transmission and perception. They provide a key to understanding the ability of some birds to participate in the milieu of human speech.

REDUNDANCY IN SOUND PRODUCTION

According to Gaunt *et al.* (1976), "the avian vocal system is both plastic and redundant. This affords considerable potential for variation both among species and at the whim of the individual."

Apparent anatomical redundancy

It has been shown that normal vocalizations can often be produced after inactivation of one or more pairs of syringeal muscles. Gross (1964) and Brackenbury (1978) found that sectioning the sternotrachealis muscle in *Gallus gallus* did not alter ability to produce normal vocalizations. Smith (1977) made the same observation in *Columba livia* and three oscine species: *Agelaius phoeniceus*, *Zonotrichia albicollis* and *Zonotrichia melodia*. Youngren *et al.* (1974) reported that inactivation of both the tracheolateralis and sternotrachealis muscles made little difference to the vocalizations of *Gallus gallus*. Gaunt and Gaunt (1977), using electromyographic techniques, observed occasional inactivity of *M. cleidotrachealis* as well as *M. sternotrachealis* during calls of *Gallus*.

Complexity, redundancy and plasticity

Despite a general coincidence of complex syringeal musculature and spectrally complex vocalizations in birds (discussed under "Syringeal morphology" in Chapter 1), several authors (e.g., Beddard, 1898) have commented on specific anomalies in the correlation of myological and acoustic complexity. Thorpe (1961) remarked that "some birds which have a very simple syringeal musculature manage to produce surprisingly complex sounds, and others

which have a most elaborate musculature produce little in the way of song". (Miskimen, 1951, however, believed that there was a "direct correlation between the degree of development of the syrinx, with respect to muscles and attachments, and the quality - variety of notes - of the song".)

The argument presented in Chapter 1 may help to clarify this issue by pointing out that the essential connexion is not one between intrinsic syringeal muscles and acoustic complexity *per se*, but between those muscles and the vocal plasticity necessary for learning (regardless of the relative complexity of the sounds learnt).

"Vocal redundancy" refers simply to a phenomenon of sound-production physiology, while "vocal plasticity" refers to a behaviour pattern involving auditory (or proprioceptive) feedback. In evolutionary terms, therefore, a capacity for redundancy in the avian sound-production system could have provided a basis for the development of audition-based plasticiy and thus vocal learning.

There is evidence to suggest that, in at least some birds which lack intrinsic syringeal muscles, and whose vocal development does not appear to involve learning, there is a capacity for redundancy in the vocalizing function of the tracheal muscles.

For example, Gaunt and Gaunt (1977) reported that there were at least two ways in which a chicken (*Gallus gallus*) could produce "wailing" sounds. In the first method, the lateral membranes are folded into a valve which opens periodically in response to increases in subsyringeal air pressure. This mechanism, discussed in some detail in Chapter 2, is similar to that employed in mammalian phonation. In the second method, described by Gaunt and Gaunt (*ibid.*) as following "the classic model for syringeal action in producing a loud sound", air flows continuously between the two oscillating membranes.

Data collected for the present study revealed evidence of an analogous redundancy in the speech imitation of all four *Melopsittacus undulatus* subjects - particularly in imitations of the speech sound [ɔ]. (The following discussion concentrates on analysis and interpretation of this particular *Melopsittacus* "speech" sound. See the Conclusion to the chapter for a discussion of why this sound is so susceptible to variations in production technique.)

Whereas vocal redundancy in chickens is a function of the tracheal muscles, in budgerigars (which possess only one pair of tracheal muscles - Evans, 1969), the intrinsic syringeal muscles must be involved.

Evidence of redundancy in speech imitation leaves no doubt that, in budgerigars, such redundancy is integrated into vocal-learning behaviour patterns.

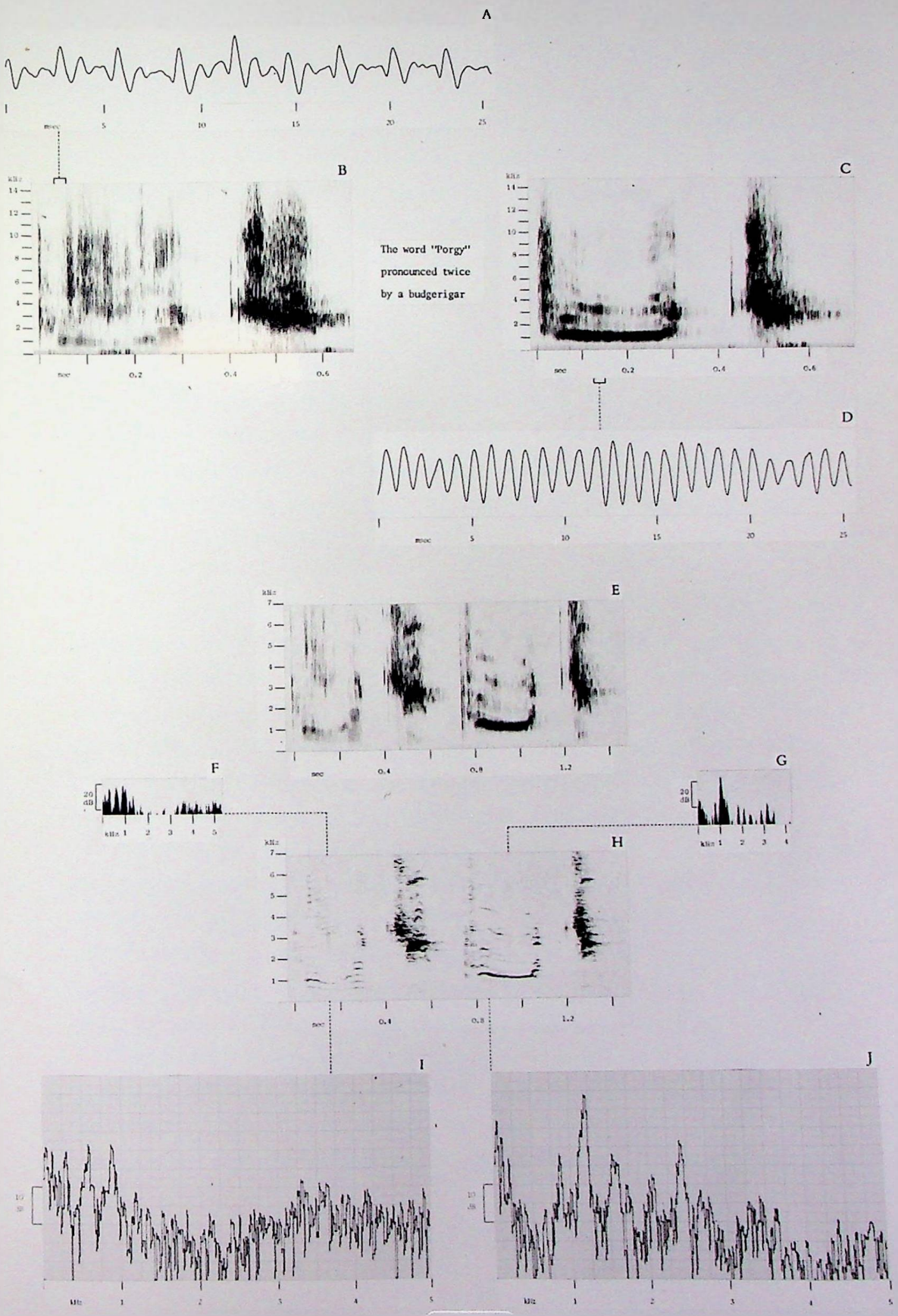
The budgerigar's [ɔ].

Plate 3.1 indicates that a budgerigar can produce imitations of the vowel sound [ɔ] using different vocal mechanisms.

In five of the ten recorded instances of the phrase "Georgy Porgy" [dʒɔdʒi pɔdʒi] by one of the budgerigar subjects, the word [pɔdʒi] is repeated immediately after its first occurrence. The resulting [pɔdʒi pɔdʒi], in all five instances, follows the pattern exemplified in Figures 3.1E (300 Hz bandwidth filter analysis) and 3.1H (45 Hz bandwidth filter analysis of the same sequence). The first version of [pɔdʒi] in the other four instances closely resembles that illustrated here (Fig. 3.1B). (Figure 3.1B represents the same sound as the first half of Figure 3.1E, analysed at half the speed.) There is a similarly close resemblance among instances of the second [pɔdʒi] - that illustrated here (Fig. 3.1C) and the other four. (Figure 3.1C represents the same sound as the second half of Figure 3.1E, analysed at half the speed.

PLATE 3.1

Two ways of imitating one word



Acoustic differences. There are obviously great acoustic differences between the first and second versions of the word, as represented in Figures 3.1B and 3.1C. The greatest differences are in the reproduction of the [ɔ] sound, which occupies most of the initial 0.3 sec in each of the abovementioned figures. The following discussion will, therefore, concentrate on this sound.

The very different acoustic properties of these two realisations of the same "phoneme" indicate that a different production mechanism is employed for each. Looking at the relevant section of Figure 3.1C, and comparing it with Figure 3.1B, its most obvious distinguishing features are its much narrower overall bandwidth and the greater intensity of its most prominent spectral peak (Fig. 3.1J). These are the acoustic characteristics discussed in Chapter 1 as often distinguishing an active production mechanism from a passive mechanism.

Acoustic similarities. Underlying these differences, however, are common acoustic features which make it possible to recognize both these sounds, in their speech-imitation context, as realizations of the same "phoneme". These common features are a spectral peak at around 1 kHz, and a frequency difference of about 300 Hz between several pairs of major spectral components.

As a working hypothesis (in the light of the apparent "active/passive" distinction mentioned in the previous section) it would be reasonable to suppose that the redundancy in question consists in an alternation of active and passive mechanisms in the production of at least one of these common acoustic features.

Spectral components of avian acoustic signals can be related to two basic kinds of membrane vibrations: (1) resonant vibrations, and (2) muscle-assisted vibrations. The capacity of syringeal membranes to be simultaneously

involved in both kinds of vibration contributes to the acoustic complexity of many avian vocalizations.

Muscle-assisted vibrations themselves can be divided into two kinds: (1) actively generated, by pulsatile myological contraction (also referred to as "myogenic" vibrations), and (2) passively generated by air-pressure and tissue-elastic forces after muscular contraction has achieved a vocalizing configuration (also referred to as "myoelastic" vibrations). (For discussion of the distinction between "active" and "passive" sound production, see "Syringeal morphology" in Chapter 1.)

Thus, our working hypothesis requires the identification of an acoustic feature, common to both [ɔ] sounds, which can be identified in one instance with a myoelastic vibration and in the other instance with a myogenic vibration.

Sound-production hypotheses: general principles

Production of the first [ɔ]. Gaunt and Gaunt (1985) asserted, and experimental manipulation of excised syringes (Chapter 2) has indicated, that contraction of *Mm. sternotrachialis* can, in at least some parrot species (including *Melopsittacus undulatus*), force the lateral tympaniform membranes to meet along the midsagittal plane of the tracheal lumen, thus completely occluding the airway. The means by which such a closure could be repeated periodically in a "phonatory" manner (i.e. according to a myoelastic mechanism) are reviewed in Chapter 2.

The EMG data of Gaunt and Gaunt (1985) showed that activity of *Mm. sternotrachialis* in *Melopsittacus* was synchronous with - or slightly preceded - sound production. This satisfied them that "the sternotrachialis functions as an activating muscle" (Gaunt and Gaunt, *ibid.*). As this muscle sometimes fires continuously during call sequences, it was evident to

Gaunt and Gaunt that "its pattern of activity is of only slight importance once sound has been started".

Also, during many *Melopsittacus* vocalizations, "the extrinsic and intrinsic muscles act in parallel" (Gaunt and Gaunt, *ibid.*). This means that the intrinsic muscles merely contribute to what Gaunt and Gaunt call "simple activation patterns": *Mm. tracheobronchialis* by assisting - and adding precision to - the mediad positioning of the membranes, and *Mm. syringealis* by controlling membrane tension. (See Chapter 2, under "Function of the syringeal muscles".)

The [ɔ] sound in the first [pɔdʒi], with a fundamental of 300 Hz and strong second and third harmonics (Fig 3.1I), has a speech-like waveform (Fig. 3.1A) consistent with such a phonatory mechanism - i.e., with a quasi-periodic occlusion of the airway by the membranes due to air-pressure and tissue-elastic forces. The vertical striations on the spectrogram (Fig. 3.1B) reinforce this picture of voice-like pulses produced in the syrinx. The emphasis of these striations in the 3-10 kHz region could indicate the excitation of acoustic resonances within this frequency range by pulses generated near the caudal end of the vocal tract.

With an effective tube length of about 3 cm (Fig. 2.1A), the lowest possible acoustic resonance frequency for the vocal tract of *Melopsittacus undulatus* would be about 2,500 Hz. (See Chapter 2, under "Acoustic resonances of the vocal tract".) Therefore, the formant-like band of energy centred on 900 Hz (Fig. 3.1B) could not be the product of <u>acoustic</u> resonance. It could, however, represent a <u>mechanical</u> resonant response of the membranes to the 300 Hz pulse, as proposed in Chapter 2 under "Theories of syrinx function". Thus the spectral component at 900 Hz (see particularly Figure 3.1F) would represent that harmonic of the forced (300 Hz) fundamental which is closest to the resonant frequency of the membrane.

It will emerge, in the course of this and subsequent chapters, that a frequency component of between about 1,000 Hz and 3,000 Hz often constitutes the major spectral peak of budgerigar "speech" sounds. This is probably the range of resonant frequencies, under variable tension, of the lateral tympaniform membranes.

The pattern, here, of 300 Hz fundamental and 900 Hz resonance is clearly depicted in the waveform (Fig. 3.1A) and the amplitude spectrum (Fig. 3.1I). (The similarity of the waveform to one shaped by acoustic resonances, as in human speech, is an indication of the success of the bird's aural illusion.) These two spectral components represent simultaneous muscle-assisted and resonant vibrations of the syringeal membranes.

Thus the first stage in the investigation has been reached - i.e., the possible identification of an acoustic feature, common to both [ɔ] sounds, which can be identified in the first sound with a passive muscle-assisted (myoelastic) vibration of the syringeal membranes. This feature is the frequency difference of about 300 Hz between several pairs of major spectral components. It remains now to test the hypothesis that, in turning to the second [ɔ] sound, this feature could relate to an active muscle-assisted (myogenic) vibration of the membranes.

Production of the second [ɔ]. Analysis of this sound reveals patterns which resemble those associated with amplitude modulation of a quasi-sinusoidal carrier. Figure 3.1J shows a dominant frequency component of 1,200 Hz placed between two other components which are 350 Hz above and below it. (The prominent second harmonic of the 1,200 Hz fundamental also has sidebands 350 Hz on either side of it.) The absence of an harmonic relationship between these three components, and the absence of a component at 350 Hz, are further indications that this pattern represents a carrier frequency with sidebands.

The smooth waveform envelope (Fig. 3.1D), and the relative weakness of high-frequency excitation throughout much of the second [ɔ]'s duration (Fig. 3.1C), indicate that, unlike the first [ɔ] sound, it does not involve the pulsatile excitation of mechanical and/or acoustic resonances. The structure of the spectrum, as described in the previous paragraph, rules out the possibility that the smooth waveform envelope is related to the falsetto-like phonation common in the "speech" of larger parrots, as such phonation gives rise to a purely harmonic spectrum (e.g., Fig. 6.5E). Thus the evidence suggests that, in this second [ɔ] sound, periodic complete occlusion of the syringeal airway (i.e. phonation) does not occur.

(Apart from the absence of pulsatile excitation, continuous opening of the syringeal valve would couple the suprasyringeal cavities with those below the syrinx, thus tending to reduce still further any possible occurrence of high-frequency acoustic resonance.)

There are several reasons for believing that the modulating vibration of 350 Hz could not be a membrane resonance. Firstly, the frequency is simply too low; secondly, there is no spectral component at 350 Hz. If this acoustic feature does not represent a membrane resonance, it must - according to the dichotomy proposed above - represent some form of muscle-assisted vibration.

Now, muscle-assisted vibrations of the myoelastic variety, relying as they do on an interaction of air-pressure and tissue-elastic forces, generally involve periodic complete occlusion of an air passage and a consequent periodic pressure increase below the closure. As such a closure apparently does not occur during generation of the second [ɔ] sound, the possibility that the vibration in question is myogenic - i.e., actively generated by an input of periodic neuromuscular pulses - must be considered. This view is corroborated by EMG data (Gaunt and Gaunt, 1985), which show periodic pulsatile

activity of the intrinsic syringeal muscles during vocalization in budgerigars.

Thus acoustic evidence supports the hypothesis that psittacine vocal redundancy - at least in the instance under examination - could involve an alternation of active and passive muscle-assisted vibrations in generating an acoustic feature which is common to both sounds.

If, as has been argued, the 350 Hz modulating frequency in the second [ɔ] is related to a muscle-assisted vibration, it follows that the other major spectral feature - the 1,200 Hz carrier - is most probably related to the other main kind of syringeal vibration - membrane resonance. The 1,200 Hz frequency is within the range of membrane resonant frequencies proposed above.

According to this interpretation, an overall symmetry between the production of the two sounds can be discerned: each comprises a resonant membrane vibration of about 1 kHz and a muscle-assisted membrane vibration of about 300 Hz.

Redundancy would consist in the employment of alternative physiological manoeuvres for the excitation of these vibrations. The next section will examine how such an alternation could most simply be effected.

Hypothetical model for the production of both sounds

The first [ɔ]. Initially, contraction of *Mm. sternotrachialis*, assisted by contraction of *Mm. tracheobronchialis*, causes adduction of the lateral tympaniform membranes. These make contact along the midsagittal plane of the tracheal lumen, thus closing the airway (Plate 2.3). Phonation - i.e., periodic reopening of this syringeal valve - ensues, resulting from an interaction of air-pressure and tissue-elastic forces, as in human phonation (Gaunt and Gaunt, 1985). Membrane

tension is controlled by *Mm. syringealis*.

The 900 Hz "formant" in the first [ɔ] is a forced resonant response of the membranes to that component of the pulsed (i.e., harmonically structured) excitation which is nearest to the membranes' natural resonant frequency. Higher-frequency spectral components represent acoustic resonant responses of the suprasyringeal vocal tract to the excitation pulses.

The second [ɔ]: myogenic vibration. Onset of the second [ɔ] is preceded, like that of the first, by contraction of *Mm. sternotrachialis* and *tracheobronchialis*, drawing the tympanum caudad and forcing the lateral membranes into the tracheal lumen. In this case, however, complete adduction of the membranes during sound production does not occur.

The prevention of periodic complete adduction of the membranes during production of this second sound is effected by the antagonistic activity of *Mm. syringealis*. (See Chapter 2 under "Function of syringeal muscles".)

It was argued above that, because the acoustic evidence indicates uninterrupted opening of the syringeal airway, the 350 Hz vibration contributing to the second [ɔ] sound could be myogenic rather than myoelastic. The Gaunts' (1985) EMG data show that such pulsatile activity is typical of the intrinsic muscles, but not of *Mm. sternotrachealis*. Their data show simultaneous neuromuscular pulses in both *Mm. tracheobronchialis* and *Mm. syringealis* during calls of several parrot species.

By such synchronised activity the opposing function of *Mm. syringealis* could prevent the membranes - when actively adducted by contraction of *Mm. tracheobronchialis* - from completely occluding the syringeal lumen. (The data of Gaunt and Gaunt are for neuromuscular pulses of not more than about 50 Hz in frequency. As explained in Chapter 2, muscular activity at frequencies considerably higher than this is probably

a resonant response to neural pulses below an upper limit of about 60 Hz.)

An alternative to the above model is simply that tension applied by contraction of *Mm. syringealis* is maintained throughout the vocalization, thus preventing complete adduction of the oscillating membranes. This would be analogous to the function of *Mm. tracheolateralis* in the non-phonatory mechanism observed in *Gallus* by Gaunt and Gaunt (1977). In reality (and this is supported by an inspection of the Gaunts' data) there are probably elements of both continuously-maintained tension and antagonistic pulsatile activity involved in preventing mutual contact of the oscillating membranes.

Thus the 350 Hz syringeal vibration involved in production of the second [ɔ] can be described as a periodic, quasi-symmetrical excursion of the opposing lateral tympaniform membranes, controlled by periodic activity of *Mm. tracheobronchialis*, towards the midline of the syringeal lumen. Complete occlusion of the airway is prevented by activity of *Mm. syringealis*, which act in opposition to *Mm. tracheobronchialis*. The function of this vibration as a modulator will be discussed below.

The second [ɔ]: resonant vibration. Turning now to the component of around 1 kHz: while, in the first [ɔ], this is interpreted as that harmonic of the fundamental pulse frequency which is nearest to the resonant frequency of the membranes, in the second [ɔ], in the absence of a fundamental pulse, it must result from a different form of resonant excitation.

Gaunt and Wells (1973) recognised that, in the absence of "phonatory" contact, membrane resonance could be excited by a periodic "pressure perturbation" in the exhalant airstream. (See "Theories of syrinx function" in Chapter 2.)

Gaunt *et al.* (1982) regarded this "pressure perturbation" as itself an acoustic vibration: an

aerodynamic whistle produced in the syrinx. They argued that this aerodynamic perturbation (a vortex street) could be adjusted in frequency and reinforced in amplitude by the natural resonance of syringeal membranes.

(The considerable difference in amplitude - about 12 dB - between the resonance-related spectral peaks of the two [ɔ] sounds, with that of the second sound being the more intense, suggests that the peak in the second sound is closer to the membranes' resonant frequency than its counterpart in the first sound. This observation supports the view that, in the second sound, the excitation frequency is influenced by feedback from the resonant response.)

Lord Rayleigh (1878), in speculating on the aerodynamic mechanism involved in the artificial "bird-call", or two-hole tone whistle, concluded that "the elucidation of this and many kindred phenomena remains still to be effected". Although some advance has been made in the theory of vortex shedding (see, e.g., Blevins, 1977; Chanaud, 1970; Morse and Ingard, 1968) and its application to such simple systems as the "Rayleigh bird-call" (Wilson *et al.*, 1971), its application to biological systems such as that involved in avian vocalization "remains [in Lord Rayleigh's words] still to be effected".

Nevertheless, some speculations have been advanced in the ornithological literature. The discussion of Gaunt *et al.* (1982) is concerned with sound production in the ring dove (*Streptopelia risoria*), which possesses medial as well as lateral tympaniform membranes. Their proposal is that resonant vibration of the medial membranes contributes to the source vibration, while muscle-assisted vibration of the lateral membranes imposes an AM envelope on the source-generated waveform. They envisage the aerodynamic perturbation as originating near the cranial ends of the bronchi, adjacent to the medial

membranes.

Parrots do not possess medial tympaniform membranes, so any resonant membrane response to an aerodynamic perturbation must be a function of the lateral membranes. Nottebohm (1976) suggested just such a mechanism in his discussion of sound production in the orange-winged Amazon (*Amazona amazonica*). He believed the source of the aerodynamic perturbation to be "turbulence past the lips of the first bronchial half-ring[s]" - i.e., in the terminology advocated in Chapter 2, the cranial edges of the syringeal cartilages.

As explained in Chapter 2, contraction of *Mm. tracheobronchialis* causes medio-caudad tilting - and thus adduction - of those edges. This produces the aperture considered by Nottebohm (*ibid.*) to be, under certain circumstances, "the main or sole sound source". (The U-shaped outline of the cartilage edges in budgerigars and some other parrots, as described in Chapter 2, could mean that, in such species, the aperture would have the sound-producing characteristics of an orifice rather than those of a slit.)

Simultaneous contraction of *Mm. syringealis* would produce a tautness in the lateral membranes, which extend craniad from the syringeal cartilages. The cartilage edges would then be sharply defined in their sound-production configuration, unobstructed by slack membrane folds. (Note that the resonant vibration in the second [ɔ] sound is of considerbly higher frequency than that in the first [ɔ]. This could indicate increased membrane tension.)

According to this hypothetical model, the antagonistic functions of the intrinsic muscles, which control the muscle-assisted, modulating vibration in the second [ɔ] sound, could at the same time control the resonant vibration.

Realising that aerodynamic and membrane vibrations are mutually dependent components of such a system, Gaunt *et*

al. (*ibid.*) concluded that "in extreme cases, designation of 'source' [i.e., as *either* membrane *or* aerodynamic vibration] becomes an exercise in sophistry".

Thus, although the nature and generation of the "pressure perturbation" in such a system remain unclear in detail, for the purposes of the present discussion the sound source can validly be regarded simply as a membrane resonance.

The second [ɔ]: amplitude modulation. According to the hypothetical model, the quasi-sinusoidal wave associated with this source vibration is amplitude modulated by periodic fluctuations in the cross-sectional area of the tracheal lumen between the lateral membranes. This is effected by periodic adduction of the membranes at a frequency of about 350 Hz, as described above. The membranes, in this case, could function as mufflers, or attenuators, as proposed by Gaunt *et al.* (1982). This would impose an AM envelope on the 1 kHz carrier - peaks in the AM envelope corresponding to periodic maximum abduction of the vibrating membranes.

As available evidence (quoted in Chapter 2 under "Theories of syrinx function") indicates that neuromuscular pulses at frequencies as high as 300 Hz cannot occur, such muscular activity must represent a resonant oscillation of the tracheobronchialis (and, probably, the syringealis) muscles in response to lower-frequency neuromuscular pulses.

Thus, according to the hypothetical model, the main difference, in physiological terms, between production of the two [ɔ] sounds is pulsatile activity of the intrinsic syringeal muscles in the second. The antagonistic functions of *Mm. tracheobronchialis* and *Mm. syringealis* prevent phonation, and allow aerodynamic excitation of resonant vibration in the lateral tympaniform membranes. Lower-frequency, muscle-assisted vibration of these

membranes which, in the first sound, is related to a phonatory mechanism, is transformed, by activity of the intrinsic muscles, into a muffling (amplitude modulating) mechanism. The frequency of oscillation in the first sound, which is controlled by an interaction of air-pressure and tissue-elastic forces, is reproduced in the second by a resonant response of the muscles to periodic neural pulses.

The excitation of a forced membrane resonance, as a component of the first sound, could prompt the bird to repeat the word, employing a more efficient[2] mechanism to produce a central component of similar frequency but far greater amplitude. This would then be an example of true vocal redundancy, as understood by Gaunt *et al.* (1976) and Gaunt and Gaunt (1977), the second sound being used simply as a substitute for the first.

An inter-species comparison

Scheich *et al.* (1977) described the acoustic structure of the Guinea fowl's (*Numida meleagris*) "kecker" call as an "harmonic spectrum with a fundamental frequency between 200 and 350 Hz". The fowl's "iambus" call, on the other hand, consisted of an "harmonic spectrum with a fundamental frequency around 1 kHz (range 900-1,300 Hz)" (Scheich *et al.*, *ibid.*). Further examination of "iambus" calls revealed that "the harmonics typically showed symmetrical high and low frequency sidebands at a distance of 200 to 500 Hz. From sonagrams and from the analog waveform this phenomenon was interpreted in terms of an amplitude modulation of the 1 kHz mechanism by another frequency generating mechanism beween 200 and 500 Hz" (Scheich *et al.*, *ibid.*). The authors noted the similar frequencies of the fundamental in "kecker" calls and the modulating frequency in "iambus" calls.

Maier (1982) described some "iambus" calls of the Guinea fowl as having an harmonic spectrum with a

fundamental of 330 Hz. More frequently however she found that, as in the analyses of Scheich *et al.*, "iambus" calls consisted of a carrier frequency at 1 kHz and sidebands - which she interpreted as resulting from amplitude modulation - 330 Hz above and below this carrier. This is, therefore, an example of vocal redundancy - the use of two different mechanisms to produce functionally identical calls.

The correspondence of low-frequency fundamental in one version of the call, and modulation frequency in the alternative version (both 330 Hz), is even more precise here than in the comparison of "kecker" and "iambus" calls by Scheich *et al.* (1977).

These comparisons within the Guinea fowl call repertoire have been adduced because of their striking similarity to the comparison of alternative versions of the speech sound [ɔ] - as produced by a budgerigar - figured on Plate 3.1.

As in the Guinea fowl sounds, the modulation frequency in the second [ɔ] is roughly equivalent to the low fundamental frequency of the first [ɔ]. Further, in the second version of [ɔ] the major component which, in the preceding version, was interpreted as a resonant response of the membrane corresponding to the third harmonic of a low-frequency (300 Hz) fundamental, has been replaced by a carrier of around 1 kHz (Figs 3.1F and 3.1G).

It should be kept in mind here that the Guinea fowl belongs to the same order - Galliformes - as the domestic fowl. The discussion in Chapter 1 under "Vocal plasticity and neural complexity" suggested that vocal redundancy implies a level of neural complexity and abstraction which may be adequate for some form of vocal learning. The fact that vocal redundancy has been observed in these two galliform species may be of significance in the light of several reports of vocal learning within the order Galliformes - see Chapter 1.

These species may be able to control tracheal and

abdominal muscles in such a way as to produce simple modulation patterns. On the other hand, precise control of complex modulation patterns in the budgerigar, as discussed in the following sections, provides evidence of intrinsic muscular control in this species.

THE PERCEPTION OF MODULATED SOUNDS

Modulation in bird sounds

Modulation frequencies around 300 Hz, such as those reported in the Guinea fowl, occur in the vocalizations of many avian species, and may reflect common properties of the neuromuscular and/or sound-perception mechanisms in birds. (See below, this chapter, and Chapter 4.) Greenewalt (1968) showed that modulation - both AM and FM - is very common in bird sounds. He observed, for example, that "for the song sparrow modulating frequencies appear to cluster about an upper limit of 300 c/s". Stein (1968) found that modulating frequencies in the songs of 12 species of songbird varied between 90 Hz and 300 Hz.

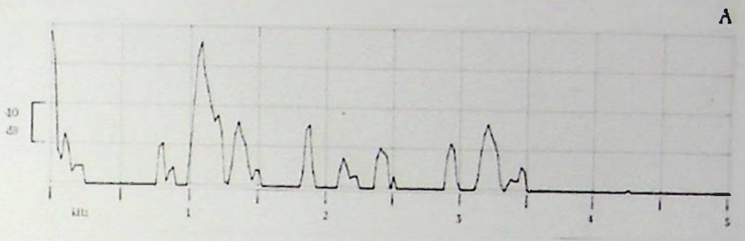
"Harmonic" relationships

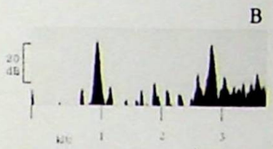
In a "phonatory" method of sound production such as that proposed for the first [ɔ] sound, the excited membrane resonance must correspond in frequency with an harmonic of the fundamental pulse rate. In the method employed in the second [ɔ], however, such an harmonic relationship need not apply between the frequency of the source vibration (the carrier) and the frequency of the membranes' oscillation (the modulator). It will be reported below (under "Anharmonic relationship") that, in the majority of such AM-structured sounds analysed, there is indeed no integral relationship between carrier and modulation frequencies.

Such a relationship is, however, possible. A spectrum (Fig. 3.2A) averaged over a 40 msec interval of the steady-state [ɔ] section in Figure 3.1C, reveals a quasi-harmonic relationship between carrier and modulation

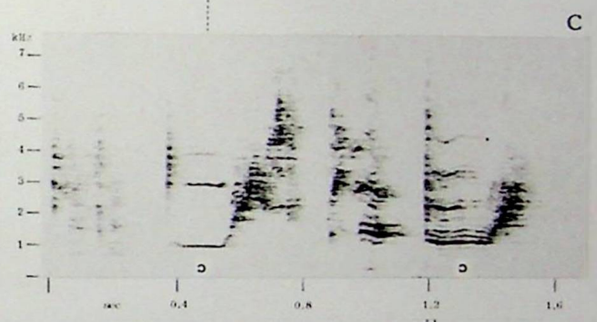
PLATE 3.2

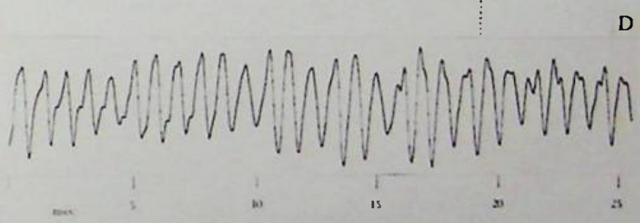
The budgerigar's [ɔ]



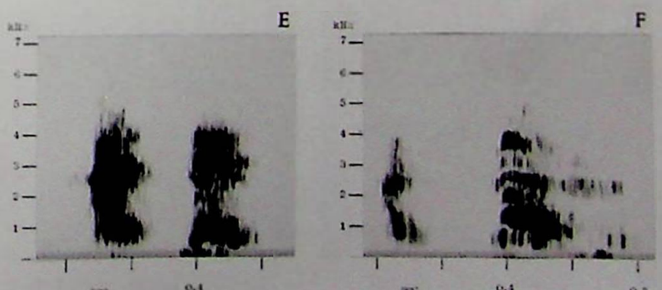


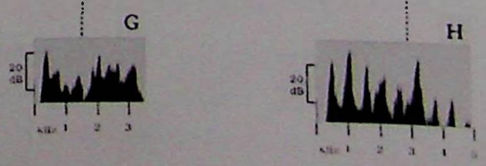
Two ways of imitating one word





A versatile raven's 'Hello'





frequencies. The spectral pattern is typical of amplitude modulation. The carrier of 1,070 Hz has sidebands about 270 Hz below and above it; the two harmonics of the carrier have sidebands at similar intervals above and below them. There is no spectral component at the modulation frequency. Unlike the spectrum - from near the beginning of the sound - in Figure 3.1J, however, the relationship between carrier frequency and modulation frequency is virtually an integral one[3]. Thus the carrier occupies the same position on the spectrum as would the fourth harmonic of a 267 Hz fundamental, while the average distance between any two closely adjacent components in the spectrum is 271 Hz.

So, looking at this second [ɔ] sound on Plate 3.1, it appears that after a period of about an eighth of a second during which the carrier frequency is changing (see Fig. 3.1H) and there is no harmonic relationship between it and the modulation frequency, there is a steady-state section of about a tenth of a second during which carrier and modulation frequencies are stable and appear to be harmonically related. This apparent harmonic spectrum could contribute to the illusion being created by the bird, as it resembles the spectra of vocalic speech sounds.

Perception of low pitch in "harmonic complexes"

Ritsma (1962) investigated the perception of three-component complex sounds in which the components could be regarded as adjacent higher harmonics of a missing fundamental. These "harmonic complexes" were produced by amplitude modulating a sinusoidal carrier. Ritsma established the boundaries within which, for stimuli of certain modulation depths, the perceived pitch of such a complex sound corresponds to the frequency of the "missing fundamental" (i.e., the frequency difference between two adjacent components). He found that, for a

modulation depth of 100 per cent, this "tonal residue" is heard when the centre frequency is lower than 5,000 Hz, and the spacing frequency greater than one-twentieth of the centre frequency. The "existence region" of the tonal residue (or low pitch) diminishes as modulation depth is reduced.

These experimental stimuli of Ritsma resemble, in both the absolute frequencies of their components and the apparently harmonic relationships of those components, sounds such as that illustrated in Figure 3.2A. (The main difference, of course, is the presence in the natural sound of the second and third harmonics of the carrier and their surrounding sidebands, produced by slight irregularities in the syringeal sound-production mechanism.)

In order to assess, aurally, the validity of the sound-production model being presented, a three-component complex was generated digitally by harmonic synthesis of three sinusoidal components (900 Hz; 1,200 Hz; 1,500 Hz) in cosine phase (Fig. 3.3E). The amplitude of the central component (1,200 Hz) was 6 dB greater than the amplitude of the other two components, thus creating a spectral pattern identical with that for amplitude modulation at 300 Hz of a 1,200 Hz carrier, with 100 per cent modulation depth. The low pitch and the timbre of this synthetic sound were similar to those of the budgerigar sound analysed in Figure 3.2A.

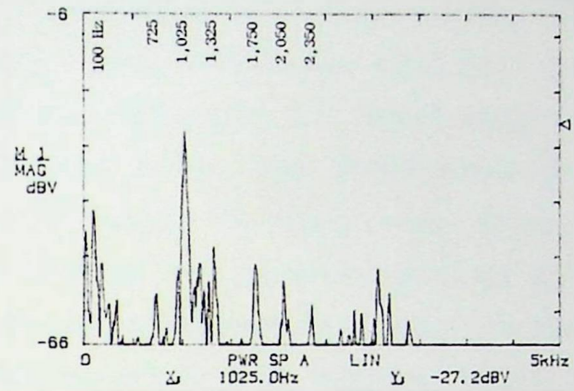
Anharmonic relationship

Analysis of other tokens of [ɔ] produced by the four budgerigars recorded shows, however, that this apparent harmonic relationship among the spectral components is the exception rather than the rule. Plate 3.3 presents spectra of four more examples of [ɔ] produced by the subject of Plate 3.1. Spectral peaks are more clearly defined in Figures 3.3A and 3.3C than in Figures 3.3B and

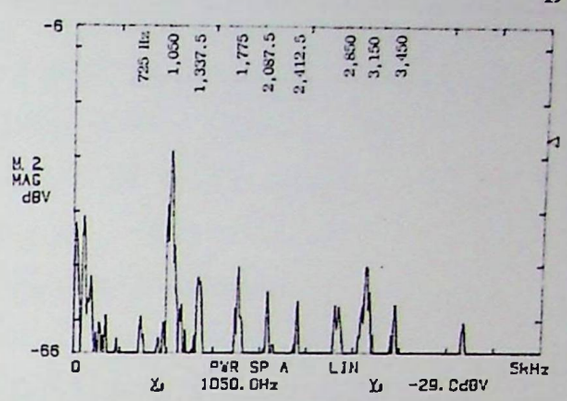
PLATE 3.3

The budgerigar's [ɔ]

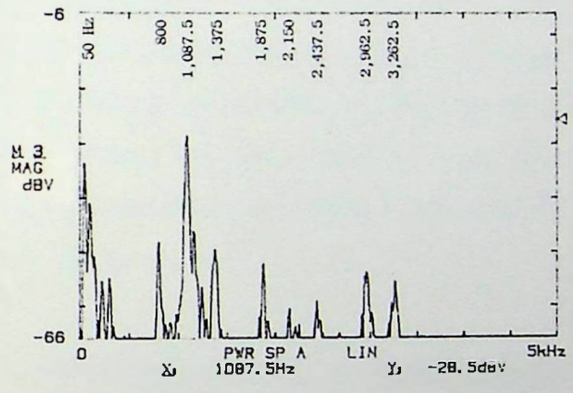
A



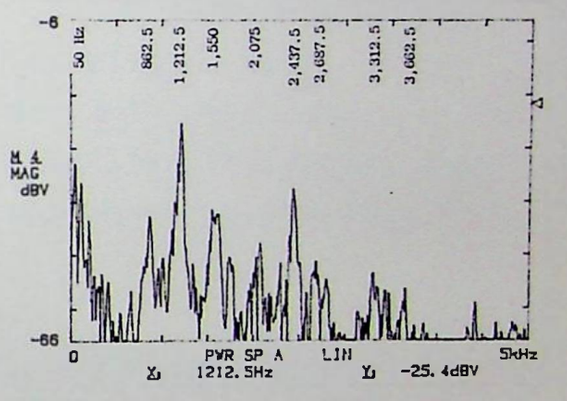
B



C



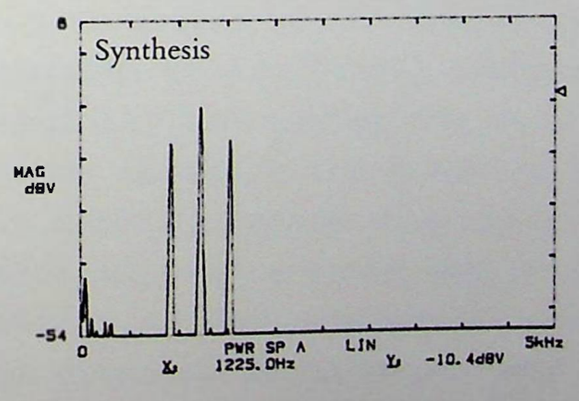
D



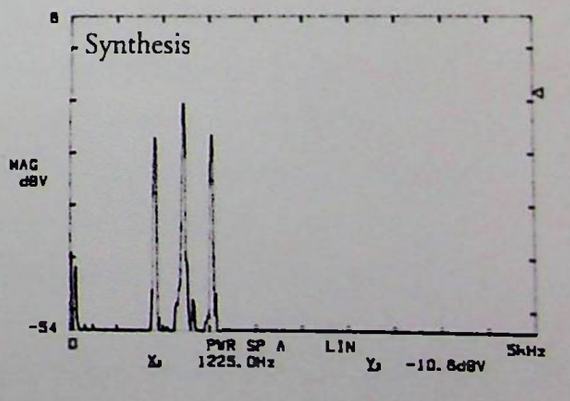
Amplitude modulation

Quasi-frequency modulation

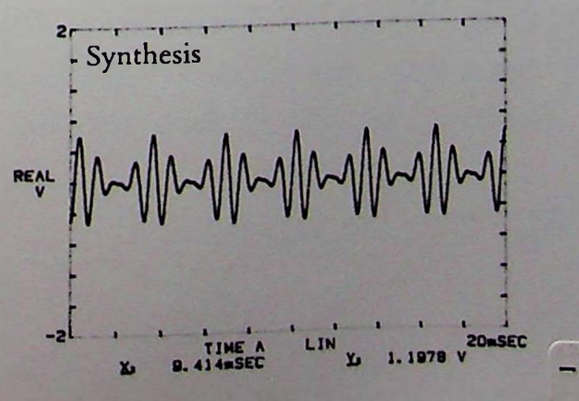
E



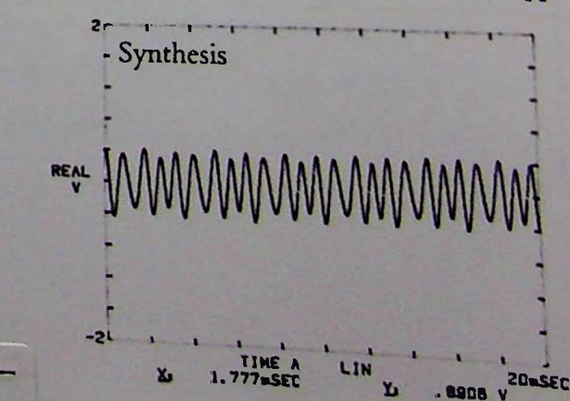
F



G



H



3.3D. Carrier and modulation frequencies in Figure 3.3A are 1,025 Hz and 300 Hz respectively, and in Figure 3.3C 1,087.5 Hz and 287.5 Hz, and so in neither case are harmonically related.

(Neither are they harmonically related in Figures 3.3B and 3.3D, although carrier and modulation frequencies are more difficult to determine in these two figures. In Figure 3.3B, and even more so in Figure 3.3D, slow modulation of vibration frequency has resulted in patterns of closely-spaced and sometimes almost indistinguishable peaks in the vicinity of the major components. See Figure 6.5B - a synthetic frequency-modulated signal - and note that the central component - the carrier - is not the component of highest amplitude. This demonstrates that the highest-amplitude peak frequency quoted on Figures 3.3B and 3.3D need not necessarily be considered the true carrier frequency. For a discussion of more rapid FM in the budgerigar's [ɔ], see below.)

Perception of low pitch in anharmonic complexes

Perceptual experiments have shown that equally-spaced components of a complex sound do not have to be harmonically related to create an impression of low pitch corresponding to the frequency spacing of the components. Schouten *et al.* (1962) extended Ritsma's work[4] by examining the perception of three-component complex sounds in which there was no harmonic relationship among the equally-spaced components. They found that the low pitch of such a sound was, once again, a function of the spacing of the components, and that it varied from the low pitch of an harmonic complex in the same frequency region according to laws which they termed "the first and second effects of pitch shift". Centre frequencies at which the low pitch was most clearly audible were between 1,200 Hz and 2,400 Hz - which overlaps the range of

centre frequencies (1,025-1,212 Hz) in analysed tokens of the budgerigar's [ɔ]. Such anharmonic complex sounds resemble, spectrally, the majority of budgerigar [ɔ] imitations analysed. (Once again, the main difference is the presence in the bird sounds of the second and third harmonics of the carrier, together with their associated sidebands.)

Frequency modulation

Avian acoustic modulation patterns are more complex than the above discussion might suggest. So far, only amplitude modulation has been discussed. Most modulated bird sounds, however, contain elements of both amplitude and frequency modulation (Greenewalt, 1968; Stein, 1968). See Chapter 2 under "Theories of syrinx function" for a full discussion of the phenomenon of AM-FM coupling, and attempts in the literature to explain this phenomenon.

To recapitulate briefly: there is a large body of data, collected by several reserchers, which demonstrates a relationship - sometimes direct and sometimes inverse - between frequency changes and amplitude changes in bird sounds. Greenewalt (1968) propounded an ingenious but (as argued in Chapter 2) fundamentally flawed mechanical explanation of this relationship, based on hypothetical changes in syrinx configuration. Gaunt and Wells (1973) suggested an alternative explanation which saw the relationship simply as reflecting the response curve (i.e., amplitude plotted as a function of frequency) of the membranes considered as resonators. While this latter hypothesis clearly represents a valuable insight into the mechanism involved, it cannot explain such rapid coupled amplitude-frequency changes as those occurring within modulation periods as short as 0.03 sec in many bird sounds.

A possible FM mechanism

Although Gaunt and his colleagues (Gaunt *et al.*, 1982) went on to develop the theory of an aerodynamic source vibration, and suggested that the sinusoidal wave associated with this vibration could be amplitude modulated by the lateral membranes acting as mufflers, they apparently did not consider the possibility of a connexion between this model and AM-FM coupling.

Experiments by Wilson *et al.* (1971) showed that the frequency of a hole-tone whistle varies directly with orifice diameter. If, then, the aperture between the cranial edges of the syringeal cartilages functions as a sound-producing orifice, as suggested by Nottebohm (1976), abduction of these cartilage edges, and thus of the attached membranes, could be associated with a rise in output frequency. Maximum membrane abduction, and therefore greater volume velocity of air flow (van den Berg, 1957) through the syrinx, would thus correspond not only to a peak in the amplitude-modulation envelope of the waveform, but to a minimum wavelength in the frequency-modulated carrier. In other words, within the modulation period, frequency rises as amplitude rises and falls as amplitude falls[5].

This possible source of AM-FM coupling is consistent with the stricture laid down by Greenewalt (1968) and endorsed by Gaunt and Wells (1973) in the light of the available data, that both amplitude and frequency modulation must be under the control of a single mechanism. The mechanism proposed here is based on the antagonistic activity of *Mm. tracheobronchialis* and *syringealis*[6].

"Primary" and "secondary" modulations

Gaunt *et al.* (1982) distinguished between two distinct kinds of amplitude modulation in bird sounds. One kind involves rapid vibrations, such as those of around 300 Hz which have been the main subject of this chapter. These vibration frequencies are far too high to be associated with fluctuations in subsyringeal pressure, and must be generated in (or cranial to) the syrinx. The second kind of AM is much slower, and has been shown experimentally to be directly linked to variations in air sac pressure (Brackenbury, 1977a, 1977b, 1979; Gaunt and Gaunt, 1977; Gaunt *et al.*, 1973, 1976; Gaunt and Wells, 1973; Suthers and Hector, 1982). Variations in air sac pressure are, in turn, directly related to activity of the abdominal muscles (Gaunt *et al.*, 1982).

Calder (1970) referred to these two kinds of modulation as "primary modulations" (syringeal mechanism) and "secondary modulations" (thoracic mechanism). He observed "secondary modulations" of up to 25 per sec in canary song, each cycle of which corresponded to an individual "mini-breath" in the respiration pattern. (Brackenbury, 1977, correlated rates of audible pulses - about 3 or 4 per sec - in the vocalizations of chickens and geese with the measured rates of panting in these species.)

Gaunt and Wells (1973) believed that "secondary modulations" in amplitude would be linked, as in "primary modulations", to variations in frequency. Although there is some experimental evidence to support this view (Gaunt *et al.*, 1982), the coupling is not as obvious as in "primary modulations", and the mechanism remains unclear (Gaunt, 1983). The slow frequency modulations noted above, particularly in relation to the closely-spaced spectral peaks in Figures 3.3B and 3.3D, are probably related to periodic changes in subsyringeal pressure.

Combinations of AM and FM

("Modulation" here refers only to "primary modulations" - i.e., in the sounds under examination, quasi-periodic variations in amplitude/frequency at rates of between about 150 Hz and 350 Hz. This excludes from the present discussion the much slower "secondary modulations" (Calder, 1970) which are evident on the amplitude spectra, and which were discussed immediately above.)

Although the amplitude spectra of sounds such as the second [ɔ] analysed on Plate 3.1 are dominated by patterns characteristic of AM (e.g. Fig. 3.2A), the waveforms of these sounds reveal the presence of FM. Its presence can be seen both in periodic wavelength variations, and in the flattening of the envelope of the waveform (Fig. 3.1D). In order to demonstrate this effect, the waveform of the synthetic AM sound described above (Fig. 3.3G) was compared with that of a sound with almost identical spectrum (Fig. 3.3F), but with an apparent element of FM (quasi-frequency modulation, or QFM) introduced by phase-shifting the carrier by 90 degrees. The waveform of the QFM sound (Fig. 3.3H), in contrast to that of the AM sound (Fig. 3.3G), exhibits the periodic variations in wavelength and the flattened envelope characteristic of the budgerigar's [ɔ] (Fig. 3.1D). (Figure 3.2D shows the waveform of an [ɔ] sound with a greater frequency modulation range than in Figure 3.1D, and in this case the periodic frequency changes can be more clearly seen.)

The two synthetic sounds - amplitude modulated, and quasi-frequency modulated - differed from each other perceptibly in timbre, the AM sound being rougher (i.e., more like a pulse train) than the QFM sound. Mathes and Miller (1947), who used such AM and QFM stimuli in their

investigation of phase perception, found that, at a sensation level of 60 dB, the two became indistinguishable when the modulation frequency was more than about 40 per cent of the carrier frequency. Data published by Goldstein (1967), who extended the work of Mathes and Miller, confirm their results, but establish the critical modulation frequency at 30 per cent rather than 40 per cent of the carrier frequency at a sensation level of 60 dB'. The perceptible difference in the two synthetic sounds under discussion here is consistent with these data, as the modulation frequency is only 25 per cent of the carrier frequency at a sensation level of 60 dB.

Zwicker (1952) observed that, at small modulation frequencies, AM was more easily detectable than FM. But, as the sidebands were placed increasingly farther from the carrier, a point was reached beyond which the just detectable modulation (in terms of sideband amplitude) was the same for both AM and FM. Thus, beyond this point, phase was no longer perceptually relevant. Zwicker *et al.* (1957) found that this critical spread of the three components corresponded with figures for critical bandwidth determined by other methods. Commenting on his critical modulation function, Goldstein (1967) remarked that "the course of the data with carrier frequency is remarkably similar to the critical band function". (For a standard review and discussion of critical bands and critical ratios, see Scharf, 1970.)

Critical band functions for the budgerigar published by Saunders *et al.* (1978) and Dooling and Searcy (1979) show that at a centre frequency of about 1,000 Hz the critical bandwidth is between 300 Hz and 500 Hz. Assuming, now, that the same relationship between critical bandwidth and phase perception applies to budgerigars as it does to man, differences in timbre between strongly amplitude-modulated sounds with centre frequencies of about 1,000 Hz, and sounds with a

structure more akin to quasi-frequency modulation, could be perceived by a bird - and thus have information-bearing significance - at modulation rates up to at least 250 Hz. The range of modulation frequencies in the sounds under discussion in this chapter is centred on about 300 Hz - which is low enough to leave the question open, if nothing more. It would be necessary to carry out behavioural experiments to come to any real conclusions about the perceptibility of modulation differences in these sounds.

A species-specific acoustic pattern contrast?

If it is accepted, though, (as, for example, by Marler, 1969) that differences in timbre resulting from different modulation patterns can have information-bearing significance in the vocalizations of wild birds, it would not be surprising if such contrasts were carried over into the imitation of non-species-specific sounds such as speech. (For a detailed study of the use of species-specific patterns in budgerigar speech imitation, see Chapter 5.)

Turning, finally, to speech imitation by another budgerigar, it is remarkable that this bird, too, regularly duplicates only those phrases containing the [ɔ] sound. Figure 3.2C is one of several similar examples of repetition of the phrase "He's a clever boy" [iz ə· klevə· bɔɪ]. Analysis has shown that each rendition of the phrase follows one or other of the two spectral patterns represented in this figure. (Unlike the [pɔdʒi] of Plate 3.1, it does not seem to matter in this case which of the versions comes first.)

Concentrating on the two renditions of the [ɔ] sound in Figure 3.2C, it can be seen that, in both, the basic pattern is similar to that of the second [ɔ] on Plate 3.1 - i.e., the pattern of a source vibration at about 1,000 Hz (with one or more harmonics) modulated to produce a

complex spectrum. (It follows, therefore, that the same basic sound-production mechanism - involving pulsatile activity of the intrinsic syringeal muscles - could be used in both instances.)

The amplitude spectrum of the first [ɔ] is shown (Fig. 3.2B), having the spectral pattern typical of a strongly amplitude-modulated sound. (Compare Figure 3.1G.) The waveform of the second [ɔ] (Fig. 3.2D) is shown; although very similar to that, for example, in Figure 3.1D, its greater periodic variations in wavelength indicate a greater frequency modulation range. This contrast is borne out in the spectrogram (Fig. 3.2C), which shows the greater number of sidebands in the second [ɔ].

The apparent modulation frequencies in the first and second [ɔ] sounds in Figure 3.2C are about 225 Hz and 175 Hz respectively - well within the half-critical-bandwidth limits established by Saunders *et al.* (1978) and Dooling and Searcy (1979). Thus the duplication of the [ɔ] sound, involving consistent variations on the same basic pattern - variations which are probably distinguishable in terms of timbre - suggest that the bird could be here introducing a meaningful contrast from its species-specific repertoire into its imitation of speech. This contrast could be under the active control of the intrinsic syringeal muscles. (See the Appendix to Chapter 6 for an analysis of the complex variable patterns which underlie a budgerigar's repetitions of a long sentence, and which may be related to species-specific temporal patterning.)

CONCLUSIONS

A consideration of the formant structure of [ɔ] provides some insight into the susceptibility of this speech sound, in *Melopsittacus* imitations, to the kind of sound-production redundancy examined above. It is the vowel in which F2 is of lowest frequency - averaging about 900 Hz in female speech (Peterson and Barney, 1952). In addition, it has a high-frequency F1 (Peterson and Barney, *ibid.*). This means, therefore, that most of the acoustic energy in the sound is contained within a relatively narrow band between about 600 Hz and 1,000 Hz. The budgerigar is able to represent this F1-F2 energy band by a spectral peak related to resonant vibration of the lateral tympaniform membranes. This resonance can be excited using a number of very different sound production techniques.

It is possible that, in examples such as that illustrated on Plate 3.1, the repetition represents an exploratory procedure, originating at an early stage of "speech" learning which has become "frozen" as part of the bird's repertoire. (See the discussion of avian "babbling" in Chapter 5.) This, then, would represent true physiological redundancy, the two sounds being simply two different attempts to reproduce the same speech sound. In other examples of [ɔ] duplication, such as that illustrated in Figure 3.2C, it is possible that a species-specific sound contrast is finding its way into the bird's imitation of speech. (This idea of a species-specific sound contrast would have to be tested by behavioural studies using the birds' natural vocalizations.)

Intrinsic syringeal muscles and vocal plasticity

This chapter has argued that it is specifically the activity of the intrinsic syringeal muscles which allows plasticity of the vocal mechanism.

Recent speculation has linked such plasticity - including the capacity for vocal copying - with the possession of intrinsic syringeal muscles. Nottebohm *et al.* (1976) found, in the canary, a direct motor pathway between the telencephalic vocal centres and the hypoglossal nuclei. They observed that the mid-brain nuclei (which control the respiratory activity of the extrinsic syringeal muscles) appear to be unimportant in song control. Thus Brackenbury (1979b, 1980) regarded the intrinsic syringeal muscles as being "emancipated" from the respiratory system. Similarly, Gaunt (1983) believed that the complex vocal behaviour of songbirds and parrots could be explained, in part, by the "dissociation" from respiratory function which intrinsic syringeal muscles allow. He also believed that, to be effective, the development of intrinsic syringeal muscles must be accompanied by a "neurological shift". In Chapter 1, under "Vocal plasticity and neural complexity", this "neurological shift" is discussed in terms of a neural "complexity threshold". The imitative abilities of songbirds and parrots are an example of such exploitation of complex syringeal musculture by central nervous systems having crossed such a "complexity threshold".

Acoustic evidence presented in this chapter supports the connexion, which emerges from the theoretical work reviewed immediately above, between the possession of an intrinsic syringeal musculature and plasticity (including copying ability) in avian vocal behaviour. The evolutionary development of these muscles must have been accompanied by a "neurological shift" which made possible their employment in a variety of complex behaviours - including vocal imitation. (Redundancy could occur when a

sound produced using this more recently evolved musculature is interchangeable, in the particular behavioural context, with a sound produced using the archaic "phonatory" mechanism.)

Speech imitation is an ideal context in which to observe these phenomena, as we can see the bird experimenting with the possibilities of physiological function and acoustic effect in attempting to reproduce sounds which are alien (or opaque) to it but familiar (or transparent) to us. Speech imitation is thus a window through which we can observe some of the mechanisms underlying avian vocal communication.

SUMMARY

A budgerigar employs, alternatively, what appear to be two radically different physiological strategies in reproducing the speech sound [ɔ]. Two possible methods of sound production are investigated with the aid of acoustic analysis and in the light of the anatomical and physiological studies reported in Chapter 2. It is argued that one method of sound production involves a "passive" vibration mechanism similar to that underlying human phonation, while the alternative method involves active muscular control of complex modulation patterns.

Synthetic sounds are used to investigate the generation of these modulation patterns, and the aural effect of speech-like low pitch (in the absence of low-frequency spectral components) which they produce. A possible mechanism of AM-FM coupling in bird sounds is outlined.

It is suggested that the actively modulated method of sound production is more recently evolved than the phonatory method. Redundancy at the physiological level in vocalization can reflect the coexistence and interchangeability of such "primitive" and "advanced" vocal mechanisms.

NOTES TO CHAPTER 3

[1]This illustrates the importance of passive membrane vibration in the vocalizations of this species. Inactivating both these pairs of muscles had less effect than section of the sternotrachealis muscles alone. The latter operation "eliminates low level vocalizations and alters the character of calls at any intensity" (Youngren *et al.*, 1974). Phillips and Youngren (1981) reported similar results in *Gallus* chicks. They interpreted this as indicating the importance of co-ordinated, antagonistic activity of *Mm. sternotrachealis* and *tracheolateralis* during vocalization. Peek and Phillips (1971) found that the "cuk" call of the same species was unaffected by section of both tracheolateralis and sternotrachealis muscles and the hypoglossal nerves.

[2]The evolution of the more efficient oscine syrinx (see Brackenbury, 1979b) saw the development and variation of this second sound-production mechanism in preference to the first. The anatomy of the songbird's syrinx (Fig. 2.3D) makes adduction of the two main vibratile membranes (the medial tympaniform membranes) impossible.

[3]Greenewalt (1968) reported instances of such relationships between carrier and modulation frequencies in the songs of some oscine species. In songbirds, however, there is no clear-cut physiological interdependence between the vibration mechanisms producing the two different frequencies, and such instances appear to be the exception rather than the rule.

[4]For a more recent perspective on, and critique of, this early work on "residue" pitch perception, see Plomp (1976) and de Boer (1976).

[5]Wilson *et al.* (1971) showed that, in such a whistle, sound pressure level varies directly with the velocity of airflow through the orifice until a certain critical velocity is reached, and then the relationship is inverse. As they found, also, that velocity varies directly with orifice diameter, and orifice diameter varies directly with frequency, it follows that the relationship between frequency and sound pressure is the same as that between velocity of air flow and sound pressure. This pattern corresponds with the data of Greenewalt (1968) and others, which show direct coupling of amplitude and frequency up to a certain critical frequency, and inverse coupling with further rises in frequency.

[6]Andrew, 1973, reported on experiments which showed that section of "tracheal" muscles in chicks eliminated frequency modulation from their calls. Phillips and Peek (1975) found that section of both tracheolateralis and sternotrachealis muscles reduced the FM range of chicken calls by about 50 per cent, and concluded that "tracheal air flow and interclavicular air sac pressures also contribute importantly".

[7]Rosen (1986) has published a comparison of Goldstein's discriminable modulation frequency data with a recent estimate of auditory filter bandwidth (i.e., critical bandwidth) which shows an even closer matching than in Goldstein's presentation.

CHAPTER 4:
PERCEPTION OF COMMUNICATION SOUNDS

INTRODUCTION

What does human speech sound like to a bird? And how are speech signals processed by the avian auditory system?

This chapter assesses information which can help answer these questions. (Such "answers" must, of course, be limited to those aspects of avian aural experience which are shared by, or are at least comprehensible to, human investigators.)

The first half of the chapter reviews experimental data on the peripheral reception of acoustic stimuli in birds, the analysis of these stimuli at the periphery and at higher levels, and patterns of neural projection from the inner ear to medullary, mesencephalic and telencephalic auditory centres. (Evidence of neural connexions between auditory and vocal centres is also reviewed.) Some significant differences in acoustic analysis between man and bird are revealed. This data is related to physical parameters of both human speech and avian "speech".

The second half of the chapter deals with pattern processing in the perception of complex acoustic stimuli. It considers the possibility that certain elements in the acoustic patterns of vocal communication signals can have interspecific significance, and suggests that common "bio-communication cues" in bird sounds and human speech are central to an understanding of avian speech imitation.

The neural correlate of an acoustic communication cue is a "feature detector" (Eimas and Corbit, 1973). A feature detector can be regarded as a selective auditory filter (Green *et al.*, 1977). A "bio-communication cue" can be understood, therefore, as an acoustic feature which is similarly selected, by higher-order neural

filtering, in species among which sound reception at the periphery can differ significantly.

Thus, although the evidence suggests that there are major differences between human and avian auditory percepts of speech, it also suggests that, at higher levels of pattern processing, certain features of the speech signal are similarly extracted in both man and bird.

PERCEPTUAL COMPLEXITY

The avian peripheral auditory system

Introduction. The spectral diversity of avian vocalizations is matched by an auditory system which responds to a wide range of frequencies. The avian ear is not mechanically tuned, as in frogs (Frishkopf and Goldstein, 1963), to respond only to certain biologically determined frequencies.

Physical constraints. Greenewalt (1968) could find no correlation between the size of a bird and its highest vocal frequency. Similarly Konishi (1970) found no relationship between a bird's body weight and its range of audible frequencies. Schwartzkopff (1955), however, noted an inverse relationship between the size of related avian species and their most sensitive audible frequencies - i.e., the larger the bird, the lower the most sensitive frequency - due, he believed, to the dimensions of vibratile elements in the ear.

Overall sensitivity is determined by the sound power transformation ratio in the middle ear - a ratio which can be simply expressed, in anatomical terms, as ear drum area/columella base area (Schwartzkopff, 1955; Pumphrey, 1961). (For reviews of outer and middle ear morphology, see Schwartzkopff, 1973, and Kuhne and Lewis, 1985). Schwartzkopff's (1955) table (which, unfortunately, does not include details of any representative of the Psittacidae) shows that small birds have larger tympanic membranes, relative to their body size, than large birds, thus ensuring their acuity of hearing. The area transformation ratio is seen to be far greater for owls than for other species listed.

The basilar membrane. Measurements of the basilar membrane in *Gallus* reported by Smith (1985) show that the width of the membrane increases in a regular manner from 0.07 mm at the proximal tip to 0.395 mm at the distal end. "It seems probable that a similar increase in width also occurs in *Columba*. However, in the budgerigar (*Melopsittacus undulatus*) the short (2.3 mm in length) basilar membrane shows little change in width over the distal 1.5 mm, that is, over 65 per cent of its total length" (Smith, *ibid.*).

The basilar papilla. The basilar papilla, which covers most of the upper surface of the basilar membrane and extends onto the superior fibrocartilaginous plate, ranges in length from slightly over 2 mm in small birds to 9.5-11.5 mm in the barn owl (Smith, *ibid.*). Variations in the width of the basilar papilla follow the same patterns as variations in the width of the basilar membrane. Thus the papilla gradually broadens towards the distal end in *Columba* (Takasaka and Smith, 1971) and *Gallus* (Tanaka and Smith, 1978). "In the budgerigar and the barn owl, however, the change in width of the papilla is not so gradual" (Smith, 1985). Boord (1969) reported that the average length of the basilar papilla in the specimens of *Columba livia* examined by him was 3.74 mm, surface area 0.89 sq mm, and maximum width 0.4 mm.

The basilar papilla is made up of sensory cells and supporting cells. For reports on the morphology and distribution of the sensory cells, and their individual nerve endings, see Takasaka and Smith (1971) and Tanaka and Smith (1978). Smith (1985) presents figures illustrating the arrangement of the three major types of sensory cells - tall, intermediate and short hair cells - into three discrete areas within the papillae of *Columba livia*, *Gallus gallus* and *Melopsittacus undulatus*. The pattern formed by these three areas seems to be species-specific (Smith, *ibid.*).

Each sensory cell within the basilar papilla contains a terminal of the cochlear part of the vestibulocochlear nerve (Takasaka and Smith, 1971). (Boord, 1969, published photomicrographs showing the entry into the basilar papilla of the nerve fibres which innervate the sensory cells.)

Cochlear vibrations. Von Békésy (1960) observed tonotopy of response peaks along the basilar membrane of *Gallus* for excitation frequencies between about 200 Hz and 3,000 Hz. The response to the lower frequencies of this range was nearer the stiffer, narrow (apical) end of the membrane, while the response to higher frequecies was nearer its softer, wider (basal) end. These observations have generally been regarded as indicative of a travelling wave along the cochlea. This inference has recently been confirmed by Smolders *et al.* (1986), who demonstrated the existence of a travelling wave along the basilar membrane - from the base towards the apex - in the pigeon. Narins and Gleich (1986) have added further corroborative evidence in relating patterns of firing of cochlear ganglion cells in the inner ear of the starling to mechanical cochlear vibrations.

Bohne and Dooling (1974) reported morphological changes to sensory cells on parts of the basilar papilla in budgerigars as a result of exposure to noise. This observation supported the general notion of place analysis of sound along the basilar papilla. More recently, data on the varying position of basilar papilla damage following tonal overstimulation in chicks of different ages have been adduced as evidence of a change in the spatial encoding of frequency along the papilla during ontogeny (Rubel and Ryals, 1983; Ryals and Rubel, 1985).

Acoustic analysis in the cochlea: historical perspective. Although much of the information summarized in this chapter is strongly indicative of frequency analysis at the periphery, there is still no thoroughgoing mechanical model of this process in the avian cochlea. Historically, the main objection (as in Schwartzkopff, 1955) to the idea of peripheral sound analysis in birds has been the shortness of the basilar membrane - e.g., 2.6 mm in *Amazona amazonica* (Denker, 1907), which has similar pitch discrimination ability to man (Schwartzkopff, 1955). (The length of the human basilar membrane is 30-35 mm.) Another apparent problem is that the basilar membrane does not increase regularly in width in some birds, and indeed, in these species, "physical properties seem to change very little over a large part of the basilar membrane" (Smith, 1985). Schwartzkopff (1955) also proposed that the peculiar structure of the avian cochlea would produce high damping of resonant vibrations[1]. Schwartzkopff's final objection to peripheral analysis was that the sensory cells, so densely distributed within the avian basilar papilla, lack the relative isolation he thought necessary for such an analysing function.

Pumphrey (1961), on the other hand, used this very fact - that a transverse section of an avian basilar papilla reveals about ten times more sensory cells than a section of a mammalian basilar papilla - in arguing in favour of peripheral analysis. He pointed out that while the basilar membrane in birds is, on average, about one-tenth the length of that in mammals, the avian cochlea has ten times as many sensory cells per unit length. He believed that this would allow more precise intensity discrimination in birds. (As reported below, however, psychophysical tests have subsequently indicated rather poorer intensity discriminaion in birds than in mammals.)

"If the avian cochlea is, in fact, ten times as sensitive to changes in amplitude as the mammalian, it is

theoretically possible for the same degree of frequency discrimination over a given band of frequencies to be achieved with one-tenth of the number of frequency sensitive 'elements'; and each element need be only one-tenth as selective and consequently *may respond ten times as fast*" (Pumphrey, *ibid.*).

Saito (1980) argued that two factors - (1) the high elasticity of the basilar membrane, and (2) the firm attachment (more firm than in mammals, apparently) of the hair cells to the tectorial membrane (which is much thicker than in mammals) - "permit frequency analysis to be much more precise and efficient along the shorter avian cochlear duct than along the mammal's" (Saito, *ibid.*).

Most authors would agree with Smith (1985), however, that "the peaked sensitivity curve and limited frequency range of most birds are reflections of their short cochleas. . . The barn owl (*Tyto alba*) provides an exception, with a long basilar membrane and a wide frequency range". Konishi (1970) believed that place analysis on the basilar membrane preserves spectral characteristics "for further analysis by more central stations".

Avian hearing: data from behavioural tests

Audibility curves. Audibility curves for the canary, *Serinus canarius* (Dooling *et al.*, 1971), the redwing blackbird, *Agelaius phoeniceus* (Hienz *et al.*, 1977), the brown-headed cowbird, *Molothrus ater* (Hienz *et al.*, *ibid.*), the house finch, *Carpodacus mexicanus* (Dooling *et al.*, 1978) and the field sparrow, *Spizella pusilla* (Dooling *et al.*, 1979) - all passerine species - are remarkably similar, and closely resemble curves derived by earlier authors for the starling and bullfinch (reproduced in Dooling *et al.*, 1971). Sensitivity in all these species is greatest between 2 kHz and 4 kHz - a

frequency range which includes the dominant spectral components of their songs.

Comparisons of average acoustic power spectra and audibility curves for the field sparrow (Dooling *et al.*, 1979) and the house finch (Dooling *et al.*, 1978), however, reveal that the spectral peak for most vocalizations is somewhat higher (3-3.5 kHz) than the peak in the audibility curve (about 2 kHz). Konishi (1970) asserted that, in all songbirds studied thus far, dominant vocal frequencies were higher than the most sensitive range of hearing, and suggested that "birds may be choosing the frequency range in which they can obtain the best signal/noise ratio".

Dooling *et al.* (1971) reported that while the dominant frequencies of canary song were within the most sensitive range of audible frequencies, dominant frequencies of canary calls extended into less sensitive ranges of hearing. They explained this by the generalization that "in the songbirds, calls are usually private signals used at close range". (Note here that speech-imitation is usually such a "private" signal.)

The audibility curve of the budgerigar (Dooling, 1973; Dooling and Saunders, 1975a; Dooling *et al.*, 1978, 1979) shows that this species is more sensitive than the songbirds throughout much of its range. Thus, while the budgerigar's most sensitive frequency falls within the same range (2-4 kHz) as that of most songbirds, it has much greater sensitivity at all audible frequencies below this peak, and is somewhat more sensitive to frequencies as high as 6 kHz. Recent research has shown that the pigeon (*Columba livia*) is sensitive to very low-frequency sound. (See Dooling, 1980, for a review, and Dooling *et al.*, 1978, for the pigeon's audiogram.) The audibility curve of the barn owl, *Tyto alba* (Konishi, 1973) shows that this species is more sensitive than all others tested at all audible frequencies, and is more sensitive within the broad range of 1-8 kHz than any other species

at that species' most sensitive frequency.

Dooling, 1980, published a table of parameters describing the audibility curves of 16 avian species.

Hearing range. Schwartzkopff's (1973) list of upper and lower limits to the hearing range of 23 avian species indicates that the lower limit of the budgerigar - 40 Hz - is considerably below that of the songbirds, and even below that of owls. The upper limit of most songbirds however (about 20,000 Hz as determined by behavioural methods), exceeds that of the budgerigar (14,000 Hz).This low-frequency sensitivity helps explain the apparent ability of the budgerigar to hear low-frequency components in the human speech signal, while the maximum sensitivity of 2-4 kHz corresponds with the F2-F3 region of the speech signal which budgerigars reproduce so well. (See Chapters 3, 5 and 6.)

Konishi (1974) pointed out that "all birds studied so far are sensitive to frequencies far below the lowest frequencies in their vocalizations". Thus, while there is a correlation between the highest audible frequency and the highest vocal frequency, there is no such relationship between the lowest audible and vocal frequencies. (In speech imitation, low-frequency fundamentals are often represented by modulations of a higher-frequency carrier - Chapters 2, 3 and 6.)

Frequency discrimination. Dooling's (1982) comparison of data from behavioural testing of five avian species showed that frequency discrimination is best in all five species between 1 kHz and 4 kHz, and that within this range birds can perceive a 1-2 per cent change in frequency - i.e., 10-20 Hz at 1 kHz, 20-40 Hz at 2 kHz, 40-80 Hz at 4 kHz. (Sinnott *et al.*, 1976, reported a just noticeable difference of 30 Hz at 4 kHz in two Icterid passerine species - *Agelaius phoeniceus* and *Molothrus ater*.) Sinnott *et al.* (1980) found that both pigeons and

blackbirds were, like the budgerigar (Dooling and Saunders, 1975a), most sensitive to frequency changes at around 4 kHz. (Blackbirds were more sensitive than pigeons to frequency changes above 1 kHz, and pigeons were more sensitive than blackbirds to frequency changes below 1 kHz.) At 4 kHz, which represents the lowest absolute threshold on the blackbird's audibility curve and a major component of its song spectrum, the blackbirds' sensitivity to frequency changes "showed no clear dependence on sensation level" (Sinnott *et al.*, 1980).

Frequency discrimination thresholds are higher than those for man at most frequencies. Dooling (1980) remarked that "these results are puzzling in view of electrophysiological data from the avian auditory system, such as tuning curve width and temporal synchrony [discussed below, this chapter], which suggest that birds appear to code stimulus frequency as well as mammals".

The optimal frequency-discrimination range of 1-4 kHz correlates well with the range of dominant spectral components in both species-specific avian vocalizations (see above) and speech-imitation sounds (Chapters 5 and 6). It also correlates well with indirect (psychophysical) evidence concerning resonant vibrations along the avian basilar papilla. (See below, this chapter.)

Intensity discrimination. Dooling (1980) compared psychophysical data from five avian species which indicate that intensity difference thresholds are similar to those for most other vertebrates tested (and somewhat higher than those for man). This is in spite of arguments (Pumphrey, 1961 - see above) that the greater number of hair cells per unit length along the avian basilar papilla compared to mammals could mean a greater power of intensity discrimination in birds.

Temporal resolution. Dooling and Searcy (1981) studied minimum temporal integration time in the budgerigar by describing sensitivity to an amplitude modulated signal as a function of modulation frequency. They reported that budgerigars can detect about a 10 per cent change in modulation rate over the range of modulation frequencies from 40 Hz to 320 Hz. Their data showed "statistically significant differences between the modulation functions of parakeets [*Melopsittacus*] and humans in the direction of more efficient temporal processing by the avian auditory system" (Dooling and Searcy, *ibid.*). Dooling (1982) quantified this difference by referring to a time constant in the budgerigar of about 1.2 msec compared to a time constant in humans of 5-6 msec.

For modulation frequencies below 40 Hz, modulation threshold was independent of modulation rate - i.e., it depended entirely on the detection of intensity changes (modulation depth). The birds' performance in this region was poorer than the humans', which led Dooling and Searcy to remark that the budgerigar appeared to compensate for poorer intensity resolving power than humans with better temporal resolving power. (That this is an oversimplification, however, can be seen from one parameter of their data, which shows that the birds maintained their level of sensitivity to modulation depth up to modulation frequencies of about 130 Hz, while human sensitivity began to decline for modulation frequencies above 28 Hz.)

In other psychophysical tests of temporal resolution, birds have performed similarly to human subjects. These tests, and the avian species involved, include auditory duration discrimination in the budgerigar (Dooling and Haskell, 1978), gap detection thresholds for the house finch, *Carpodacus mexicanus* (Dooling *et al.*, 1978), two-click thresholds for the pigeon, *Columba livia*, bullfinch, *Pyrrhula pyrrhula*, and greenfinch, *Carduelis chloris* (Wilkinson and Howse, 1975), and temporal

integration thresholds (i.e., maximum integration time) for the field sparrow, *Spizella pusilla* (Dooling, 1979), and the budgerigar (Dooling, 1979; Dooling and Searcy, 1985a).

Dooling and Searcy (1981) remarked that "there are many ways to vary the temporal characteristics of a signal, and some of these may be more salient to the avian ear than others". The present study has shown (Chapters 3 and 6) that amplitude modulation is an important ingredient of psittacine acoustic signals. It has also brought forward evidence (in Chapter 3) which suggests that differences in timbre between amplitude-modulated and frequency-modulated sounds with modulation frequencies of 150-350 Hz may have information-bearing significance for budgerigars. It is possible that birds perceive the acoustic correlates of low-frequency peaks (e.g., fundamental and first formant) in the speech spectrum as amplitude modulations of a higher-frequency (i.e., second-formant) carrier - see below, this chapter, under "Single-formant avian 'speech'".

A special instance of the precise timing of intervals within avian sound sequences - the reaction time in duet singing - was studied by Thorpe (1963), and used as an indication of the temporal resolving power of the avian auditory system. Thorpe found that reaction time (i.e., the interval between the beginning of one bird's sound and the subsequent beginning of its partner's) in the duet of a species of East African shrike - the black-headed gonolek (*Laniarius erythrogaster*) - was both three times faster and three times more consistent than was possible for human subjects given a comparable task. From his data, he estimated that the mean reaction time was somewhere between 90 msec and 135 msec. Snow (1977) recorded reaction times of between 40 msec and 240 msec in duets by species of the genus *Chiroxiphia*[2]. (For reviews of avian duetting behaviour, see Farabaugh, 1982; Thorpe, 1972.)

Frequency selectivity

Introduction. Psychophysical studies have provided valuable data on the analysing capabilities of the avian auditory system, and some insights into the nature of the analysing mechanism itself. Most of the published psychophysical data relate to the budgerigar (*Melopsittacus undulatus*), one of the main species under investigation in this study.

Critical bandwidths. Saunders *et al.* (1979) compared the bandwidths of internal auditory filters[3] in the budgerigar as determined by two psychoacoustic procedures - psychophysical tuning curves (Saunders, Else and Bock, 1978) and critical bands (Saunders, Denny and Bock, 1978)[4]. They found that the shape of the curve describing "Q" (an index of filter sharpness) as a function of the centre frequency of the filter was the same for the results of both procedures. Their data showed that "in a narrow range of frequencies (3-4 kHz) the selectivity of parakeet [*Melopsittacus*] hearing is similar to or even better than that observed in human listeners" (Saunders *et al.*, 1979). (Dooling and Saunders, 1975a, had shown that *Melopsittacus* critical bandwidths were narrower between about 3 kHz and 4 kHz than any critical bandwidths measured in man or other mammals.)

Nonsimultaneous masking. Saunders *et al.* (1979) pointed out, however, that because of complex interactions within the cochlea during simultaneous masking such as they had used, frequency selectivity was probably more finely tuned than indicated by their experiments. Dooling and Searcy (1985b) tested this hypothesis using nonsimultaneous (forward, backward, and forward-and-backward) masking techniques and found that, in the spectral region of 2-4 kHz, the differences

between *Melopsitacus* and human were even more pronounced than had previously been supposed. "Indeed", they wrote, "the budgerigar shows more than an order of magnitude greater frequency selectivity than the human in the backward masking condition." In all three masking conditions, considerably less masking occurred for budgerigars than for humans at all masker frequencies.

Also, for the budgerigar, the degree of frequency selectivity (as expressed by the Q ratio) differed by a statistically significant amount among the three masking conditions, whereas there was no significant difference for humans.

The Q ratio for budgerigars was greatest in backward masking, indicating greater selectivity in this condition (Dooling and Searcy, 1985b). These authors (Dooling and Searcy, 1980) showed that there is no masking effect in budgerigars when the backward masker occurs 10 msec or more after the probe tone, whereas in humans there is a masking effect at delays of up to 20 msec. This suggests either that masking is interfering with later, more central analysing processes in the human (Dooling and Sercy, *ibid.*), or that the neural process of audition is, as a whole, faster in budgerigars than in humans.

Critical ratios. The shape of the budgerigar's CR curve is markedly different from CR curves for mammalian species, including man. In these other species, the critical ratio rises at a rate of about 3 dB per octave over much of its length (Dooling and Saunders, 1975a). In *Melopsittacus*, however, such a slope occurs only between about 2 kHz and 4 kHz (Dooling and Searcy, 1979). As it has been generally agreed, after von Bekesy (1960), that resonant frequencies are organized in logarithmic progression along the basilar papilla, it follows that an increase of 3 dB per octave (equivalent to an intensity ratio of 1:2) in the critical ratio function represents a series of equally-spaced critical bands within the

cochlea (Scharf, 1970). This has led to speculation (Dooling and Searcy, 1979) that logarithmic organization of resonant frequencies along the budgerigar's basilar papilla may occur only within the 2-4 kHz range.

Patterns of auditory projection

The cochlear nuclei. Boord and Rasmussen (1963) observed orderly patterns of projection of primary acoustic fibres from the papilla on to both the angular and the magnocellular cochlear nuclei. The fibres' order of entry into the *medulla oblongata* depends on whether they originated on the basal, middle or apical third of the basilar papilla. These three groups of fibres terminate in units having high, intermediate and low characteristic frequencies respectively (Konishi, 1969).

The tonotopic organization of acoustic neurons in the cochlear nuclei (Konishi, 1970) supports the view of a "point-to-point projection of the basilar membrane on the hind brain" (Konishi, 1974). As the relationship between the characteristic frequencies of neurons and their spatial separation in the cochlear nuclei is nonlinear, however, Konishi (*ibid.*) argued for a nonlinear arrangement of resonant frequencies along the basilar membrane, as in mammals. Following von Bekesy (1960), he (Konishi, 1970) argued that the avian basilar membrane functions as a mechanical wave analyser "and that the place principle proposed for the mammalian cochlea may be applicable to birds".

Responses of single auditory neurons in the cochlear nuclei have been observed in several avian species. Konishi (1969, 1970, 1974) found that the relative thresholds of these neurons, when plotted against their characteristic frequencies, formed a pattern remarkably similar to the audibility curve obtained by behavioural methods. (See above.) Such a direct correlation between behavioural perceptual performance and neuronal response

is unusual, and indicates that the auditory system is sending most of the frequency-dependent acoustic information received at the periphery directly to the brain (Konishi, 1969b).

Konishi (1970) presented a table of highest and lowest characteristic frequencies for 10 species of songbird. Lowest CFs are between 135 Hz and 655 Hz; highest CFs are between 6,200 Hz and 9,382 Hz. The range of hearing is greater than the range of characteristic frequencies, however (see above), and depends, towards its upper limit, on the intensity of the acoustic signal.

Konishi (1969b) found that "the greatest number of units occurs within the frequency range where unit thresholds are lowest", thus providing a channel for the processing of species-specific sounds within this range. He (Konishi, 1970) graphically compared vocalization spectra of six songbird species with single auditory neuron threshold distribution, showing correlations among dominant acoustic frequencies, lowest unit thresholds, and densest unit distribution.

Rubel and Parks (1975) reported similar tonotopic organization of *nucleus laminaris* (NL) and *nucleus magnocellularis* (NM) in chickens. As *nucleus laminaris* is innervated primarily or entirely by *nucleus magnocellularis*, this demonstrated a further extension of point-to-point projection from the cochlea. Parks and Rubel (1975) found that "neurons in NM project in a very discrete topographic, tonotopic and symmetrical fashion to NL on both sides of the brain, contributing to the binaural response properties and tonotopic organization of neurons in NL".

The mesencephalon and telencephalon. Boord (1965, 1968) traced efferent auditory fibres (in *Columba*) from the cochlear nuclei to the *Nucleus mesencephalicus lateralis dorsalis* (MLD) in the midbrain. Karten continued the tracing of ascending auditory projections

in *Columba* - from the MLD to the thalamic *Nucleus ovoidalis* in the diencephalon (Karten, 1967), and from the *Nucleus ovoidalis* to field L of the caudal neostriatum (Karten, 1968). This author remarked that "the organization of the avian ascending auditory pathway is remarkably similar to that of its mammalian counterpart" (Karten, 1968). (Boord, 1969, published a useful diagrammatic representation of the ascending auditory pathway in *Columba*.)

Measurements by Cobb (1964) showed a wide variation in the relative and absolute size of MLD in 27 avian species, and correlated these variations with degrees of behavioural dependence on auditory cues. Karten (1968) reported a correspondingly greater development of *Nucleus ovoidalis* and telencephalic field L in those species with larger MLD.

In observing tonotopic organization of single units in field L of the zebra finch (*Poephila guttata*), Zaretsky and Konishi (1976) demonstrated a point-to-point projection from the avian cochlea, along the ascending pathway outlined above, to the auditory area of the telencephalon. Similar tonotopic organization has been found in the primary auditory cortex of mammals (e.g. Evans and Whitfield, 1964). Thus, "not only is there anatomical homology between the avian and mammalian auditory pathways, but also physiological homology at the level of functional architecture of auditory neurons" (Zaretsky and Konishi, 1976).

Bonke, Scheich and Langner (1979) reported that the area of the neostriatum in the Guinea fowl (*Numida meleagris*) receiving most auditory axons from the *nucleus ovoidalis* - corresponding to field L as delimited by earlier writers - was bounded dorsally and ventrally by two other auditory areas. While extending the term "field L" to embrace the whole of this three-layered structure, they referred to the layers specifically as (dorsal to ventral) L_1, L_2 and L_3. They found an overall tonotopic

organization of L_2, isofrequency lines corresponding with the course of fibres arriving from *nucleus ovoidalis*. Isofrequency lines also ran perpendicular to L_2 through L_1 and L_3. It is believed that there is input to L_1 and L_3 both from L_2 and directly from *nucleus ovoidalis* (Scheich *et al.*, 1983).

The orderly projection of auditory fibres from the cochlea to the telencephalon, revealed by the studies summarized above, supports the view of a mechanical analysis at the periphery, the results of which are then transferred to successively higher centres. This process would seem to involve a progressive abstraction in encoding. "It now appears that at least a portion of the [avian] auditory system is a hierarchically organized analysing network in which is found a progressive degree of abstraction of the acoustic signal as one proceeds centrally along the auditory pathway" (Sachs *et al.*, 1980).

(Such a direct relationship between mechanical events in the cochlea and neuronal firing in the telencephalon is supported by the psychophysical data, derived from masking experiments, reviewed earlier in the chapter.)

The model, developed below, of signals received somewhat differently at the periphery in man and bird being subject, subsequently, to similar neural filtering processes, is consistent with such a "hierarchical" system.

Voco-auditory integration. Nottebohm *et al.* (1976) identified the *hyperstriatum ventrale, pars caudale* (HVc) in the canary (*Serinus canarius*) as one of the telencephalic nuclei involved in the control of song, and noted the close proximity of HVc to field L. They suggested that HVc might be particularly involved in the process of song learning - i.e., processing perceived patterns into motor commands. Zaretsky (1978) located an auditory region in the telencephalon of the zebra finch

and the canary, distinct from and superficial to field L, which he identified with HVc, thus supporting the proposed projection of auditory fibres to this song-control centre.

Kelley and Nottebohm (1979) confirmed this connexion between auditory and vocal telencephalic centres in songbirds by tracing projections of field L to both HVc and the robust nucleus of the archistriatum (RA) - the other song-control centre identified by Nottebohm *et al.* (1976). Saito and Maekawa (1981) traced an anterograde projection to RA from a nucleus in the diencephalon of the mynah (*Gracula religiosa*) immediately ventral to *nucleus ovoidalis* - the auditory nucleus of the thalamus - and regarded this as further evidence of central voco-auditory integration in songbirds.

Bonke, Bonke and Scheich (1979) observed connexions between field L in the Guinea fowl (*Numida meleagris*) and two telencephalic areas thought to be involved in vocal control - the hyperstriatum and the dorsal neostriatum. (The dorsal neostriatum is located at the posterior edge of the *hyperstriatum ventrale* - i.e., in approximately the same position as HVc in songbirds.) "From our data, it can be concluded that the connexions of the auditory neostriatum are widespread and complex and include several connexions to presumed vocal-motor areas" (Bonke *et al.*, *ibid.*).

This anatomical evidence of voco-auditory neural interaction corroborates developmental data on vocal learning in birds (reviewed in Chapter 1) and speech-imitation data presented throughout this thesis.

Patterns of neuronal response

Relationship to cochlear events. Konishi (1969a) recorded impulses in single units within the *nucleus angularis* of sparrows and pigeons. He found that, when stimulated by a single click, a unit produces several impulses separated by an interval which is the reciprocal of the unit's best frequency. He interpreted this phenomenon as representing resonant oscillations of the basilar membrane in response to the click.

A comparison of data from songbirds and cats (Sachs *et al.*, 1974, 1980; Woolf and Sachs, 1977) indicates a higher degree of frequency selectivity in auditory-nerve fibres in birds than in mammals. The most obvious difference consists in the lack of a low-frequency "tail" in the avian tuning curves. This has been interpreted as reflecting a sharper and more symmetrical travelling wave envelope along the avian basilar papilla (Dooling, 1980, 1982), as first suggested by Saunders, Bock and Fahrbach (1978).

Psychophysical tuning curve data obtained in masking experiments (Dooling and Searcy, 1985b; Saunders, Bock and Fahrbach, 1978; Saunders, Else and Bock, 1978; Saunders *et al.*, 1979) corroborate the neurophysiological evidence in demonstrating less spread of masking to frequencies higher than the masker frequency in birds than in mammals. "This may be part of the mechanism by which birds attain excellent frequency resolving power in spite of a short receptor surface" (Dooling, 1980). (Recent reports on basilar membrane vibration, however - e.g. Narins and Gleich, 1986, and Smolders *et al.*, 1986 - have agreed that basilar membrane tuning curves are less sharp than psychophysical or single-unit tuning curves, and that further filtering by the hair cells themselves, in addition to the membrane resonance, may sharpen the response.)

The range of frequencies within which such sharp place analysis occurs is more limited than in mammals, and probably does not extend below about 1 kHz (Dooling, 1980). At frequencies below about 1.5 kHz mechanisms in the time domain (i.e., periodicity analysis) probably predominate (Konishi, 1969a; Langner, 1981).

As in mammals, the firing of auditory-nerve fibres is phase locked to the waveform of a sinusoidal acoustic stimulus between frequencies of about 1 kHz and 6 kHz (Sachs *et al.*, 1980; Woolf and Sachs, 1977). Frequency discrimination data (see above) have shown this to be the range most sensitive to frequency changes - lending weight to the argument for a finely-tuned place analysis along the basilar papilla within this range.

Thus, as Sachs *et al.* point out, "although there are structural differences between the avian and mammalian auditory peripheries, there do not appear to be gross qualitative differences in neural encoding at the level of the auditory nerve". (One difference, reported by Gross and Anderson, 1976, in relation to *Columba*, is the presence of auditory-nerve fibres which respond solely to "sweep frequencies". Although cortical neurons have been discovered in mammals which similarly respond only to stimuli of changing frequency - Whitfield and Evans, 1965 - this may be an indication of more peripheral processing in birds. Backward masking data also indicate this possibility - see above.)

Periodicity analysis. Konishi (1969a) found that, when stimulated by a train of clicks, some units in *nucleus angularis* exhibited nearly 100 per cent time locking at click frequencies of up to about 1,500 per sec.

Langner (1981) studied units in the auditory midbrain nucleus (MLD) of the Guinea fowl (*Numida meleagris*) which were responsive to amplitude modulated tones, and found that some of these units were selectively tuned to respond to particular modulation frequencies. This led

him to believe that periodicity analysis (i.e., analysis in the time domain) could be a factor in avian pitch perception. This idea is supported both by Konishi's (1969a) time-locking data recorded in the cochlear nucleus, and by phase-locking data on response to sinusoidal stimuli at the level of the auditory nerve (Sachs *et al.*, 1980; Woolf and Sachs, 1977). Scheich *et al.* (1977) found midbrain units responsive to harmonic spectra in the Guinea fowl, and speculated that "harmonic analysis could come about by some mechanism of periodicity coding in peripheral neurons each locking to a particular harmonic".

"Simple" and "complex" neurons. Scheich *et al.* (1977) recorded responses to species-specific calls from single units in the MLD of the Guniea fowl. They found "simple neurons", from which responses to the spectrally complex calls could be predicted from the units' best frequencies as determined by pure-tone testing. (Each of these units responded simply to a particular narrow band of frequencies, with centre frequencies ranging from about 100 Hz to 9,000 Hz.) They also found "complex neurons", from which responses to the species-specific calls could not be so predicted. (Indeed many of these neurons failed to respond at all to pure tone stimuli.) These neurons were responsive to one or more of the following cues: multiple "formant" structure and harmonic spectra (spectral cues); temporal cues such as pulse repetition rate; amplitude and frequency modulation. (Biederman-Thorson, 1970, had reported a similar dichotomy between neurons with simple and complex response patterns in the diencephalic *nucleus ovoidalis* of the ring dove.)

The most sensitive spectral region for all mid-brain auditory neurons tested was 1-4 kHz (Scheich *et al.*, 1977) - the range of dominant spectral peaks in Guinea fowl vocalizations (and the range of F2 and higher formant frequencies in human speech). One unit tested was

responsive to a tone at 320 Hz ("a common fundamental in these calls" - Scheich *et al.*, *ibid.*) and its second and third harmonics. Just such a three-harmonic-component spectral pattern, with a fundamental at about 300 Hz, is common in budgerigars' speech imitations. See, for example, Figure 3.1J. A fundamental of 300 Hz is near the middle of the range of female speech fundamental frequencies.

In spite of the demonstrated selectivity for acoustic cues of mid-brain auditory units, discrimination of call types at this level was poor (Scheich *et al.*, *ibid.*). Units responded equally to different call types having at least one common cue (e.g., similar harmonic spectra). The authors speculated that generalization, and thus recognition of call types, might be a function of the telencephalic auditory centres.

(Biederman-Thorson, 1970, had found that auditory units in field L of the ring dove, unlike units in *nucleus ovoidalis*, were more responsive to certain complex acoustic stimuli than to pure tones. Leppelsack and Vogt, 1976, recorded the responses of telencephalic auditory neurons in the starling, *Sturnus vulgaris*, to species-specific vocalizations and found that some neurons responded to a single parameter, others, in some cases referred to as "feature detectors" - see "Feature detection" below, this chapter - responded to a combination of parameters, and yet others responded to the entire complex of parameters describing the sound.)

Bonke, Scheich and Langner (1979) and Scheich *et al.* (1979) confirmed the predictions about telencephalic discrimination of call types in the Guinea fowl by locating and investigating auditory units in field L which are responsive only to a particular call type - the "iambus" call (described in detail by Maier, 1982). Bonke, Scheich and Langner (1979) found that units in L_2, which receives most of the direct input from *nucleus ovoidalis* (see above), did not distinguish iambus calls

from other calls with at least one similar acoustic cue, while certain units in L_1 and L_3 were highly selective for iambus calls. (This lends weight to the argument that L_1 and L_3 may be "higher order auditory areas" - Bonke, Bonke and Scheich, 1979.)

"Formant" detectors. Scheich *et al.* (1979) investigated the basis of this telencephalic selectivity for iambus calls. They found that iambus-responsive units were concentrated in that area of the "tonotopic map" in which the best frequency of units was 1-2 kHz. As this is the range of the dominant spectral band in iambus calls, it can be said that sensitivity to this "formant" plays an important part in iambus discrimination. However, Scheich *et al.* were careful to point out that "this sensitivity should not be confused with simple tuning, since units usually had several sensitive bands as shown by using tones and complex synthetic stimuli".

The iambus calls of *Numida meleagris* are often strikingly similar to certain speech-imitation sounds of the budgerigar discussed in detail in Chapter 3. (See, for example, the figures of budgerigar sounds on Plate 3.3, and compare with Guinea fowl iambus spectra in Maier, 1982, and Scheich *et al.*, 1979.) The spectra of these calls, in both species, consist of a dominant component (carrier frequency) at around 1-2 kHz, with two less intense peaks - one above and one below the dominant component - representing amplitude modulation of the carrier. (This pattern is often repeated higher in the spectrum around harmonics of the carrier.) Scheich *et al.* (*ibid.*) found that an additional cue for iambus recognition was the fine structure of the spectrum - i.e., the relationship of sidebands to carrier - within the dominant spectral band. Thus iambus-responsive units are responsive to the entire "frequency complex" (spectral envelope peak plus fine structure - see below, this chapter, under "Vowels") within the dominant band.

(Such complex feature extraction has also been shown to be an important aspect of the mammalian auditory system - e.g., Wollberg and Newman, 1972.)

Scheich, Bonke and Langner (1979) pointed out that while the dominant "formant" frequency of a call type explained the restriction of responsive neurons to a particular region on the tonotopic map, variations in other cues - such as the frequencies of subsidiary spectral envelope peaks - could determine preferences of particular units within the sensitive area for particular variants of the call type.

Scheich *et al.* (1977) had remarked on the significance of "formant" structure (relative position of spectral envelope peaks) in vertebrate vocalizations generally and, in this later paper (Scheich *et al.*, 1979), specific reference was made to the formant structure of human vowels and the ability of other species to discriminate among them. (See below, this chapter, under "Vowels".) "The preference of selective units for formants appears to be a relevant result with respect to the general question of how wide band calls may be analysed at upper levels of the auditory system" (Scheich *et al.*, *ibid.*).

Response to vowel stimuli. Scheich and his colleagues extended their investigations, in a manner of great significance for the present study, by observing the responses of neurons to both natural and synthetic vowel sounds in telencephalic field L of speech-trained mynahs. Langner *et al.* (1979) reported 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. Langner *et al.* (1981) elaborated on these initial observations. They reported that a unit whose two excitatory bands matched the two formants of synthetic vowel [e], and which responded to either of the formants alone as well as to the two-formant vowel, failed to respond to synthetic vowel [o],

even though [e] and [o] shared a common F1 frequency. They concluded that, as F2 of [o] lay between the unit's two excitatory bands, it was probably inhibitory. They also found units which responded categorically to a [ba]-[ga] continuum.

Langner *et al.* (*ibid.*) presented further data illustrating selectivity of field L neurons in vowel recognition. They reported tonotopic organization of auditory neurons, and a predominance of vowel-selective neurons in L_1 and L_3.

They also investigated the significance of relative formant amplitudes, finding that although responses were best if the A1:A2 ratio was less than 15 dB, responses did occur to stimuli having an A1:A2 ratio of up to 60 dB. (In some units in which one formant of a stimulus excited a response and the other formant was inhibitory, response was still strong when the A1:A2 ratio was 0; in other units, response was abolished when the amplitude of the inhibitory formant rose to a level 30 dB below that of the excitatory formant.)

These authors (Langner *et al.*, *ibid.*) discovered a variety of excitatory-inhibitory interaction patterns in the responses of individual neurons to two-formant stimuli. When plotted against axes calibrated to represent F1 frequencies (abscissa) and F2 frequencies (ordinate), each unit's responsive area formed a rectangle with each boundary parallel to one or other of the axes. They interpreted this as indicating independent neural assessment of each of the two formants. (Psychophysical tests with human listeners using the same two-formant stimuli produced vowel-discrimination boundaries which were similarly parallel to either the F1 or the F2 axis - Hose, Langner and Scheich, 1983 - indicating a similar neural perception mechanism for these human sounds in mynah and in man.)

They concluded that "the increased selectivity for vowels in L_1 and L_3 is related to increasing inhibition

and multiple tuning of units in these layers" (Langner *et al.*, 1981).

Such demonstrations of human-like neural response patterns in the perception of vowel sounds by a bird (albeit a "talking" bird) have important implications for the development of this thesis. In order to understand these implications more fully, it will be necessary to take a general look at the acoustic patterns of speech, and their processing by human and non-human auditory systems.

Vowels

Spectral features. "The parameters in complex sounds which are available for conveying information are of two major types: temporal and spectral" (Newman *et al.*, 1977). In speech, the vowel sounds have traditionally been regarded as discriminable on the basis of static spectral features. (See Rosen and Fourcin, 1986, for a review and critique of this position.) The study of Langner *et al.* (1981), discussed immediately above, described neuronal response to vowel sounds in telencephalic field L of *Gracula religiosa* as a function of spectral peak frequencies. In the synthetic stimuli used by Langner *et al.*, these spectral peaks corresponded to the first and second formant frequencies of natural vowels.

Animal studies. 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). Discrimination of vowel sounds by non-human mammals has also been shown to be based, as in human listeners, on relative formant frequencies. (See Miller, 1977, for a review.)

Burdick and Miller (1975) trained chinchillas to distinguish between natural [a] and [i] (despite speaker sex difference, and therefore differences in absolute formant frequencies among the tokens), and found that they transferred successfully to the labelling of synthetic stimuli which differed only in relative formant frequencies. Baru (1975) trained dogs to distinguish between four-formant synthesized [a] and [i], and found

that this discrimination was unimpaired either by changes in fundamental frequency, signal intensity, or duration (within specified limits), or by removal of the third and fourth formants. He also found that the dogs transferred well to stimuli synthesized according to formant frequencies typical of female speech, and to randomized presentation of the male and female stimuli. Both these studies concluded that the animals used primarily the same cue as human listeners - relative frequencies of the first and second formants - in distinguishing vowels.

(Dewson, 1964, trained cats, and Dewson *et al.*, 1969, trained monkeys to discriminate between natural vowel sounds.)

Klatt and Stefanski (1974) and Thorpe (1967) noted that a mynah's speech imitation involves such relative (rather than absolute) formant frequency relationships, and the detailed analyses in Chapter 6, below, provide further evidence of this.

The ability of birds to reproduce perceptually important contrasts in formant relationships is perhaps most evident in their imitations of the diphthongs - vowels of changing quality. For instance, the budgerigar's [ɑ] in [pɑɪ] (Fig. 6.9G) and [krɑɪ] (Fig. 6.9Q) follows the human pattern in its steeply rising F2 transition from [ɑ] to [ɪ], and the simultaneously falling F1 transition (which disappears, however, before an appropriate level for F1 of [ɪ] is reached - a feature of budgerigar [ɪ] imitations discussed, particularly, in Chapter 5). A more completely speech-like pattern for [ɑɪ] can be seen in the grey parrot's version illustrated in Figure 6.6B. (Although the frequency scale is more compressed in this figure, and the closely-spaced F1 and F2 of [ɑ] appear like a single band of energy centred on about 1 kHz, the divergence of these two formants for [ɪ] can be clearly seen.)

Ambiguities in interpretation: spectral envelope shape. Many avian vocalizations are, like human speech, wide-bandwidth sounds which show consistent patterning both of spectral envelope peaks and of individual frequency components. Scheich *et al.* (1977), in their study of wide-band calls by the Guinea fowl (*Numida meleagris*), referred to the spectra of these sounds as "frequency complexes". Parameters used in the specification of such a spectral "complex" include those associated with total bandwidth, fine structure (e.g. frequencies of harmonics or sidebands), and overall energy distribution (i.e. the shape of the spectral envelope). Scheich *et al.* (1979) attributed the fine structure of such complex spectra to the vocal source mechanism, and the envelope shape to acoustic resonances of the vocal tract. In this they followed the source-filter model of human speech production established by Fant (1960). (See Chapter 2 for a discussion of acoustic resonances in the avian vocal tract.) It has been shown, however, that multiple peaks in the spectral envelope of some bird sounds can be source-generated - see particularly Chapters 3 and 6.

Such source-produced frequency complexes are susceptible to perceptual processing by human listeners according to the same patterns as the acoustically-filtered complexes of speech. This perceptual ambiguity is exploited by "talking" birds.

Ambiguities in interpretation: fine structure. The source-generated harmonics which constitute the fine structure of human vowel spectra need not be present for the correct identification of the sound. In whispering, the harmonic spectra of vowels are replaced by glottal noise energy. Some "talking" birds - mainly the larger parrots, which make the greatest use of acoustic resonance (see Chapter 2) - can reproduce human

whispering (Chapter 6, Fig.6.5G), while many more can exploit this aspect of spectral redundancy in their imitation of vowels. The budgerigar's [ɪ], for instance, has a noisy spectrum, in which the vowel-like spectral envelope shape (of F2 and higher formants) compensates for the irregularity of fine structure (Fig. 6.8H). Compare this with the spectrum of the corresponding mynah utterance (Fig. 6.8G), in which the harmonic fine structure is the most striking feature and the envelope peaks are relatively small. Now turn to a spectrum of the same vowel sound produced by a human speaker - Figure 6.3F: here there is both an harmonic fine structure and a strong formant pattern.

Variations in the fine structure of human voiced spectra depend on the fundamental frequency of glottal excitation, and the mode of vibration of the vocal folds. That birds are capable of producing similar variations is demonstrated by the analyses on Plate 3.2 illustrating variable reproduction of the word "Hello" by a raven. Figure 3.2G is a section through the voiced transition from [l] to [əʊ] in one version of the word. The spectrum shows little apparent harmonic structure, and considerable noise energy - reminiscent of human irregular and "breathy" voicing. Figure 3.2H is a section through the [əʊ] sound of another version of the word, showing the high-frequency fundamental, well-defined harmonic spectrum and absence of noise energy typical of a rather falsetto-like human register.

Irregularities of "harmonic" structure in Figure 3.2G indicate that the production mechanism in this case is not similar to that of the human voice. (See Chapters 2 and 6 for discussions of the oscine sound-production mechanism.)

Ambiguities in the perceptual interpretation of spectral fine structure by human listeners, as in the interpretation of envelope shape, allow the bird to employ such a non-human mechanism successfully. For

instance, it has been shown in Chapter 3 that a budgerigar can create the aural illusion of an harmonic spectrum by modulating a carrier frequency. This can mean the absence of many "harmonic" components, including the "fundamental" (as in Fig. 3.2A), from the spectrum. These components are thus seen to be non-essential in the gross identification of the sound. The bird, in effect, exploits this redundancy in the fine structure of the vowel spectrum just as it exploits redundancy in the configuration of the spectral envelope.

Patterning and redundancy in vowel sounds. The data of Peterson and Barney (1952) provided a framework for subsequent research on the pattern processing of spectral envelope shape in vowel perception. These authors made a survey of the extent to which natural vowel sounds could be specified in terms of envelope peaks representing the first and second formants. They plotted second-formant frequency as a function of first-formant frequency for ten vowels each produced by 76 speakers. They found that most instances of each of the ten vowel sounds fell within one of ten discrete areas on the graph (although there was some overlapping where "the distribution of measured formant values [did] . . . not correspond closely to the distributions of phonetic values" - Peterson and Barney, *ibid.*). This confinement of most vocalic tokens within discrete areas on the F1/F2 graph demonstrates the relative constancy of the F1/F2 ratio. The importance of the F1/F2 ratio - rather than absolute formant frequencies - in identifying vowels is evident in the phenomenon of normalization (Fourcin, 1968, 1980), which allows vowel sounds produced by speakers with differing vocal tract dimensions - i.e., men, women and children - to be identified as realizations of the same phoneme.

The redundancy of the higher formants in vowel identification has been demonstrated in many perceptual

experiments using synthetic two-formant vowels. (See, for example, Joos, 1948; Delattre *et al.*, 1952.) The capacity of vowels to be represented by two-formant synthetic sounds was studied in detail by Carlson *et al.* (1975). They found that "all Swedish vowels could be matched by two-formant approximations", and that the "F2'" of these approximations "was placed close to F2 in back and midvowels, in between F2 and F3 in non-high or rounded front vowels, and in the region of F3 or higher for a typical [i] vowel"[5]. Thus the importance of the two-formant model can be seen in the human perception of vowel sounds, and an insight can be gained into the ability of birds to imitate these sounds by using a simplification of their formant structure.

This simplification reaches its extreme in those imitations in which all the spectral energy in concentrated in "F2'" (or simply F2). Delattre *et al.* (1952) remarked that [i], with its extremely high F2, "can be very closely approximated by a single high-frequency formant". (See the budgerigar's imitation of high front vowels figured and discussed in Chapter 5, and the discussion of relationships between F2 and perceived pitch in Chapter 6.)

Stop consonants

Complexity and redundancy in the temporal domain.
Reaction to a single transient sound - a pulse - can be interspecific, the transient taking on the significance of a particular species-specific signal for members of each species affected (Busnel, 1968). Pulse repetition rate, however, "is a very specific characteristic of a signal and may even serve in taxonomic determination" (Busnel, *ibid.*).

In many cases, the total duration of a pulse train (i.e., the number of pulses) seems to be irrelevant to the message. Thus two apparently redundant temporal

parameters - the overall duration of the call sequence and the number of constituent pulses - have been introduced with the increased complexity of pulse repetition (and the meaningful parameter of repetition rate).

The inter-pulse interval can be systematically varied, to produce what Brackenbury (1978a), in discussing the pulsed songs of the grasshopper warbler (*Locustella naevia*) and the sedge warbler (*Acrocephalus schoenobenus*), referred to as a "rhythmic pulse code". Even in sound sequences consisting of non-pulsatile elements of varying duration, "the interval separating the calls has also a semantic value" (Busnel, 1963). For instance, Emlen (1972) found that altering the duration of intervals between the elements of recorded indigo bunting (*Passerina cynea*) song had a much more deleterious effect on species recognition than any other form of signal alteration he tried (such as rearrangement and selective deletion of elements). Jouventrin *et al.* (1979) showed that partner-recognition among emperor penguins (*Aptenodytes forsteri*) is possible when only part of the partner's song is played back, but that "modification of the song's rhythm (i.e. temporal phrasing) prevents all identification"[6].

Voice onset time. Analysis of phrase and sentence imitations by a budgerigar, discussed in Chapter 6, has shown that while the overall duration of an imitation often differs from that of its human model, the relationship of temporally-perceived features within the utterance (i.e., its rhythm) is usually well preserved and, moreover, some individual temporal features are accurately reproduced. (In the human/budgerigar comparison figured on Plate 6.2, both overall duration, and the duration of component segments, are of almost equal duration. See the discussion of this plate in Chapter 6.) The most outstanding example of precise

imitation of a temporal feature is in voice onset time (VOT).

Voice onset time, as observable in physical analyses of speech sounds in the time domain, is an acoustic consequence of differences in relative timing of laryngeal and supralaryngeal articulations in syllable-initial prestressed stop consonants (Lisker and Abramson, 1964). On spectrograms of English stop consonants, this difference is most clearly reflected in the interval between the plosive burst and the subsequent onset of F1. (The spectral components of aspiration noise are of relatively low amplitude at F1 frequencies, and are usually invisible on spectrograms - Rosen and Fourcin, 1986.)

Labelling tests with synthetic stimuli by Liberman *et al.* (1958) showed that a three-formant synthetic sound perceived as a voiced-plosive-initial syllable, such as [ba], was perceived as [pa] after "cutback" of the rising first-formant transition by an apparently critical interval (between 20 and 30 msec). While they recognized the possibility of spectral cues (e.g. differences in onset frequency of the cut-back F1 transition) for the voicing distinction, a further experiment showed that the temporal cue alone (i.e. delay in the onset of the complete F1 transition relative to F2 and F3) was sufficient for the distinction in at least some consonant-vowel combinations.

Using the stimuli from this second labelling experiment (i.e., with VOT as the only variable), Liberman *et al.* (1961) found a peak in the discrimination function which corresponded to the previously determined labelling boundary. Such a correspondence of labelling and discrimination data is regarded as sufficient evidence of categorical perception (Studdert-Kennedy *et al.*, 1970).

Spectral cues for the voicing distinction. Liberman *et al.* (1958), and before them Cooper *et al.* (1952), had observed that "the transitions of the first formant appear to contribute to voicing of the stop consonants" (Cooper *et al.*, *Ibid.*). More recently, attention has been increasingly directed towards the possible significance of such spectral cues in the voicing distinction. Stevens and Klatt (1974) suggested that the mere presence or absence of an F1 transition might be the primary cue; Lisker (1975) and Summerfield and Haggard (1977) investigated the significance of F1 onset frequency. (The rising F1 transition is related to the transition of the vocal tract from the stop to the vocalic configuration.)

The Summerfield and Haggard experiments indicated that F1 onset frequency did contribute to the perceived distinction. Indeed, their results showed that this spectral cue and the temporal cue of VOT were complementary - i.e., there appeared to be an inverse relationship, in plosives categorized as voiceless, between F1 onset frequency and VOT. (Covariation of VOT and F1 onset frequency is a result of source-filter interaction - Rosen and Fourcin, 1986.)

Soli (1983), on the basis of his perceptual experiments, went even further in asserting that spectral cues predominate in the categorization of these sounds. (See Rosen and Howell, 1987, for a survey of the continuing debate on cues for the voicing distinction.)

Avian preference for the temporal voicing cue. The study of "talking" birds can contribute to this debate by pointing to temporal and/or spectral features thought to be related to the voicing distinction which are preserved in discriminable avian imitations (and those which are omitted with no apparent loss of intelligibility).

For example, the budgerigar, as will be shown in Chapters 5 and 6, has difficulty with F1 - omitting low

F1 altogether (as in [i] - see Figures 6.2G and 6.2H for a comparison of human and *Melopsittacus* [tʃiki]) and, in imitations of sounds with higher F1, raising its frequency still further. Voice onset time, on the other hand, is precisely reproduced by the bird.

(Indeed, it seems possible from a consideration of avian audiograms - discussed above - that spectral cues in the first-formant range are generally less significant for birds than for man. On the other hand, on the basis of psychophysical studies such as that of budgerigar amplitude modulation thresholds by Dooling and Searcy, 1981, temporal cues are probably more significant. Some birds may perceive the fundamental frequencies of speech as pulsed bursts or amplitude modulations of a second-formant carrier - see below, this chapter, and Chapter 6.)

For an example of the extreme condition of a budgerigar's reliance on the temporal (VOT) cue in its production of the voicing distinction, see the second [pɔdʒi] in Figures 3.1E and 3.1H, in which F1 is entirely absent. Here there is an obvious interval of minimal amplitude at around 1 kHz between the initial plosive burst and the onset of the intense second "formant", the onset time (about 50 msec) being appropriate for [p]. Note that in, for example, [pu] (Fig. 6.3L), there is an imitation of aspiration noise, following the human model (Fig. 6.3K), at second-formant frequencies before the onset of F1; in [pɔ], with no F1, there is no apparent aspiration noise at F2 frequencies during the onset of "voicing" - the interval of virtual silence, transferred from F1 to F2, being the preferred cue.

Although it would be necessary to conduct labelling tests with a continuum of such "F2-delayed-onset" sounds in order to demonstrate conclusively the relative importance of temporal and spectral acoustic cues for both man and bird, the mere fact that an imitation is intelligible in the complete absence of the relevant

spectral cues is good evidence for the greater significance of temporal factors in the voicing distinction.

Systematic VOT variations. VOT boundaries for the voicing distinction have been found to vary according to whether the voiced/voiceless pair are bilabials [b/p], the alveolars [d/t], or the velars [g/k]. Liberman *et al.* (1958) reported the shortest critical VOT for bilabials. Subsequent researchers (e.g., Abramson and Lisker, 1970, 1973) have agreed that the bilabial, alveolar and velar boundaries - typically in that order - are located between about 20 and 45 msec VOT.

Such systematic variation in VOT according to place of production can be observed in avian imitations. For instance, in the phrase [pʊsi kæʔ pʊsi kæʔ] (Plate 6.3) the human trainer's VOT figures of 47 msec for the voiceless bilabial [p] and 53 msec for the voiceless velar [k] are accurately reproduced by the bird. (See Figures 6.3A and 6.3B for trainer and bird respectively, and Figures 6.3K and 6.3L for time-expanded versions of the second [pʊsi kæʔ] in each case. These figures are discussed in greater detail in Chapter 6.)

Longer onset times can also be faithfully reproduced: see, for example, the rendition of voiceless velar [k] in [tʃiki] by a mynah in Figure 6.8C, VOT being 111 msec, and compare with his trainer's version of the same sound, having a VOT of 100 msec, in Figure 6.7G.

Cross-language and infant studies. That a critical VOT of around 20 msec could be common to all languages which have chosen to use the voiced/voiceless distinction as a phonemic contrast is indicted in the comparative study of Lisker and Abramson, 1964. Eimas *et al.*, 1971, showed that this pattern applies in the perception of speech by humans who are not even speakers - i.e., pre-speech infants. Lasky *et al.*, 1975, Streeter, 1976, and Eilers

et al., 1979, provided evidence for a universal human sensitivity to this critical onset time by demonstrating that infants exposed to only Spanish or Kikuyu - a language which, like Spanish, employs the prevoiced/voiced stop distinction rather than the voiced/voiceless distinction of English - showed increased discriminative sensitivity across the non-native boundary distinguishing English [b] from [p].)

Animal studies. Behavioural experiments with non-human mammals - rhesus monkeys (Waters and Wilson, 1976), macaques (Kuhl and Padden, 1982,), and chinchillas (Kuhl, 1976, 1979, 1981; Kuhl and Miller, 1975a, 1975b, 1978) - have shown responses to synthesized (voiced/voiceless initial plosive) VOT continua, in both labelling and discrimination tests, similar to those for human subjects. (See reviews by Kuhl, 1987; J. D. Miller, 1977; Snowdon, 1979.) It seems highly likely, therefore, that the animals were responding to the same cue or cues as human listeners. In all the animal experiments, the test stimuli used were of the "F1 cutback" variety (Liberman *et al.*, 1958), and so spectral as well as temporal cues could have been involved in labelling and discrimination (see above). However, the correspondence of these results from entirely non-linguistic listeners with the results from tests of human subjects using non-speech stimuli varying only in onset time (see above) strongly suggests the involvement of a psychophysical cue related to VOT.

The agreement between labelling and discrimination data for the chinchilla (Kuhl, 1981) and the macaque (Kuhl and Padden, 1982) is considered sufficient evidence (Studdert-Kennedy *et al.*, 1970) of categorical (but, or course, non-phonetic) perception of these speech sounds by a rodent and a monkey.

The data of Kuhl and Miller (1975b, 1978) showed a shift in the VOT boundary for the voicing distinction in chinchilla perception according to place of production

similar to that reported for man - see above. (Waters and Wilson, 1976, however, found a contrary variation in VOT in speech perception by rhesus monkeys: bilabial boundaries were higher than velar boundaries.)

That a temporal cue similar to VOT in speech may be used in the communication sounds of some non-human primates is suggested by the data of Green (1975) on Japanese macaques (*Macaca fuscata*) and Lillehei and Snowdon (1978) on stumptail macaques (*Macaca arctoides*).

Perception of non-speech stimuli. Perceptual tests on the relative onset time of non-speech sounds have been widely interpreted as supporting the case for VOT as an important cue in the voicing distinction.

The work of Hirsh (1959) on the perception of temporal order had shown that, while an interval of only about 2 msec is sufficient for the perception of two short non-simultaneous sounds as separate events, an interval of about 20 msec is necessary for the correct determination of their temporal order. This interval which, according to Hirsh, is independent of the spectral qualities of the sounds involved, is close to that of the VOT voicing boundary as determined in discrimination tests such as those of Liberman *et al.* (1961).

In the light of Hirsh's work, and of published VOT data, Miller *et al.* (1976) and Pisoni (1977) carried out experiments on the perception of onset time differences in non-speech sounds. Miller *et al.* used a wide band of noise and a complex periodic sound (a "buzz"), and varied the time by which the onset of the noise preceded that of the "buzz" within the range of -10 to +80 msec; Pisoni used two tones in an otherwise similarly designed experiment. In both experiments, the two components of each stimulus, although having variable relative onset, ended simultaneously.

Similar results were obtained in both studies: a difference of about 20 msec was required before the non-

simultaneity of onset was perceptible.

In effect, these two experiments (like that of Hirsh, 1959) are to do with the perception of temporal order - of one event leading or preceding another (Kuhl, 1981). As Hirsh remarked: "The length of the required temporal interval suggests that the judgment of order requires other mechanisms than those asociated with the peripheral auditory system" (Hirsh, 1959). All these experiments, then, indicate that the VOT boundary associated with the voiced/voiceless distinction could reflect an inbuilt limit of the human central nervous system in the perception of temporal order - a psychophysical rather than a purely linguistic cue'. The animal studies reviewed above suggest that this cue could be interspecific. (The possible involvement of such cues in the evolution of communiction sounds, and thus their entitlement to the designation of "bio-communication cues", will be discussed below under "Feature detection".)

Perceptual tests on pre-speech infants using two-tone stimuli with differing relative onset time (Jusczyk *et al.*, 1980) revealed three well-defined categories along the continuum "corresponding to leading, simultaneous, and lagging events" (Jusczyk *et al.*, *ibid.*). These results provide further support for the hypothesis that the perception of VOT is related to the perception of the temporal order of component events within the acoustic stimulus.

The possibility remains, however, that a more peripheral mechanism - simply a limitation of the auditory system's temporal resolving power - is involved. This limitation appears to vary with the stimuli involved (Divenyi and Sachs, 1978). Stevens and Klatt (1974), using an initial noise burst and a succeeding "buzz" (having a complex periodic waveform) - a stimulus basically similar to a plosive-initial speech syllable - found that the noise burst and the "buzz" could not be

perceived as having separate onset at intervals of less than about 20 msec. Some authors (e.g. Bailey, 1979) have continued to believe that such a limit to temporal resolution is an important factor in the location of phoneme boundaries.

Speech-sound perception tests using birds as subjects could help clarify this issue. Auditory temporal resolution in birds has been shown (at least under some experimental conditions - see above) to be better than in mammals. Thus, for instance, an experimental demonstration of a categorical boundary in the avian perception of a VOT continuum at roughly the same place as the human boundary (and that for monkeys and chinchillas) would provide significant evidence of a more central, temporal order processing constraint underlying VOT.

Transition slope and direction. The three-formant synthetic syllables used by Kuhl and Miller (1978) in their experiments with chinchillas, following the specifications of Abramson and Lisker (1970) for stimuli in similar experiments with human listeners, differed among bilabial, alveolar and velar initial plosives in both the slope of the rising F1 transition and the direction and slope of the F2 and F3 transitions. Earlier perceptual experiments with human subjects, such as those reported by Liberman *et al.* (1954) and Liberman (1970), had shown that formant transitions - particularly that of F2 - could convey sufficient information to distinguish place of production.

Morse and Snowdon (1975) used three-formant synthetic syllables in which the starting frequencies of F2 and F3 differed for stimuli representing [bæ], [dæ] and [gæ]. They found that their rhesus monkey subjects were better able to discriminate between stimuli from different phoneme categories than between acoustically different stimuli from one category[8].

In their discrimination tests with macaques, Kuhl and Padden (1983) used speech-sound pairs drawn from a synthetic two-formant continuum. The acoustic feature varied was the onset frequency of the second-formant transition. Their results "demonstrated that peaks in the discrimination functions occur for macaques at the 'phonetic boundaries' which separate the /b-d/ and /d-g/ categories for human listeners" (Kuhl and Padden, *ibid.*).

Rowell and Hinde (1962) noted a resemblance between differences in the initial slopes of calls by the rhesus macaque, and initial second-formant transitions in speech.

Miller and Kuhl (1976) tested chinchillas' discrimination among syllables of natural speech, and found that bV/dV was an easy discrimination, while dV/gV was "nearly impossible". The two possible explanations for this are that differences in VOT are greater for bV/dV, and that the bV/dV contrast has a rising F2 transition in bV.

Transition length. Liberman *et al.* (1956) varied the length of initial transitions in two-formant synthetic syllables, and found that they ceased to be perceived as [be] after the transitions reached durations of between 30 and 40 msec, and were perceived as [we]. ([we] became [ue] after transitions reached about 50 msec.) In syllables with the same vowel (i.e., the same formant frequencies), changing the duration of transitions necessarily involves changing their slope. In order to test whether duration or slope was the more significant cue in their categorization, Liberman *et al.* tried the same experiment using a variety of vowels, and concluded that transition duration rather than slope is the significant cue. (Their results were basically confirmed by Mack and Blumstein, 1983, who argued, however, that the pattern of the amplitude envelope was the critical cue to the stop/glide contrast.)

Nabelek and Hirsh (1969) investigated the perception of frequency glides (i.e., transitions), and found that discrimination among transitions differing in rate of change was best for transitions around 30 msec duration. In explanation of this, they postulated that there are two mechanisms involved in discriminating the rate of frequency change - one which discriminates fast changes (below 30 msec) covering relatively large frequency intervals, as in some speech sounds, and one which discriminates slower changes, especially within smaller frequency intervals. Heightened sensitivity to the formant transitions of speech, which are typically about 30 msec in duration, would thus be a result of this discontinuity in perceptual mechanisms. It would thus be another possible example - like VOT - of essential contrasts in speech patterns being grouped around such discontinuities (Stevens, 1972).

The data of Nabelek and Hirsh support those of Pollack (1968), who found that at very short transition times (below about 30 msec) sensitivity to changes in the rate of transition was directly proportional to the frequency range subtended, while at longer transition times this sensitivity was almost independent of frequency range. Thus these data on noticeable differences in rate of frequency change, and those on the discrimintion of such changes, all point to the employment of two distinct modes of operation in the perception of such transitions.

The significance of birds' impressively precise reproduction of speech formant transitions is discussed below, this chapter, and in Chapters 5 and 6. At the neurophysiological level of single-unit response, Whitfield and Evans (1965) have reported that some auditory cortical neurons in cats respond to frequency sweeps, and Gross and Anderson (1976 - see above) found some neurons in the eighth nerve of the pigeon which responded solely to frequency sweeps.

Discrete and graded signals

Spectrographic analysis of bioacoustic communication sounds - particularly, in this respect, those of birds and non-human primates - has revealed a distinction between (1) acoustically "discrete" systems, comprising a repertoire of highly stereotyped and mutually distinct signals, and (2) acoustically "graded" systems, in which certain tokens of the repertoire - with respect to a particular physical parameter - can be regarded as adjacent components of an acoustic continuum. (See Marler, 1975b, 1977b for reviews and general discussion. For a more recent discussion on grading and discreteness in bird sounds in particular, see Morton, 1982.)

It must be emphasised that the discrete/graded distinction is based on physical analyses of recorded animal vocalizations as interpreted by human investigators (Snowdon and Pola, 1978). Whether species-specific vocalizations are perceived continuously or categorically - see below, next section - by individuals of the species concerned can be determined only by techniques such as playback experiments (Snowdon and Pola, *ibid.*).

(In spite of their introductory cautions, Snowdon and Pola themselves, in the course of discussion, fall into the trap of confusing the term "graded" - which refers to acoustic structure - with "continuous" - which refers to mode of perception.)

The acoustic communication systems of some species appear to include both physically discrete and graded elements (Gautier and Gautier, 1977; Morton, 1982; Oppenheimer, 1977; Rowell and Hinde, 1962).

Gautier and Gautier (1977) and Marler (1975b, 1977b) associated discrete signals among primates with inter-group (i.e. long-range) communication, and graded signals with the subtler, close-range communication possible

(and, perhaps, necessary) within the complex structure of a large social group. A comparable distinction is evident in avian vocalizations (Marler, 1976). For example, the data of Morton (1982) show an inverse relationship between structural variation, and the distance from sender to receiver, in vocalizations of the Carolina wren (*Thryothorus ludovicianus*).

Acoustic grading and vocal plasticity. Graded vocal repertoires - in species, at least (such as *Macaca fuscata* - Green, 1975), in which such slight signal variations have been correlated with consistent differences in the message conveyed - imply both flexibility and precision in neuromuscular control of the sound-producing organs.

It was argued in Chapters 1 and 3 that such flexibility and precision in psittacine sound production can be attributed to the activity of the evolutionarily "advanced" intrinsic syringeal muscles.

It is possible that slight acoustic differences between repetitions of the [ɔ] sound in Figure 3.2C could reflect gradations of species-specific *Melopsittacus* vocalizations particularly associated - by speech-trained birds - with the spectral pattern of [ɔ]. (A detailed study of species-specific *Melopsittacus* vocalizations is called for in this connexion.) That such slight acoustic variations in tokens of budgerigar "speech" reflect gradations of species-specific significance is indicated by the highly regular patterns of such variations in repetitions of a long speech phrase, as analysed in the Appendix to Chapter 6.

Vocal imitation is another avian phenomenon which has been attributed - on the motor level - to the activity of intrinsic syringeal muscles (Nottebohm, 1972b; Gaunt, 1983; see Chapter 1). There have been very few reports of imitative learning in species other than those possessing such muscles - i.e., the psittacids and passerines

(Gaunt, 1983; see "Vocal plasticity and neural complexity" in Chapter 1).

Vocal imitation, like gradation of signals, is thought to have its origin in social, as opposed to territorial, communiction (Bertram, 1970; Nottebohm, 1970; Thorpe, 1969; Thorpe and North, 1965; see "Social factors" in Chapter 1).

(All the authors mentioned in the previous paragraph, in discussing the social function of vocal imitation and plasticity, give special mention to the speech-imitation abilities of parrots and mynahs.)

This connexion between graded species-specific signalling and speech imitation in birds is of particular interest in the present context because speech itself is a predominantly graded system (Liberman *et al.*, 1967).

Continuous and categorical perception

A graded series of acoustic stimuli can be perceived either continuously, as in human vowel perception, or categorically, as are the stop consonants (Liberman *et al.., ibid.*).

Morton (1977) presented a convincing case for believing that some of the principal acoustic parameters along which continuously-perceived continua vary are interspecific. He argued that natural selection has resulted in the structural convergence of animal sounds used in similar motivational contexts. Unfortunately there is some confusion in his discussion between structural (e.g. spectral) and perceived (e.g. qualitative) features: one of the "structural" features he deals with is "harsh quality" (presumably as perceived by human listeners). Nevertheless, it is clear from his evidence that sounds produced by many birds and mammals in the motivational context of aggression have physical characteristics related to relatively low perceived pitch and harsh quality; at the opposite extreme, sounds

produced in the context of submission have physical characteristics related to relatively high perceived pitch and tone-like quality. The physical parameters relating to pitch and quality can be varied independently, expressing subtle gradations between these two extremes.

Another apparently interspecific graded parameter dealt with by Morton is the slope of frequency transitions: sounds of rising frequency are associated with increasing appeasement or fear; sounds of falling frequency with increasing hostility.

In categorical perception, on the other hand, an acoustic continuum is segmented into a limited number of categories, the constituents of each category being - according, at least, to the classical definition[9] - indiscriminable one from another. Human perception of stop consonants is the best-documented example of categorical perception. (See Rosen and Howell, 1987, for a recent review.)

Categorical perception of species-specific sounds has been reported in several mammalian species (reviewed by Ehret, 1987).

A useful way of considering continuous and categorical perception is in terms of "analogue" and "digital" models of information processing. Thus, in continuous perception, a token is perceived according to an analogue process: it is related to the two extremes of a continuum, and given a uniquely-defined grading which is an analogue of that relationship. On the other hand, in categorical perception, the token is labelled simply as falling on one or other side of a categorical boundary, thus following the digital processing model.

Marler (1975b) proposed that the categorical perception of certain speech-sound continua in man represents, paradoxically, a reversion to more "primitive" neural mechanisms - associated with the processing of discrete signals - in the perception of

graded communication sounds. This new emphasis on acoustic categories rather than on semantically significant gradations would have made possible the purely phonemic contrasts necessary for the development of language. Snowdon (1987), commenting on Marler's theory, remarked that "although the details of his phylogenetic classification have been subsequently shown to be wrong (he argued that human beings were the only species to divide continua into categories), the general thrust of his argument is probably correct. The development of discrete categories along acoustic continua represents a sophisticated stage in the evolution of vocal communication" (Snowdon, *ibid.*).

(To return briefly to the analogue/digital analogy, it is significant, in the light of Marler's evolutionary theory, that digital processing is the more flexible system.)

Marler's evolutionary hypothesis at least provides a theoretical framework for the comparison of data on the categorical/continuous perception of speech sounds by non-human species (reviewed above), and data on those species' perception of their own communication sounds.

Feature detection

Much of the work reviewed above can be interpreted as pointing to the existence of constraints in auditory processing. Such auditory constraints could account for discontinuities in the perception of stimuli which constitute an acoustic continuum (Stevens, 1981). Such auditory discontinuities, by allowing binary contrasts within the range of an acoustic cue (such as VOT), could correspond, on the linguistic level, to phoneme boundaries. This contrastive function is thought to have led to the selective use in speech of acoustic cues which involve such discontinuities (Stevens, *ibid.*).

Such cues, when occurring contrastively in speech

sounds, have been termed "linguistic" or "phonetic" features (Eimas and Corbit, 1973), and the hypothetical auditory mechanisms, whose special sensitivity to such features is at least partly a function of their perceptual discontinuity, have been called "linguistic feature detectors" (Eimas and Corbit, *ibid.*). At the level of single-unit response to species-specific vocalizations, feature detectors have been reported in birds (Leppelsack and Vogt, 1976; Scheich, 1977). The animal experiments reviewed above have shown not only that non-human mammals are sensitive to such linguistically relevant features as VOT and formant transition slope, but that their perception of these cues can be marked by discontinuities similar to those found in human perception. Such "linguistic" features could thus be regarded, more broadly, as "bio-communication" features. (Snowdon, 1979, discussed possible correspondences between the speech cues involved in the animal tests and acoustic cues apparently of significance in the animals' own vocalizations.)

In evolutionary terms, an auditory discontinuity in an ancestral species could have led to the development of significant contrasts across that discontinuity in vocal communication sounds. The importance of those contrasts to communication - and hence survival - could have ensured the preservation of the original auditory discontinuity - and its significance to communication - throughout the subsequent ramifications of phylogeny.

While "bio-communication cues" used in categorical auditory perception may be the result of <u>divergent</u> evolution, as outlined immediately above, those used in continuous auditory perception - the "motivation-structural" cues of Morton (1977) - may be the result rather of <u>convergent</u> evolution, as argued by Morton (*ibid.*).

Non-human perception and reproduction of speech.

Feature detection can be regarded as selective auditory filtering (Green *et al.*, 1977). Avian speech imitation provides a comparative approach to the study of such filtering, as the birds' imitations of speech sounds reveal the results of filtering these sounds through a non-human auditory system.

Formant transitions - especially F2 transitions - are among the most conspicuously speech-like elements in analyses of avian speech imitation. Much of Chapter 5 is devoted to a description of the budgerigar's "chrp" - a call in the natural repertoire of the species, characterized by glides of rapidly changing frequency - and the transformation of this sound into speech-like [i] which, in medial and final position, involves a steeply-rising F2 transition.

Some imitations of whole phrases, especially by the smaller parrots, concentrate almost exclusively on the second formant and its transitions - capitalizing on the birds' natural virtuosity at glissando whistling. (See Chapter 2 for the physiology of whistling.) For example, Figure 5.2B shows a speech-imitation phrase which consists mainly of such "F2-transition" glissando elements, followed immediately by a whistled (non-speech) sequence incorporating all the glides used in the speech imitation. (See below, this chapter, and Chapter 5 for discussion.)

In human speech, with its complex and continuously changing formant patterns, there is no simple relationship (direct or inverse) between, for example, changes in F1 and changes in F2 frequencies. F1 of (voiced) plosive-initial syllables typically has a rising initial transition, reflecting the movement from the stop towards the vocalic configuration of the vocal tract. Initial F2 transitions, on the other hand, can either rise or fall, and have a variable slope, depending on the plosive's place of articulation and the quality of the

succeeding vowel (Delattre *et al.*, 1955). Birds can reproduce these complex patterns, even though their methods of producing them are very different from those involved in human speech. (See Chapter 2.) In the budgerigar's [pɑɪ], for example (Fig. 6.9G), F1 can be seen to rise initially (after an appropriate interval of VOT), while F2 maintains a virtually flat contour throughout the first half of the syllable.

(Another place cue in human speech perception is the frequency range of the plosive burst which is the spectral correlate of the release of the stop consonant. Liberman *et al.*, 1952, showed that, in synthetic stimuli without formant transitions, this cue was sufficient for place information. Its position is not fixed, however, as it varies in relation to the formant structure of the following vowel. There is a discussion in Chapter 5 of the way in which a budgerigar adjusts the frequency range of a noise burst in a non-speech sequence to produce the impression of an affricate in a speech-like sequence - Figure 5.2D.)

An important point to remember in any discussion of acoustic cues in speech perception is their variety and interdependent variability. As expressed by Rosen and Fourcin (1986):

"All speech sound contrasts can be based on a subset from a number of acoustic cues. The specific cues used and their relative importance vary within the particular context, and across listeners."

Neural lateralization

Lateralized perception of species-specific communication sounds in man and animals. The question of lateralized neural function is of particular significance in the context of avian vocal communication because of the demonstrated asymmetry of vocal control in songbirds. (See "Human/avian similarities in vocal ontogeny and

control" in Chapter 1.)

"Nottebohm's discovery that the neural control of some birdsongs is lateralized should alert us to the possibility of other non-human parallels to hemispheric dominance" (Marler, 1975b).

Marler (*ibid.*) continued: "It is important to find out whether nonhuman primates, especially those with graded vocal repertoires, show cerebral dominance, and if so, how vocal control and sound processing differ in the two hemispheres."

The results of primate studies which begin to answer these questions will be reviewed in the following section.

It is equally important to ask similar questions about avian processing of communication sounds. It was argued in Chapter 1 that, in "talking" birds, speech sounds are processed according to the same neural patterns as species-specific sounds. The present chapter has argued that such processing is facilitated by the presence in the speech signal of interspecific "bio-communication cues". Detection of a mechanism such as hemispheric lateralization in the processing of communication sounds would be of great value in testing these hypotheses.

As far as I know, there has been no experimental investigation of possible hemispheric lateralization in avian auditory perception[10].

The following discussion will review evidence of special, lateralized processing, in man and some non-human primates, of acoustic communication cues which the present study of avian "speech" indicates could be of comparable communicative significance to birds.

Broca's (1865) observations correlating left-hemisphere lesions and language disorders established language as a classic instance of cerebral dominance in human behaviour.

Many subsequent observations to this effect have been directly supported by experimental data such as those of

Wood *et al.* (1971), which showed that patterns of evoked potentials over the left hemisphere (but not over the right hemisphere) during auditory discrimination tasks involving linguistic criteria were significantly different from evoked-potential patterns during non-linguistic auditory discrimination.

The notion of left-hemisphere dominance in linguistic behaviour has been supported indirectly by perceptual experiments which demonstrate a right-ear advantage in both the recognition (e.g., Broadbent and Gregory, 1964) and recall (e.g., Kimura, 1961a) of speech-sound stimuli. (The link between right-ear advantage and left-hemisphere dominance in mammalian hearing has been demonstrated in experiments with cats by Rosenzweig, 1951, and Hall and Goldstein, 1968, which showed that monaural stimulation is more strongly represented in the contralateral than in the ipsilateral hemisphere.)

Several studies (e.g. Kimura, 1964) have revealed a left-ear (right hemisphere) advantage in the perception of certain non-speech stimuli. (See Molfese *et al.*, 1983, for review.)

Kimura's (1961b) tests indicated that the contralateral auditory pathway is dominant even when language is represented in the right hemisphere: in these subjects there is a left-ear advantage in the recall of speech-sound stimuli. Kimura (1961a) interpreted her results as reflecting an indirect auditory pathway from the disadvantaged ear - first to the contralateral (linguistically subordinate) hemisphere and then across the cerebral commissures to the dominant hemisphere where most linguistic processing takes place.

A study by Milner *et al.* (1968), on ear advantage in patients who had suffered complete midline section of the cerebral commissures, supported Kimura's hypothesis by showing that these patients were unable to report verbal input to the left ear. (This was in dichotic testing. It must be pointed out here that right-ear advantage - even

in subjects with complete surgical disconnection of the cerebral hemispheres - is observable only in dichotic testing: i.e., when different acoustic stimuli are presented simultaneously to the right and the left ears. With monaural stimulation, subjects show equal accuracy of report for the two ears. Dichotic presentation is thought to overload the perceptual system, resulting in suppression of the subordinate ipsilateral auditory pathways - see Studdert-Kennedy and Shankweiler, 1970, for references. The data of Springer, 1973, conform to this model by showing that reaction time in the recognition of dichotically presented speech syllables was faster for stimuli presented to the right ear - following the proposed direct contralateral pathway to the dominant hemisphere - than to the left ear - following an indirect contralateral then cross-hemisphere pathway. For a review of dichotic listening data from brain-damaged patients, see Molfese *et al.*, 1983.)

Perceptual experiments reported by Petersen *et al.* (1978) and Beecher *et al.* (1979) revealed a right-ear (left hemisphere) advantage in the discrimination of a communicatively relevant acoustic feature (identified by Green, 1975) in species-specific vocalizations by the Japanese macaque (*Macaca fuscata*). This feature is the relative position of a peak in the fundamental frequency contour of the call. On the other hand, the Japanese macaques discriminated among variations in the initial perceived pitch of the call - a communicatively irrelevant acoustic feature (Green, *ibid.*) - with either left-ear advantage or no ear advantage. Subjects belonging to several other Old World monkey species, in whose own vocalizations the placing of a frequency peak had no communicative significance, showed no ear advantage in the discrimination of this feature in the Japanese macaque calls. "The results suggest that Japanese macaques engage left-hemisphere processors for the analysis of communicatively significant sounds that

are analogous to the lateralized mechanisms used by humans listening to speech" (Petersen *et al.*, 1978).

Zoloth *et al.*, (1979) reported that discrimination performance by the Japanese macaques was best for the communicatively relevant peak-position feature, while performance by the other monkeys was best for the more acoustically prominent starting-pitch feature. "The species-specific predisposition of Japanese macaques to discriminate groups of vocal sounds on the basis of the somewhat complex and subtle cue of frequency peak position parallels the evidence from field study suggesting that they use this cue for encoding information in their natural communictive repertoire" (Zoloth *et al.*, *ibid.*).

Ehret (1987) discovered a left-hemisphere preference for recognition by mice of ultrasonic calls emitted by their young. "The results suggest that lateralization of this function evolved early in mammals and emphasize that an innte predisposition for perceiving communication sounds is connected with a left-hemisphere advantage in procesing them" (Ehret, *ibid.*)

Lateralization of temporal auditory processing in man and animals. There are some experimental data on the perception of non-verbal acoustic stimuli (e.g., Halperin *et al.*, 1973) which suggest that the left hemisphere predominates in the analysis of sequential patterning. For instance, trained musicians have a right-ear (left-hemisphere) advantage in the recognition of melodies, while musically naive listeners show a left-ear advantage (Bever and Chiarello, 1974; Gordon, 1975). These researchers interpreted their results as indicating left-hemisphere dominance for analytical processing (by trained musicians) and right-hemisphere dominance for "holistic" processing.

Such data have led some researchers to believe that left-hemisphere dominance for language represents a

special case of left-hemisphere dominance for sequential analysis generally. Thus, despite Tsunoda's (1969) finding of a right-ear advantage in the perception of Morse code signals by Morse code operators, and a left-ear advantage for naive listeners, the work of Papcun *et al.* (1974) on Morse code perception indicates that it is sequential analysis rather than information recognition which underlies this perceptual lateralization.

(Dewson, 1975, 1977, published experimental results demonstrating left-hemisphere dominance for short-term memory function in monkeys. There is an obvious connexion between short-term memory and sequential processing.)

The complex sequential structure of bird song, and the known neural lateralization of song production, may thus mean that birds are particularly well equipped with central neural pathways along which the sounds of speech can be processed, at a sub-linguistic level, in a manner analogous to that employed in human speech perception.

To complicate matters, it appears from auditory evoked response (AER) data that the right hemisphere of human subjects is dominant in the categorical perception of temporal duration in speech patterns. Molfese (1978a) recorded response variations over the right hemisphere during identification of synthesized stop-consonant stimuli which corresponded to the "manner" (voiced/voiceless) judgement of the subject. The stimuli differed only in VOT. Responses over the left hemisphere did not correspond to this categorical labelling distinction. That this is an auditory rather than a phonetic process was indicated by results of Molfese (1980a), who recorded a similarly lateralized AER component using the non-speech, two-tone stimuli of Pisoni (1977). (See above, under "Perception of non-speech stimuli".) In this case, initial bilateral processing of the temporal cue was followed by the right-lateralized process.

Morse *et al.* (1987) have recently recorded categorical

right-hemisphere responses during discrimination of voicing contrasts by rhesus monkeys.

"Monkeys as well as humans appear to process voicing contrasts categoricaly in the right hemisphere. This finding suggests that human and nonhuman primates have in common not only the behavioural but also the underlying neuro-cortical control processes involved in the perception of voicing contrasts" (Morse *et al.*, *ibid.*).

Another AER component recorded exclusively over the right hemisphere during speech perception by human subjects was correlated with another temporal duration cue - the length of the F2 transition in stop-initial CV syllables (Molfese *et al.*, 1985). This right-hemisphere temporal processing was in addition to left-lateralized "place" (i.e. formant-transition) discrimination. (See the next section, below.) "It would seem that . . . the left hemisphere actively discriminated the [b] from [d] stimuli while the right hemisphere somewhat later in time detected differences in the length of the consonant transitions" (Molfese *et al.*, *ibid.*).

Lateralization of transitional cues: "encoding" and "phonetic processing". Shankweiler and Studdert-Kennedy (1967a) found a significant right-ear advantage, both in recognition and retention, for synthetic[11] consonant-vowel (CV) syllables but not for steady-state synthetic vowels.

AER data have conformed with this result, giving a consistent picture of overall left-hemisphere dominance in the perception of stop consonants (see below) but no hemispheric asymmetry in the perception of vowels (Molfese and Erwin, 1981).

Studdert-Kennedy and Shankweiler (1970) reported that lateralization of stop consonants (in CV syllables) varied with manner (voiced/voiceless) and place (bilabial/alveolar/velar) of production. These authors related the difference between stop consonant perception

(right-ear advantage) and vowel perception (no right-ear advantage) to the corresponding dichotomy of categorical perception for stops and continuous perception for vowels (Liberman *et al.*, 1967). They also drew attention to the "encoded" nature of CV syllables (which is related to the phenomenon of categorical perception) - i.e., the lack of a one-to-one relationship between stimulus and percept.

Darwin (1971) and Haggard (1971) developed the connexion of "encoding" and right-ear advantage by showing that there was a right-ear advantage for steady-state vowels themselves if an element of encoding (sometimes called non-invariance) was introduced. They did this by using vowel stimuli produced by speakers with different vocal tract dimensions (men, women and children), the non-invariance thus lying in the differing values of formant frequencies. Haggard's experiments led him to believe that right-ear advantage for encoded stimuli depended simply on a recognized "trigger feature" rather than on the amount of decoding involved.

Despite the right-hemisphere components discussed above in connexion with temporal auditory processing, AER data support the notion of left-hemisphere phonetic processing of consonants (Molfese, 1980b; Molfese and Schmidt, 1983). In particular, such data have provided evidence of left-hemisphere processing of formant-transition (i.e. "place") cues in stop consonant perception (Molfese, 1978b, 1984).

The experimental results of Studdert-Kennedy *et al.* (1972) indicated that the improvement in the identification of dichotically presented stop consonants when they share place (e.g. [ba-pa]) or voicing (e.g. [ba-da]), over presentations in which neither feature is shared (e.g. [ba-ta]), is an effect of phonetic rather than auditory processing. On the other hand, dichotic listening tests by Blumstein (1974) indicated that sharing distinctive features (Jakobson *et al.*, 1963) does not enhance the perception of vowels. Thus "the

perception of vowels seems to be based upon their acoustic rather than their linguistic structure" (Blumstein, *Ibid.*).

Wood (1975), who performed perceptual experiments designed to compare the processing of CV stimuli at phonetic and purely auditory levels, published evidence of left-hemisphere lateralization for phonetic processing. He also believed, on the basis of reaction-time data, that this phonetic processing can occur in parallel with central processing of a purely auditory nature (Wood, 1974).

Cutting (1974), as a result of his perceptual experiments with speech and non-speech sounds, believed that there were both auditory and phonetic left-hemisphere mechanisms in the processing of acoustic stimuli. For instance, he found that three-formant synthetic sounds with non-phonetic transitions (initial falling F1 and F3) are perceived with a right-ear advantage, possibly through an auditory mechanism sensitive to stimuli of changing frequency, while "stimuli that are classifiable as speech yield larger right-ear advantages [through phonetic mechanisms] than sine-wave stimuli which resemble speech tokens but which are not identifiable as speech" (Cutting, *ibid.*).

Indications of a mechanism in the left (language-dominant) hemisphere sensitive to frequency transitions *per se* (see Darwin, 1971, for further evidence) are of significance in the consideration of data presented throughout this thesis. Chapter 5, for example, is particularly concerned with the way in which a budgerigar uses - with slight modifications - a species-specific call of sharply rising frequency to represent the steeply rising F2 transition of medial or final [i] or [ɪ]. Formant transitions are particularly well reproduced throughout the avian "speech" repertoire. (See below, this chapter, for a general discussion, and Chapter 6 for specific instances.)

The presence of mechanisms in both man and bird which respond, at the auditory level, to an acoustic feature - the ("F2") frequency transition - common to both speech and bird call, could help to explain the readiness with which birds imitate this feature. It could also indicate that this feature functions as an acoustic cue in avian vocal communication in much the same way as it does in human speech. This feature could then be said to have a common function in man and bird at a higher level - that of vocal communication - as well as at the simply auditory level. (Note, too, the similarity to the communicatively significant cue of frequency peak position in some monkey vocalizations mentioned above.)

Single-formant avian "speech"

Sufficiency of F2. Mol (1971, 1972) noted that, in a grey parrot's imitation of the vowel [ɑ·], there was apparently only one resonance - at 1,500 Hz, an appropriate frequency for F2 of the vowel - excited by the low-frequency fundamental. He interpreted this as an acoustic resonance - the first formant of the parrot's tracheal tube - responding to a phonatory pulse-train excitation. (It has been argued above, however - in Chapters 2 and 3 - that F2 in psittacine speech imitations typically corresponds to a mechanical resonance of the syringeal membranes.) He observed that the parrot made no attempt to imitate F1 of the vowel.

Avian perception and reproduction of F2. Mol went on to suggest that the bird, because of its acuity of auditory temporal resolution, may not perceive this lower-frequency component as such (implying, presumably, that it perceives the F1/fundamental as an amplitude modulation of the second-formant carrier). Sedlacek (1971) had said much the same thing when he suggested that birds might perceive vowels as a series of "very

short bursts of the formant [i.e. F2] frequency".

This is an important point which is central to the argument of this thesis: whereas humans perceive vowel sounds primarily as patterns of resonant response (spectral "modulation" patterns - Plomp, 1983) to an harmonic excitation (the fundamental) in the frequency domain, there is considerable evidence for believing that birds perceive them primarily as patterns of amplitude modulation of a carrier (usually the second formant) in the time domain.

Evidence for avian analysis of such signals in the time domain has been provided by the psychophysical data of Dooling and Searcy (1981) and the electrophysiological data of Langner (1981) discussed above. These perceptual data are supported by analyses of avian "speech" (particularly in Chapters 3 and 6) which demonstrate that parrots, at least, reproduce speech sounds according to this amplitude-modulation pattern.

To anticipate, briefly, the argument of Chapter 6, the reader is referred to Plate 6.4, which illustrates the imitation of a speech phrase by a small parrot (*Nymphicus hollandicus*). A comparison of Figure 6.4A (the human model) and Figure 6.4C (the parrot imitation) shows that the bird reproduces only F2 of the model[12], modulating this in both amplitude (Fig. 6.4E) and frequency (from the 0.4 sec mark on Figure 6.4C) to give the impression of a speech-like timbre. This bird sacrifices intonation entirely, however, as the perceived intonation of the imitated phrase follows the frequency contour of the second formant, whereas the intonation pattern of the human model is independent of (and contrary to) the F2 contour. (See Chapter 6 for further discussion.) The budgerigar "speaker" of Figure 5.2B is more fortunate, however, as the instructor's intonation of the phrase "Pretty pretty boy" [pɹɪtɪ pɹɪtɪ bɔɪ] follows the pattern of the F2 contour[13]. Not only does this bird concentrate exclusively on the second formant in his imitation, but

he follows the speech imitation with a whistled sequence which largely reproduces this F2 pattern. Considering this figure (Fig. 5.2B) it is remarkable that the first half represents an unmistakably speech-like utterance, while the second half bears no resemblance, perceptually, to speech. One important difference between the two is the discontinuities in the speech-like sequence representing consonantal stops. It can be seen that, for instance, if the gap in the first [pɹɪtɪ] were filled by a trace of rising frequency, it would be almost identical, in spectrographic pattern, to the first element of the whistled sequence. It has been pointed out above that in one-formant imitations, budgerigars transpose VOT from the missing F1 to the imitated F2, thus telescoping these two important cues - VOT and F2 contour - into the same narrow spectral band.

Single-formant human speech

Whistled languages. There is a form of human speech which consists solely of a sinusoidal wave of changing frequency within the F2 range. This is the phenomenon of whistled speech, as described by Classe (1956), Busnel *et al.* (1970) and Busnel and Classe (1976). In whistled speech, all the essential linguistic information is conveyed by an orally produced aerodynamic whistle - a quasi-sinusoid. Whistled languages have an average frequency range of about 1,000-2,500 Hz (Busnel and Classe, *ibid.*), which is the same as the F2 range of normal (female) speech, although maximum whistle frequencies can be as high as 3,500-4,000 Hz.

Whistled languages are of two distinct kinds: articulated whistled languages, in which the whistle basically follows the F2 contour of the spoken equivalent (as explained below), and the whistled version of tone languages, in which the whistle follows the intonation contour of the spoken equivalent. Unless otherwise

indicated, the discussion below will be concerned only with articulated whistled speech.

Whistled vowels are differentiated by relative pitch, instead of contrasting timbre as in normal speech. As contrasts in vocalic timbre are the perceptual correlate of variations in F2 frequency in relation, principally, to variations in F1 frequency (the F1/F2 frequency ratio discused above), all three articulated whistled languages studied by Busnel and Classe rely on the obvious cue of relative pitch (representing the F2 frequency contrasts of the corresponding spoken sounds) for vowel contrasts.

In the vowel sounds of speech, the vocal-tract resonances are excited by a source vibration (in the larynx) caudal to the resonating cavities, the spectral envelope of the radiated speech signal representing the transfer function of the vocal tract as a whole. Although Fant (1960) demonstrated that a vowel's formant pattern represents mutually dependent peaks in the transfer function of a system of coupled resonators, it can be stated as a guiding principle that F2 frequencies are primarily determined by the length of the cavity between the palato-lingual constriction and the lips (Stevens and House, 1955). Changes in F2 frequency thus reflect, at least partly, changes in the position of this constriction - i.e., articulatory movements.

In whistling, the aerodynamic vibration at the labial orifice (in a bilabial whistle, as used in tone-language whistling) or the more complex anterior constriction of articulated whistling (as described by Busnel and Classe) is tuned by the resonant properties of the cavity immediately posterior to this constriction (Wilson *et al.*, 1971) - the cavity which is particularly associated with F2 in speech.

Articulatory movements during whistle production can thus effect changes in whistle frequency corresponding to changes of F2 frequency in normal speech. This is basically what the whistler does: he reproduces as

closely as possible - under the constraints of the anterior constriction - the articulatory gestures associated with the spoken equivalent.

(The production of a tone-language whistle, on the other hand, is similar to that of a whistled tune: movements of the tongue are unrelated to articulatory gestures, their function being simply to control the volume of the buccal cavity and thus the pitch contour of the whistle. For the whistled version of tone languages such as Mazateco - see Cowan, 1948 - in which the pitch contour of the whistled signal follows the fundamental frequency contour of the spoken equivalent, this pattern has to be transposed into the higher frequency range of whistled speech. In articulated whistled languages, on the other hand, in which the pitch contour of the whistled signal follows the second-formant frequency contour of the spoken equivalent, no such transposition is necessary.)

Not only, then, is there a general correspondence between second-formant and whistle frequency ranges, but the whistle contour, like the F2 contour, can contain encoded linguistic information relating to articulatory gestures. Thus consonants are cued primarily by "second-formant" (whistle) transitions which have been shown, in the experimental work reviewed above, to contain sufficient linguistic information for phonemic contrasts in speech. The fact that articulated whistled speech retains sufficient linguistic information for the transmission of any combination of speech sounds is an impressive demonstration of the importance of these F2-related cues.

Whistles as "bio-communication". Note that in articulated whistling, as opposed to tone-language whistling, prosody has to be sacrificed, as the second-formant contour of speech is independent of the fundamental frequency contour. This is what happens in

those avian speech imitations, such as that of *Nymphicus hollandicus* discussed briefly above and again in Chapter 6 (see Plate 6.4), which concentrate almost exclusively on F2 representation. In this case it is the F2 contour rather than the fundamental frequency contour of the trainer's utterance which the bird follows. As pointed out above, it is possible that birds hear speech as a modulation of a second-formant "carrier".

Whistled speech demonstrates the human ability to transmit and perceive the phonemic contrasts of speech in a physically simplified acoustic signal which can be represented by a single spectral peak of changing frequency. Such encoding of speech contrasts into a quasi-sinusoidal representation of F2 produces a speech signal which roughly corresponds to the proposed avian speech percept. This correspondence partly explains the intelligibility to human listeners of such simplified avian speech imitations as that represented in Figure 5.2B.

Certain segments of the quasi-sinusoidal whistled speech waveform could correspond to the inter-specific "bio-communication" cues hypothesized above.

Baylis and McConnell (1983) speculated on the existence of such inter-specific cues (which would represent "a general property of vertebrate acoustic signals" - Baylis and McConnell, *ibid.*) as a result of their work on the whistled signals by which shepherds communicate with their sheep dogs. These researchers found, throughout their recorded signals of shepherds from Europe and North America, a strong correlation of signal function and acoustic structure in both the "functionally stimulatory" signal for "Fetch" - two short whistles involving a rise in frequency - and the "functionally inhibitory" signal for "Down" - a single whistle of rising and then falling frequency. The common information-bearing significance of these auditory cues in both man and dog, facilitating a form of inter-species

communication, provide good evidence for the "motivation-structural rules" hypothesized by Morton (1977).

Wild, town-dwelling blackbirds, which can have constant daily exposure to human speech sounds, and which are known to mimic other sounds in the environment, have never been reported to copy speech. It is conceivable that if they perceive speech as a carrier at F2 frequencies, and reproduce this as a whistle without the modulation patterns which impart a speech-like quality to avian imitations, any such copies (which, in wild birds, would be the product of "mimicry" rather than "imitation" - see Chapter 1) would go undetected by human listeners.

Busnel and Classe (1976) point out the cryptic nature of whistled language: to the non-whistler it is undetectable as language, and is perceived simply as part of the environmental background noise. They also report that on La Gomera in the Canaries, which is one of the homes of whistled speech, wild blackbirds commonly copy these human whistles. (Classe - see Busnel and Classe, *ibid.* - found that wild British blackbirds readily copied his renditions of whistled speech.) Thus the environment of towns and cities could abound in avian "speech copying" which would be undetectable to human listeners.

SUMMARY

A review of psychophysical and neurophysiological data on avian hearing reveals an auditory system well adapted for processing certain important components of the human speech signal.

Spectral analysis. Birds typically are most sensitive to frequencies within a range of 1-4 kHz, which constitutes, too, their range of optimal frequency discrimination.

The second (and higher) formants of speech fall largely within this range, the second formant (and its transitions) being the single most important spectral feature in the human recognition and discrimination of speech sounds.

Frequency selectivity in budgerigars, as determined by masking experiments, is more finely tuned than in man between 2 kHz and 4 kHz. Comparative neurophysiological data support the picture of generally finer frequency resolution in birds than in mammals.

Thus the region of the second (and higher) speech formants is subject to precise spectral analysis by birds. This is reflected in acoustic analyses of avian "speech", which show detailed spectral correspondences with the human models in this region. (Compare, for example, Figures 6.2G and 6.2H.)

Temporal analysis. All published data indicate that avian auditory analysis in the time domain is excellent. For example, psychophysical tests on budgerigars using amplitude modulated signals have revealed more efficient temporal processing than in man.

Some auditory-nucleus units in the Guinea fowl are selectively tuned to respond to particular amplitude-

modulation frequencies.

These findings raise the possibility that the fundamental frequencies of speech may be perceived (in the time domain) as amplitude modulations of a second-formant "carrier". This possibility is strengthened by acoustic analysis of budgerigar "speech" (discussed in Chapters 3 and 6) which indicates that some of the sounds are reproduced using an amplitude-modulation mechanism.

Pattern processing. Signal redundancy, and comparable pattern processing at central levels, allow mutual recognition of speech/"speech" sounds between man and bird despite audio-vocal differences at more peripheral levels.

Behavioural testing of non-human mammals has shown categorical discrimination of speech (and speech-like) stimuli comparable to that in human listeners.

Acoustic analysis reveals that some of the cues important for human and animal discrimination of speech sounds are particularly well reproduced in avian "speech".

There is a demonstrable correspondence between these cues (e.g. second-formant transitions - see Chapter 5) and basic vocal patterns of the avian imitator. In the light of this observation, and of the argument (in Chapter 1) that avian speech imitation is an essentially communicative behaviour, it is suggested that these cues may be of communicative significance to both birds and man.

Indeed, considering the animal speech-discrimination data, they may have an even more general significance.

It may thus be possible to broaden the concept of a "speech mode" of perception into that of a "bio-communication mode" - i.e., the recognition and interpretation of species-specific communication signals (including speech) by response to acoustic cues which can themselves be inter-specific.

NOTES TO CHAPTER 4

[1]Elsewhere, he (Schwartzkopff, 1973), suggested that this high damping of the basilar membrane might account for the high degree of temporal resolution in avian hearing.

[2]There is an interesting correspondence between these figures and those for inter-click intervals in the echolocating signals produced by some cave-dwelling birds: the South American oil bird, *Steatornis caripensis* (Griffin, 1953), and some of the South-east Asian cave swiftlets of the genus *Aerodramus* (formerly *Collocalia* - Griffin and Suthers, 1970). Medway and Pye (1977), reviewing the published data on echolocation in species of *Aerodramus*, concluded that ranges of recorded inter-click intervals differed somewhat among species, but that they were all within a range of about 50-200 msec. Recorded inter-click intervals in *Steatornis* (reviewed in Pye, 1980) vary between about 80 msec and 250 msec.

[3]Fletcher (1940) introduced the concept of a series of auditory filters ("critical bands") with bandwidths, on a logarithmic scale, representing divisions of equal length (0.5 mm) along the human basilar membrane. He showed that the curve relating minimum perceptible pitch changes in human hearing to a range of audible frequencies (Shower and Biddulph, 1931) is the same shape as the curve relating critical ratios (i.e., critical bands as inferred from broadband masking data) to the same range of audible frequencies. Critical bands have also been used in the interpretation of data on loudness summation, phase sensitivity and musical consonance (Scharf, 1970).

[4]Dooling and Saunders (1975a) determined critical ratios (from broadband masking data) for the budgerigar. Saunders, Denny and Bock (1978), measured critical bands directly using band-narrowing techniques, and found that the function relating these critical bandwidths to centre frequencies was parallel to that for critical ratios (see Footnote 2, above) but that CB values were about 2.5 times higher than CR values. Dooling and Searcy (1979) followed Bilger (1976) in arguing that "the somewhat arbitrary assumption that at masked threshold the signal power and the noise power in the critical bandwidth are equal is faulty". They showed that the discrepancy between CR and CB values in the budgerigar data could be resolved by understanding that in a broadband masking test "the task is one of detecting the difference in intensity between a critical band of noise alone and a critical band of noise plus signal" (Dooling and Searcy, 1979).

[5]"When the first formant of the complete and reduced vowel are at the same frequency, F2′ is the frequency of the second formant in the reduced vowel that gives the best match in quality to the complete . . . vowel" (Rosen and Fourcin, 1986).

[6]In the song of the chiffchaff (Phylloscopus collybita) however, Schubert (1971) found that all the information was contained in each separate unit, and all units had identical information content.

[7]A similar limit seems to apply in discriminating among VOT-like intervals of different duration. Divenyi and Sachs (1978) showed that when two tone bursts are of unequal frequency or intensity the ability to discriminate changes in their temporal separation decreases when that interval is less than about 25 msec.

[8]Sinnott, Beecher, Moody and Stebbins (1976) found that their Old World monkeys could learn to discriminate a /ba-da/ phonemic contrast. They also found, however, that while the latency function for human discrimination among stimuli from a [ba-da] continuum was discontinuous - i.e., all latency times for discriminations across the phoneme boundary were considerably longer than all latency times for discriminations within a phoneme category - latencies for the monkeys increased linearly as stimulus difference decreased. They argued, on the basis of these results, for a unique speech processing mechanism in humans.

[9]For a modern critique of the classical conception of CP, see Harnad (1987).

[10]Left-hemisphere dominance for auditory habituation learning has been demonstrated in chickens by Rogers and Anson (1979).

[11]Shankweiler and Studdert-Kennedy, 1967b, and Springer, 1971, obtained a similar right-ear advantage using spoken CV syllables.

[12]This can be fully appreciated by placing a strip of paper horizontally along the bottom of Figure 6.4A so as to cover the first 800 Hz.

[13]This is a common phrase in the repertoires of several of the budgerigars recorded for the present study. This probably has something to do with the fact that the usual fundamental frequency (intonation) contour does largely coincide with the F2 contour, thus making it possible for the bird to reproduce both these features with a single modulated sinusoid.

CHAPTER 5:
TRANSFORMATION OF BIRD SOUNDS INTO "SPEECH"

INTRODUCTION

Ornithological literature contains a few general observations on acoustic similarities between the species-specific vocalizations of some "talking" birds and human speech. This chapter will present a detailed analysis of the way in which some species-specific sounds of the budgerigar (*Melopsittacus undulatus*), because of such acoustic similarities, can be transformed into speech-imitation sounds.

Bertram (1970) drew attention to certain speech-like features of the most common vocalization produced by wild hill mynahs in Assam. He named this vocalization the "um"-sound, describing its apparent function in terms of "close-range social contact". "Um"-sounds are soft, broad-band sounds. Although most of their energy lies between 2 kHz and 5 kHz they have a low perceived pitch which is reflected in the spacing of vertical striations on wide-band spectrograms. This low pitch, according to Bertram, helps them stand out against forest background noise. Other acoustic qualities, including sudden onset and wide frequency range, make them easy to locate. Bertram commented on the variability of "um"-sounds - they "may vary in loudness, duration, pitch and quality". As well as the vertical striations typical of voiced speech sounds, spectrograms of "um"-sounds show formant-like horizontal concentrations of intensity.

Nottebohm (1976) recorded some "soft, mumbling vocalizations" produced by wild orange-winged Amazon parrots (*Amazona amazonica*), and speculated that the presence of such vocalizations in their repertoires could have predisposed parrots to imitate speech sounds. Spectrograms of these "mumbling vocalizations" show the vertical striations and formant-like concentrations of

energy typical of human vowels (Nottebohm, *ibid.*).

Brockway (1964, 1969) distinguished 10 types of vocal behaviour in the budgerigar, and Wyndham (1980) eight. One of these vocal patterns - the one she called "loud warble" - Brockway (1964) described as "the 'chatter' so commonly associated with this species". The use of the word "chatter" suggests the speech-like quality of budgerigar warble, and Wyndham's (1980) spectrograms confirm this similarity. Warble contains broad-bandwidth sounds with formant-like concentrations of energy - particularly in the region of human second-formant frequencies (about 800-2,500 Hz).

Loud warble, as described by Brockway, is primarily a courtship vocalization, but is not restricted to this context. It "may involve a thwarted sexual drive" (Brockway, 1964). Wyndham's (1980) observations suggest that warbling may also serve for individual identification, and in enticing the female from the nest.

Warbling is most commonly performed by males, but can also be performed by females. It is highly mimetic: one budgerigar is stimulated to warble by hearing another. Brockway reported that "performers . . . show no specific orientation to other birds" (Brockway, 1969). However, "under appropriate environmental conditions, the hearing and/or performance of warbling promotes full gonadal activity in both sexes" (Brockway, *ibid.*). The budgerigar was the only species of bird known to her in which specific male vocal behaviour produced this effect.

"Loud warble . . . contains recognizable components of many of the other vocalizations of this species, combined in a highly varying pattern" (Brockway, 1964).

These characteristics of the budgerigar's warble have their parallels in the "verbal" (i.e., speech-imitation) behaviour of the birds. Male budgerigars are recognized 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 (although not in the sense of "mimicry" as defined in Chapter 1). The budgerigars recorded during this project 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 other talking species, such as the larger parrots and the 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 his 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. This distinction, related to that between song and calls, has been discussed in Chapter 1, under "Calls and song".

ONTOGENY OF AVIAN "SPEECH"

Two of the four *Melopsittacus* 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 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", contained speech phrases and syllables which were not part of the performance repertoire[1]. This conforms with the eclectic nature of warble as described by Brockway and Wyndham. It also suggests a possible connexion with the subsong of songbirds, which includes a wide variety of vocalizations and, in some cases, imitations of other species (Marler, 1956; Thorpe, 1955; see "'Imitation' and 'mimicry'" in Chapter 1). Physically, subsong has a broader, more speech-like spectral envelope than full song. (See spectrograms in, for example, Thorpe and Pilcher, 1958.) Like warbling and speech imitation in budgerigars, it is mainly performed by males, but can be performed by females.

This pre-performance warble can be appropriately termed "babble", as it has affinities with the babbling behaviour of infants. In budgerigars, as in human infants, babbling can be seen as initially an exploration of the possibilities of the sound-producing mechanism, and then increasingly a narrowing down of these possibilities to those associated with the communication system prevailing in the environment.

"In babbling the child explores the possibilities of his vocal apparatus and so produces an enormous diversity of sounds, out of which a comparative few will ultimately form the speech-sounds of his mother tongue" (Lewis,

1963).

Thorpe (1958) made a similar comparison between the babbling of infants and the subsong of the chaffinch. Thorpe and Pilcher (1958) follow Darwin in *The Descent of Man* (1871) in quoting Daines Barrington's observation that subsong "may be compared to the imperfect endeavour of a child to babble". See "Human/avian similarities in vocal ontogeny and control" in Chapter 1 for further references in the literature to this analogy.

Thorpe (1955) speculated on the possibility of an element of "play" in subsong - an element which could also be involved in the babble of human infants and budgerigars. In babbling, the human infant is "playing with his vocal organs in the same way as he plays with movements of his fingers and toes" (Lewis, 1963).

Poulsen (1951) showed that sub-song is induced by "the incipient effect of male sex hormone". This weak and unoriented sexual stimulus could be analogous to Brockway's (1964) "thwarted sexual drive" in the warble of wild budgerigars, and the speech-imitation behaviour of pet budgerigars. Verbal chatter could indicate a redirection of the "sex drive" in the context of the human environment.

The speech-like babble of pet budgerigars has its parallels in the babbling behaviour of other "talking" species. Foss (1964) noted that his captive mynah birds, after exposure to speech, developed vocalizations which "contained the phonemes of human speech but no words. From a distance it was indistinguishable from human conversation."

West *et al.* (1983) described the speech-like sounds produced by their captive Eurasian starlings, after exposure to speech, as being "characterized mostly by mimicry of the cadence of human conversation. . . . To listeners, it sounded like a person talking in a muffled voice." Their spectrograms of these starling sounds show formant-like concentrations of energy at appropriate

frequencies, but none of the fine structure of speech.

Considered as a whole, a bout of pre-performance babble by a pet budgerigar, which becomes increasingly more speech-like throughout, can be seen to reproduce, in a condensed form, the "speech"-learning process in the bird. The two carefully-trained birds, once they had begun their repertoires proper, recited them usually in the order established during training, and with no interjection of bird sounds. Once the repertoire had been completed, there was another bout of babble before it was repeated. The two birds which had not been carefully trained to repeat a fixed repertoire interspersed each speech-imitation phrase, throughout the recordings made, with babbling sounds.

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". Some examples of analogous intermediate forms in the vocal output of "talking" budgerigars will be discussed in the next section of this chapter.

TRANSFORMATION OF NATURAL SOUNDS INTO "SPEECH"

"Voicing"

The vocalic sounds of speech consist basically of patterned irregularities (imposed by supralaryngeal acoustic resonances) in the spectral envelope of a train of pulses produced at the glottis. While such a "frequency complex" is typical of some bird sounds (Scheich *et al.*, 1977, 1979), the majority of these sounds lack the low-frequency, highly-damped pulsatile component which forms the fine structure of human vowels. If the spectral envelope of a species-specific avian sound were to have spectral-envelope peaks in the regions, say, of F2 and F3, the addition of such a pulsatile component would be an important step in the transformation of the species-specific sound into a speech-like sound. The ability of the syrinx in at least some psittaciform species (including *Melopsittacus undulatus*) to effect quasi-periodic complete occlusion of the airway in much the same fashion as the human vocal folds during phonation (see Chapters 2 and 3) means that these species are peculiarly adapted to perform such transformations.

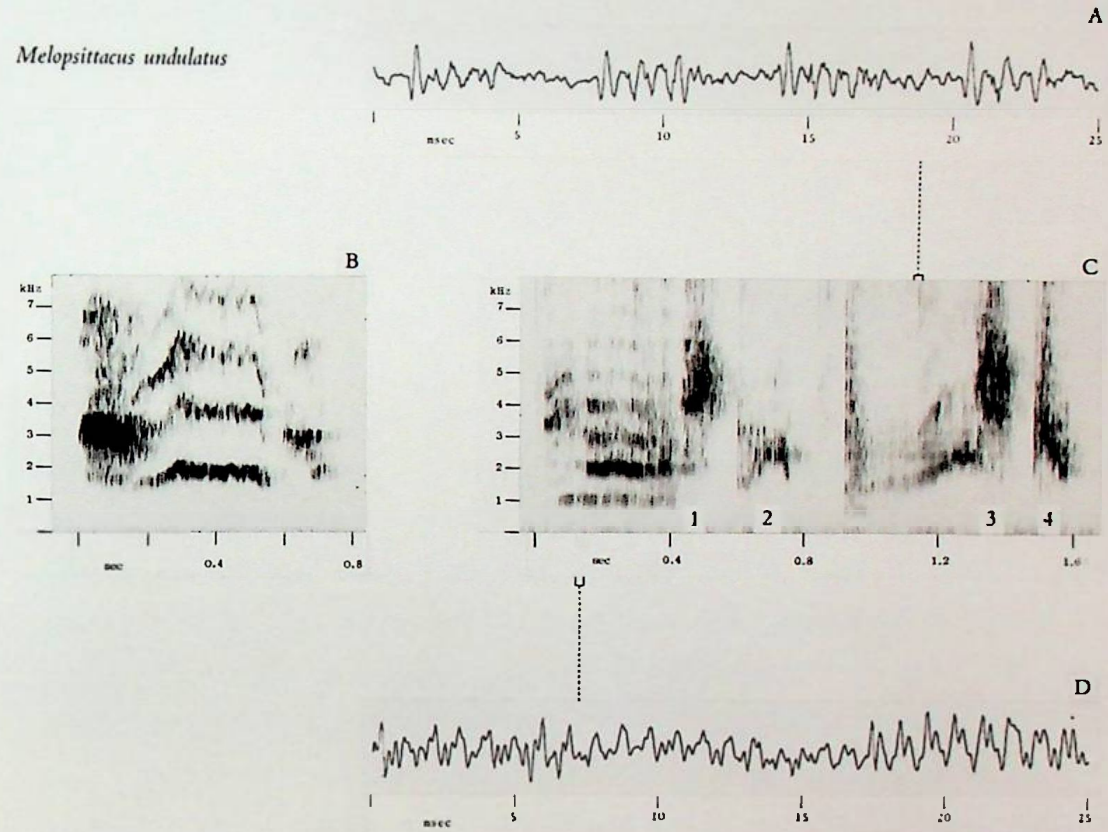
A common sound in the warble of wild budgerigars[2] is a rather strident vocalization consisting of an array of harmonically related components. This may be a version of the "brr" sound which Brockway (1964) observed in the warble of her laboratory-kept birds; it will be referred to hereafter as "harmonic warble". Figure 5.1B is a typical harmonic warble by a wild budgerigar. It has a fundamental frequency - producing an intense spectral peak - at about 1,900 Hz (i.e., within the range of the second formant of speech).

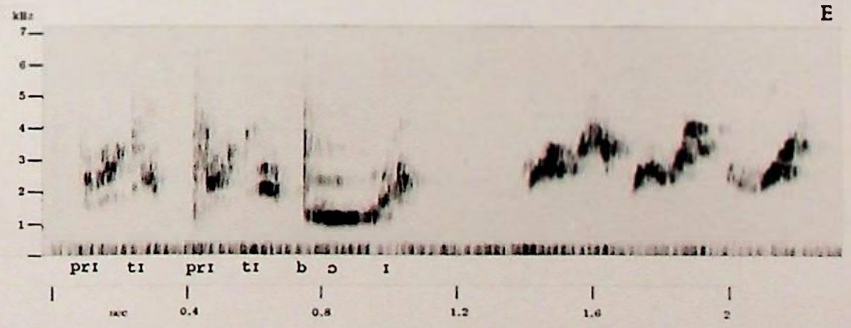
Figure 5.1C shows two sound sequences, each of about

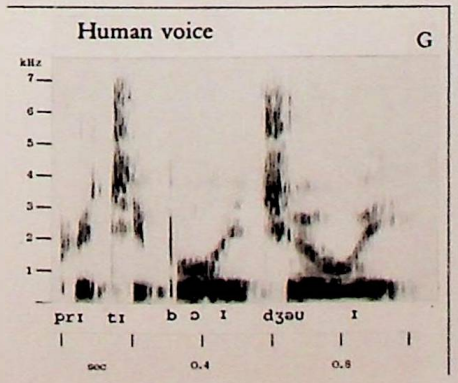
PLATE 5.1

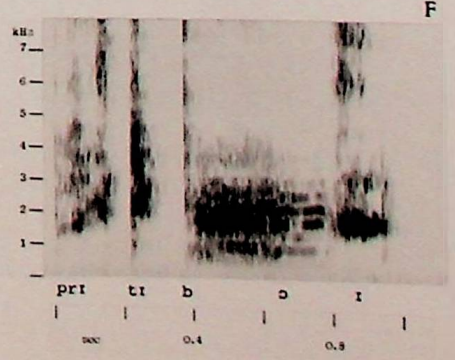
How bird sounds
become
speech sounds

Melopsittacus undulatus









0.75 sec duration, produced consecutively during pre-performance warble by a speech-trained pet budgerigar[3]. The first sequence contains a section resembling harmonic warble with a fundamental component of about 1,000 Hz and an intense second harmonic. The waveform (Fig. 5.1D) of a 25 msec section near the beginning of this sound (just before the onset of the harmonics at about 0.15 sec) shows the 1,000 Hz fundamental, as well as the rather speech-like "formant" rising from about 3,000 Hz to 4,000 Hz (Fig. 5.1C).

Unlike the harmonic warble of wild birds, however, the harmonic spectrum is here associated with a pattern of vertical striations which occur at a frequency of about 125/sec (i.e., within the range of speech fundamental frequencies) from about 0.15 to 0.4 sec on the spectrogram (Fig. 5.1C). Chapter 6 (under "The parrot's vocal mechanism") will include discussion of a possible mechanism for the production of such an acoustic pattern in psittacine "speech". Here it is necessary only to consider the possibility that the imposition of this speech-like fine structure on a species-specific sound having some of the spectral-envelope characteristics of speech (e.g. the second-formant-like peak at 2,000 Hz) is a vocal adaptation in the direction of "speech"[4].

Note that the duration of each of the two sequences in Figure 5.1C is almost identical (about 0.75 sec) with that of Figure 5.1B. Here is a further indication that the two sequences produced by the pet bird may be adaptations of a natural sound sequence such as that represented by Figure 5.1B.

As well as in duration, the second sequence in Figure 5.1C resembles the first in the presence of corresponding pairs of "chrps" (numbered 1 and 2 in the first sequence, 3 and 4 in the second) at the end of each sequence, and in the presence of pulse-like striations. In the second sequence the striations are more regular and speech-like than in the first. They occur at a frequency of about

140/sec, and this low-frequency, speech-like pulse is an obvious feature of the waveform (Fig. 5.1A). (This waveform is similar to that in Figure 6.6J - a "speech" sound produced by a grey parrot. The production mechanism discussed - in Chapter 6 - in relation to the *Psittacus* vocalization probably applies also to the budgerigar.)

On the spectrogram of the second sequence, the harmonic pattern has been replaced by a more formant-like pattern (complete with "formant transition") at frequencies appropriate for F2.

Thus the three sound sequences in Figures 5.1B and 5.1C could represent stages in the transformation of the natural harmonic warble into a speech-like sound.

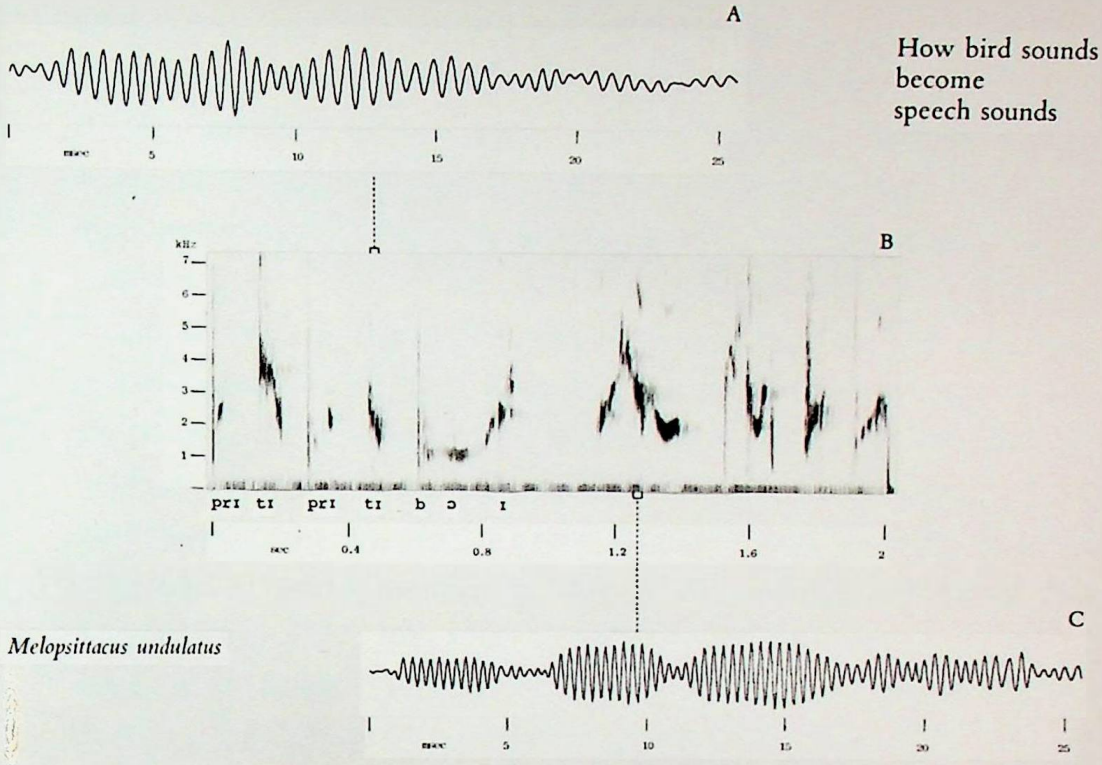
Spectral envelope shape and fundamental frequency contour

Figure 5.2D demonstrates a transition from 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 (1) is followed by an intermediate stage (2) and then the speech-imitation proper (3). It is obvious from the spectrogram that, in terms of acoustic patterning, the bird has modified (1) only minimally to produce the very different perceptual effect of (3).

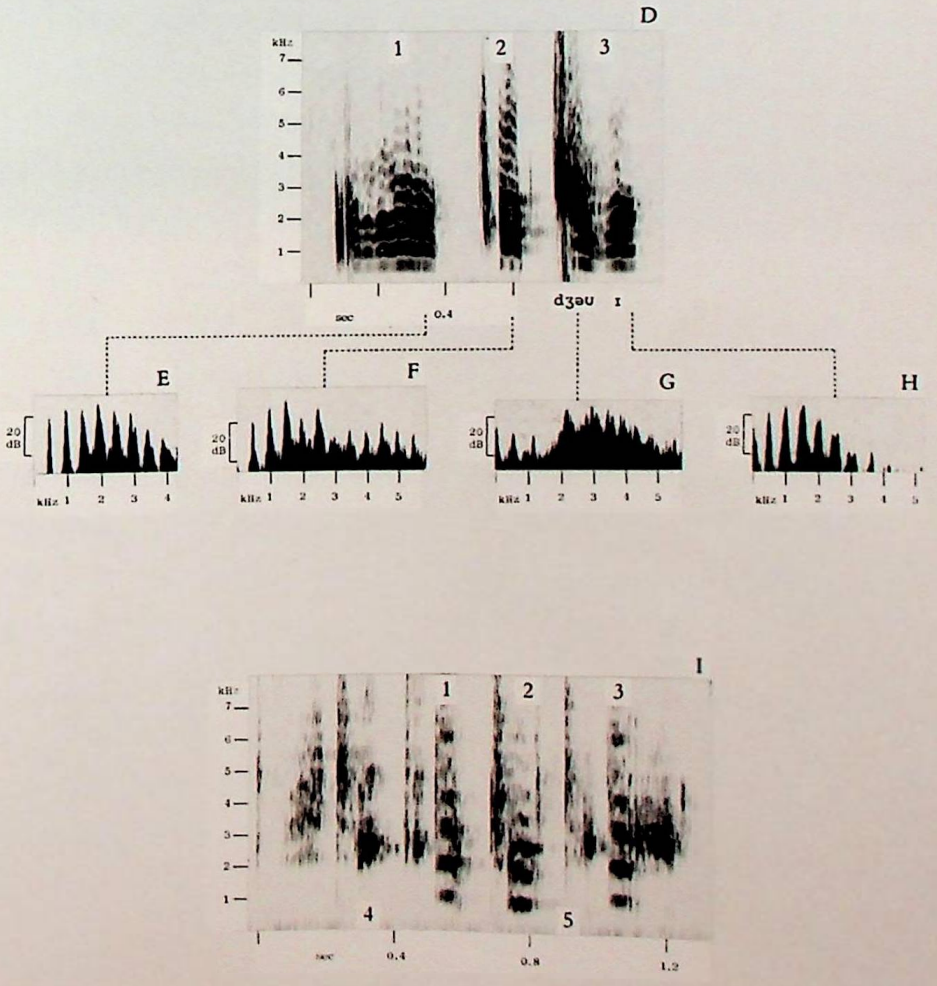
The syllabic structure of (2) - plosive burst + aspiration noise + vocalic segment - is consistent with the speech-like impression it produces. The presence of low-frequency energy in the plosive burst, and the length of the interval between it and the onset of the vocalic segment (54 msec) contribute to the impression of a [p] sound. The noise energy following the burst, although more intense and of higher frequency than aspiration noise in normal speech, must also be a contributing factor.

PLATE 5.2

How bird sounds become speech sounds



Melopsittacus undulatus



Differences between the spectral envelope shape of (1) and (2) account for the more vowel-like quality of (2). The spectral envelope of (1) (Fig. 5.2E) is a broad, gently sloping curve. The spectral envelope of (2) (Fig. 5.2F) on the other hand contains well-defined peaks at around 1,500 Hz and 2,500 Hz - frequencies appropriate to the second and third formants of vowels such as [ə˞], [ɜ] and [ʌ].

A plosive burst at the beginning of (3) is not as apparent as in (2), but there is obviously a sudden release. The fricative noise here is of similar frequency range (mainly between 2,000 Hz and 4,500 Hz), intensity, and duration (about 46 msec) to that in the same word, [dʒəʊɪ], pronounced by the bird's trainer (Fig 5.1G). In the first part of the vocalic segment in (3) the falling second-formant transition of the human counterpart (Fig. 5.1G), falling from about 2,000 Hz to 1,000 Hz, is reflected in the pattern of components at appropriate frequencies. A gap between the two parts of the vocalic segment signals the end of the diphthong, as in speech.

The second part of the vocalic segment in (3) is spectrally similar (Fig. 5.2H) to the original rasping sound, but with a modified fundamental frequency contour. This contour includes an initial rise in fundamental frequency which, in the higher harmonics, resembles the second-formant transition to [ɪ] in Figure 5.1G (the voice of the bird's trainer).

Thus the frequency and duration of initial fricative noise, and the changing frequency of a comparatively weak fundamental with stronger overtones in an ensuing vocalic segment, combine in
(3) to produce the effect of an English word - the bird's name, "Joey".

Formant pattern

Figure 5.2I, taken from a recording of a different budgerigar, shows some speech-like sounds apparently superimposed on three quite audible bursts of high-pitched harmonic warble. (1, 2 and 3)[4]. The perceptual effect, too, is one of superimposition - of the harmonic warble sounds and the speech-like sounds occurring simultaneously. The speech-like sounds include rapid glides like formant transitions (4), clicks like plosive bursts (e.g. 5) followed by high-frequency noise like frication, and concentrations of energy in the F2/F3 region of the spectrum between 2 kHz and 3 kHz.

The spectrogram suggests that the bird is able to use some of the harmonically related spectral components of the rasping sound in reproducing the formant patterns of speech (as in Fig. 5.2D). The component of greatest amplitude in (1) and (3) is about 2,200 Hz - a frequency which would be appropriate for the second formant of some high front vowels - and the first three components of (2) (800 Hz, 1,600 Hz and 2,400 Hz) could be made to represent the formant patterns of a vowel such as [æ], in which the first three formants are separated by roughly these harmonic intervals. (See Plate 6.3, and the related discussion in Chapter 6, illustrating a budgerigar's imitation of the vowel [æ] using just such an array of harmonic components.)

Single formant "talking"

Budgerigar babble also includes sounds which resemble quasi- sinusoidal whistles. Only minimal alterations are required to transform these whistles into a crude representation of speech sounds. (The capacity of such "F2 whistles" for transmitting speech information has been discussed in Chapter 4, under "Single-formant human

speech".)

Figure 5.2B shows a budgerigar's imitation of a phrase - "pretty pretty boy" [pɹɪtɪ pɹɪtɪ bɔɪ] - followed by whistles which look remarkably similar, on the spectrogram, to the phrase. The rapid release and initial upward glide of the first whistle become the plosive burst and rapid upward second-formant transition of [pɹɪ]. The transition - 2,000 Hz to 2,600 Hz - matches exactly the frequency excursion of the whistled counterpart, and is similar to that in the human-voice [pɹɪ] produced by the bird's trainer (Fig 5.1G).

The rapid release and downward glide in the first whistle become the plosive burst, aspiration noise, and falling second-formant transition of [tɪ].Once again, there is a close correspondence of frequencies. (For the remarkable similarity of the human version, see Figure 5.1G.) The second [pɹɪtɪ] is a simplified version of the first, lacking as it does the rapid upward transition typical of an [ɪ] sound, and the aspiration-like section after the release of [t].

In the third word, the rising second-formant transition of [ɔɪ] has its counterpart in a remarkably similar excursion over an identical frequency range (1,200 Hz - 3,300 Hz) in the final whistle. These bear a strong resemblance to the second-formant transition in the human-speech version of Figure 5.1G.

This whistle-like speech imitation is an extreme example of "single formant talking" by a parrot (Mol, 1970, 1971). Here, a sinusoidal wave in the frequency region of human F2 is being modulated to suggest important speech cues such as second-formant transitions and first-formant frequencies. (Compare the discussion of human whistled speech in Chapter 4, and see Chapter 6 for an analysis of speech imitation by a true "single formant talker", *Nymphicus hollandicus*.) Waveforms of downward glides in both the speech-like and whistled sections of Figure 5.2B (Figs 5.2A and 5.2C respectively) show that,

although the carriers are of different frequencies, modulation patterns are similar in both.

("Single formant talking" implies that the listener is able to adjust the acoustic framework within which meaningful contrasts are perceived, so that it encompasses only the second-formant range of the frequency spectrum. See below, this chapter, under "Perception of avian 'speech' by human listeners", for evidence that once such a framework has been established for listening to budgerigar "speech" sounds, first-formant information is not essential.)

Figure 5.1E was made from a recording of a different budgerigar which also produces this phrase and also follows it by three whistles resembling it in both temporal and spectral dimensions. The phrase is rather more speech-like than that in Figure 5.2B, particularly because of more aspiration noise (such as that following the release of the second [p]), a broader spectrum generally, and the more speech-like slope of the transitions (most noticeably in [ɔɪ]). The whistles, too, follow this pattern of a generaly broader spectrum than those in Figure 5.2B, and a more speech-like frequency range and rate of change. The phrase [prɪtɪ bɔɪ] is peculiarly suited to budgerigar imitation because of the predominance of high front vowels (with high-frequency second formants - see next section), and the birds' predilection for the sound [ɔɪ], discussed in Chapter 3.

High front vowels

Perhaps the most convincing evidence for the conversion of a natural sound pattern into an imitation of speech is to be found in those sound sequences in which speech sounds have replaced only part of the total pattern, or, conversely, in which a natural sound has been reintroduced into the imitated word or phrase. In the budgerigar, this is most common in the case of the

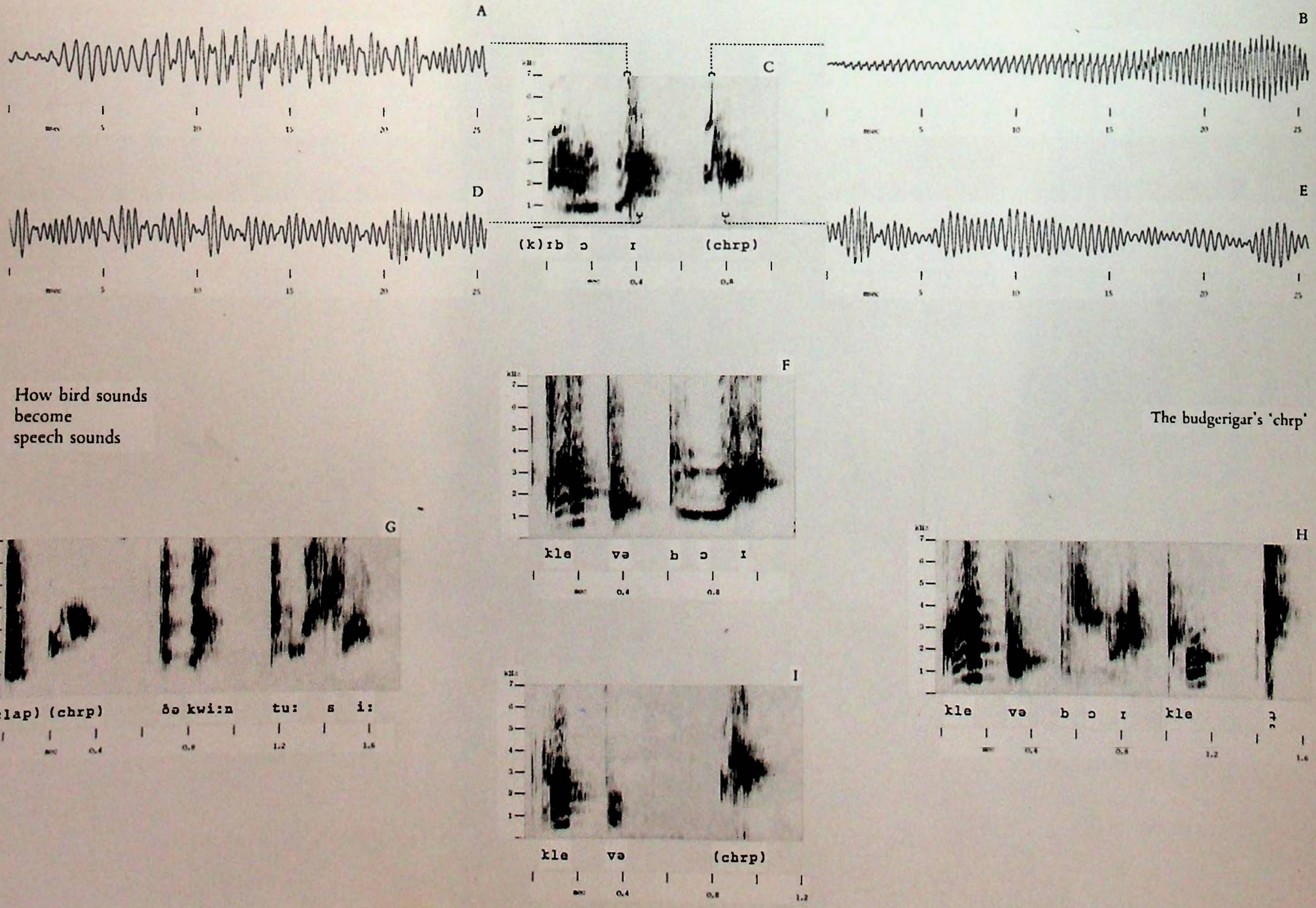
high front vowels [ɪ] and [i], which share some important acoustic characteristics with the most common call of the budgerigar, the "chrp" (Wyndham, 1980). These vowels, like the budgerigar's "chrp", have a strong concentration of energy (high F2 + F3) at about 2,000 Hz - 3,500 Hz. The "chrp" is also characterized 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 are so chirp-like, if isolated, that they have been labelled "chirps" in the literature of phonetics. (See Mattingly *et al.*, 1971. This perceptual phenomenon will be discussed later in the chapter.)

Figure 5.3I shows the insertion of a "chrp" in place of the [I] sound in the phrase "clever boy" [klevə· bɔɪ]. Comparison of this phrase, having a mixture of speech-like and avian elements, with a completely "spoken" version produced by the same budgerigar (5.3F) 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". Compare Figure 5.3F with the human-voice version of [bɔɪ] in Figure 5.1G, and note the similarity of F2 transition pattern.

Further evidence that the "chrp" in Figure 5.3I represents the final section of the word [bɔɪ] is the presence on the original spectrogram (although not clearly visible on the photograph reproduced 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 5.3I occurs 0.83 sec after the initial plosive burst in [klevə·], and exactly the same interval separates the beginning of the [ɪ] sound in Figure 5.3F from the beginning of the phrase. Duration of the [ɪ] and "chrp"

PLATE 5.3

How bird sounds
become
speech sounds



The budgerigar's 'chrp'

sounds themselves is similar.

Figure 5.3H shows the phrase [kleva· bɔɪ] followed immediately by [kle] and then a sound similar to the human [ɻ̊]. This sound resembles the "chrp" in some respects, but it is noisier and has its main energy concentration about 1,000 Hz higher than a "chrp". It is another common [ɪ] surrogate in the budgerigar's repertoire - one which characteristically lacks an initial glide, but which can also be used in the context of [bɔɪ].

In the case of the phrase "to see the Queen" [tu: si: ðə· kwi:n] (Fig. 5.3G), the words [tu: si:] have been replaced, initially, by sounds from the bird's species-specific repertoire, and then reproduced, finally, in speech-imitation form. The first sound in Figure 5.3G, a loud clap, resembles the speech sound [tu] in its sharp onset and ensuing noise energy. The "chrp" is obviously similar in many respects to the [i] in [si:], the important upward transition being present in both. The most important modification in the speech-like version is represented, on the spectrogram, by a pattern of vertical striations corresponding to the perceived speech-like quality of the "chrp".

Figure 5.3C 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".

Waveforms of the speech-like and avian glides (Figs 5.3A and 5.3B respectively) reveal similarities in the transitional section between low and high frequencies (especially noticeable between the 10 msec and 20 msec calibration marks in both figures). The steady-frequency section following the glide has similar modulation

patterns in both speech-like sound (Fig 5.3D) and "chrp" (Fig 5.3E).

Figure 5.1F concludes this analysis of the transformation of natural budgerigar sounds into speech imitation sounds. It shows the phrase "pretty boy" [prɪtɪ bɔɪ] in a transitional version which clearly includes several of the natural sound patterns discussed above. Most obvious are the harmonic warble at the end of [ɔ], and the completely unmodified "chrp" which serves as [ɪ]. Although this sound may be "heard out" as a "chrp", it is interpreted as [ɪ] because of the context in which it occurs. With only slight modification it would be actually perceived as [ɪ] in this context. The following section examines the effect of a speech-like context on the perception of such slightly modified sounds.

Importance of context: an exploratory experiment

Several studies (reviewed in Shankweiler *et al.*, 1977) have shown that labelling isolated vowels is more difficult than labelling the same vowels in a syllabic context. The importance of such a context suggests that vowels which are inadequately specified or have undergone acoustic degradation or distortion, and are therefore not identifiable in isolation, may be identifiable in an appropriate syllabic or verbal context. Conversely, in the case of avian speech imitation, segments which are perceived by the human listener as vowels in the context of a syllable or word, may sound more like bird sounds in isolation. An exploratory experiment was carried out to test whether this is so.

Stimuli. The stimuli for the experiment were five "chrps" and five vowel-imitation sounds, all from a recording of one budgerigar. The choice of stimuli for the experiment was based on the speech-like/non-speech-like contrast exemplified in such doublets as the [ɔɪ]/"chrp" sequence of Figure 5.3C. and the apparently interchangeable [ɔɪ] and "chrp" in Figures 5.3F and 5.3I. Accordingly, these four sounds were used as the nucleus of the test stimuli, and three more vocalic sounds and three more "chrps" were then chosen with reference to the spectral and aural qualities of the original four.

(The choice of "chrps" as non-speech stimuli in the present experiment was influenced by the work of Mattingly *et al.* (1971). These researchers used isolated synthetic second-formant transitions (which they called "chirps") as non-speech-like stimuli in their work on perception differences for speech-like and non-speech-like sounds. They synthesized their speech-like sounds by

adding a first formant with rising transition to the second-formant "chirp", and following the "chirp" with a steady-state second-formant section[5].)

These sounds were sampled on to computer disk from audio tape at 20 kHz after low-pass filtering at 9 kHz. Appropriate sections were then isolated, and stored separately.

All of the five speech-like stimuli were isolated from contexts involving the imitation of high front vowels. All of the "chrp" stimuli, like the two already mentioned, were associated with the imitation of high front vowels in "pre-performance warble". For instance, "chrp" stimulus 3 was the "chrp" in Figure 5.3G, while "chrp" stimulus 4 (not figured here) occurred on the original tape recording only about a second after the sequence in Figure 5.3C. ("Chrp" stimulus 5 - pictured on Plate 5.4 - is discussed in Footnote 4 at the end of this chapter.)

An attempt was made, in selecting and editing the five vowel-imitation stimuli, to create a speech-like continuum - one stimulus being designated "most speech-like" and the other four deviating progressively from that extreme. The order of the stimuli within this continuum was arrived at on the basis of both spectral and aural considerations. Spectrograms of these five sounds, presented in this experimental order, appear on Plate 5.4.

The "most speech-like" stimulus, numbered "1", was derived from the [ɔɪ] sound depicted in Figure 5.3F. Here the spectral components typical of the second formant of this diphthong in human (female) speech are clearly visible. There is a steady-state "formant" at about 1,000 Hz for the [ɔ] segment, followed by a rising transition. A steady-state "formant" at about 2,500 Hz represents the [ɪ] segment. (Compare this second-formant structure with that of the human - female - [ɔɪ] in Figure 5.1G. The patterns are basically the same.)

At first sight, the short, formant-like component at 3,000 Hz near the beginning of the sound, with its falling transition to the 2,500 Hz "formant", looks like an unadulterated avian ingredient. On examination of spectrograms of human [ɔɪ] in which F3 is clearly represented, however (see Potter *et al.*, 1947), it can be seen that this diphthong involves a downward transition of F3 having a similar slope and frequency excursion to that in the figure numbered "1". This emphasises the fact, discussed earlier in this chapter and in Chapter 4, that the concentration of energy around 3,000 Hz in the budgerigar's imitation of high front vowels, in which F2 and F3 frequencies are close together, represents both of these formants of the human speech sound.

This first speech-like stimulus was derived by deleting the beginning of the [ɔɪ] sound in Figure 5.3F to leave a steady-state [ɔ] "formant" of 24 msec duration.

The next speech-like stimulus, numbered "2" on Plate 5.4, was derived from the [ɔɪ] sound depicted in Figure 5.3C. Having a steady-state [ɔ] "formant" of only 16 msec, and a less distinct [ɪ] "formant" at 2,500 Hz, this stimulus was considered, on spectral as well as aural grounds, to be rather less speech-like than the first. (These two stimuli will henceforth be referred to as "speech-like stimulus 1" and "speech-like stimulus 2" respectively.)

Speech-like stimuli 3, 4 and 5 (see Plate 5.4) all lack the [ɔ] "formant", the impression of this part of the diphthong being conveyed entirely by their rising F2 transitions. This deficiency is the main cause of their sounding less speech-like than stimuli 1 and 2. The order of these three stimuli within the speech-like continuum can be correlated with the 2,500 Hz [ɪ] "formant", which is well-defined in bandwidth and relatively long in duration (0.1 sec) in stimulus 3, and shorter (0.04 sec) in stimulus 4. In stimulus 5 this formant-like component

is not distinguishable on the wide-band spectrogram from the wide-bandwidth concentration of energy centred on about 3,200 Hz which represents both F2 and F3. Also, speech-like stimulus 5 shows only a faint trace of the rising F2 transition, its aural impression being more that of [ɪ] than [ɔɪ]. The spectrogram of this sound bears a striking similarity to that of the "chrp" in Figure 5.3I, and the sound is, indeed, very "chrp"-like. In fact, the progressive deviation from the speech-like qualities of stimulus 1, traceable on the spectrograms of the five sounds through the progressive elimination of speech-cue components, means that their "chrp"-like qualities are correspondingly enhanced. That listeners perceived such a progression within these five sounds is confirmed by the negative slope of the categorization curves (Plate 5.4) discussed below.

The "chrp" (in Figure 5.3I) associated on Plate 5.3 with speech-like stimulus 1 (the [ɔɪ] in Figure 5.3F) was designated "'chrp' stimulus 1", and an attempt was made, in choosing the four other "chrp" stimuli, to grade them progressively away from this first stimulus according to several acoustic parameters, thus creating an acoustic continuum within the "chrps". The parameters involved were, firstly, dominant spectral frequency (related to perceived pitch), and, secondly, the relative importance, in perceptual terms, of downward glides within the "chrp". Thus "chrp" stimulus 5 (figured on Plate 5.4) sounds lower pitched, with a predominantly falling intonation pattern, in comparison with "chrp" stimulus 1, which is higher pitched, with rising intonation. This lowering of perceived pitch throughout the "chrp" continuum produced the aural effect of an increasing resemblance to the speech-like stimuli. (See Footnote 8.)

It must be emphasised, however, that the ten stimuli (five speech-like stimuli and five "chrp" stimuli) constituted in fact not one acoustic continuum, but two (a speech-like continuum and a "chrp" continuum), with a

distinct break, in terms of all acoustic parameters, between speech-like stimulus 5 (the "least speech-like" of the speech-like stimuli) and "chrp" stimulus 5 (the "least 'chrp'-like" of the "chrp" stimuli). Thus, as it is the perception of the speech-like stimuli which is of interest in this study, categorization functions for only the speech-like stimuli are presented on Plate 5.4.

Perception test. The previously isolated sounds were played out from computer disk (at 20 kHz, with 9 kHz low-pass filtering) in the order described below, and recorded on to audio tape for presentation to listeners.

The test was in the form of a forced choice, subjects being presented with 100 groups of three sounds. The first two sounds in each group - the exemplars - were "chrp" stimulus 1 and speech-like stimulus 1 in that order, and the third sound was any one of the 10 stimuli, 10 of each of which were randomly distributed in this third place. The seven subjects were asked to decide whether the third stimulus in each group resembled the first more than the second, or vice versa. They were not told anything about the sounds they were listening to.

Results. The categorization function for the five speech-like stimuli is labelled "(a)" on Plate 5.4.

Excepting the exemplar itself (speech-like stimulus 1 which, of course, had 100 per cent correct categorization), and the stimulus acoustically and perceptually closest to it (speech-like stimulus 2, which had 94 per cent correct categorization), 80 per cent of instances of the three remaining speech-like stimuli were grouped with the "chrp" exemplar, showing that in isolation they resembled "chrps" more than speech-like sounds. On the other hand, with the exception of the "chrp" exemplar and the stimulus closest to it ("chrp" stimuli 1 and 2), only 35 per cent of "chrps" were grouped with the speech-like exemplar. Subjects thus

showed much less confusion in the categorization of "chrps".

Thus the demonstrated perceptual similarity of isolated high front vowel imitations and "chrps" gives an indication of the importance of frame-of-reference in the perception of such sounds.

The fact that the budgerigar has to alter the species-specific sound so little in order to create the impression (in the appropriate context) of a speech-like sound suggests that the dominant acoustic feature - a rising frequency glide - may have interspecific significance in the communication sounds of both man and bird.

Listeners' expectations: further exploratory experiments.

The following question concerning the speech-like stimuli then arose: Are these stimuli sufficiently like speech for the provision of information about their linguistic content to produce a significant improvement in their correct categorization? (This involves a perceptual question: Is the listener's conceptual expectation about what he is to hear, as well as the acoustic context in which he hears it, a significant factor in his perception of speech-like sounds? Fourcin, 1968, showed that the categorization curve for single-formant (F2) synthetic speech syllables with variable transition slope could be significantly modified according to the subjects' inferences about the "speech" source - in this case, child versus adult male speaker - based on a single, naturally-spoken precursor word.)

Perception tests. In an attempt to answer these questions, it was decided to run the above experiment three more times with three different groups of subjects, giving each group of subjects a different amount of information (included in the written instructions) about the sounds they were listening to. The group already tested, which had been given no information about the sounds, was called Group (a). Group (b), which functioned as a control group, was told the sounds were made by a bird; Group (c) was told the first exemplar was a bird chirping, and the second exemplar was the same bird imitating a speech sound. In addition to the information given to Group (c), Group (d) was told that the speech sound in question was [ɔɪ]. Thus Group (c) was given only the most general information about the linguistic content of the speech-like exemplar, in contrast to the phonemic information provided to Group (d). There were seven subjects in each group - 28 in all.

Results. Categorization functions of responses by the four groups to the five speech-like stimuli are shown on Plate 5.4.

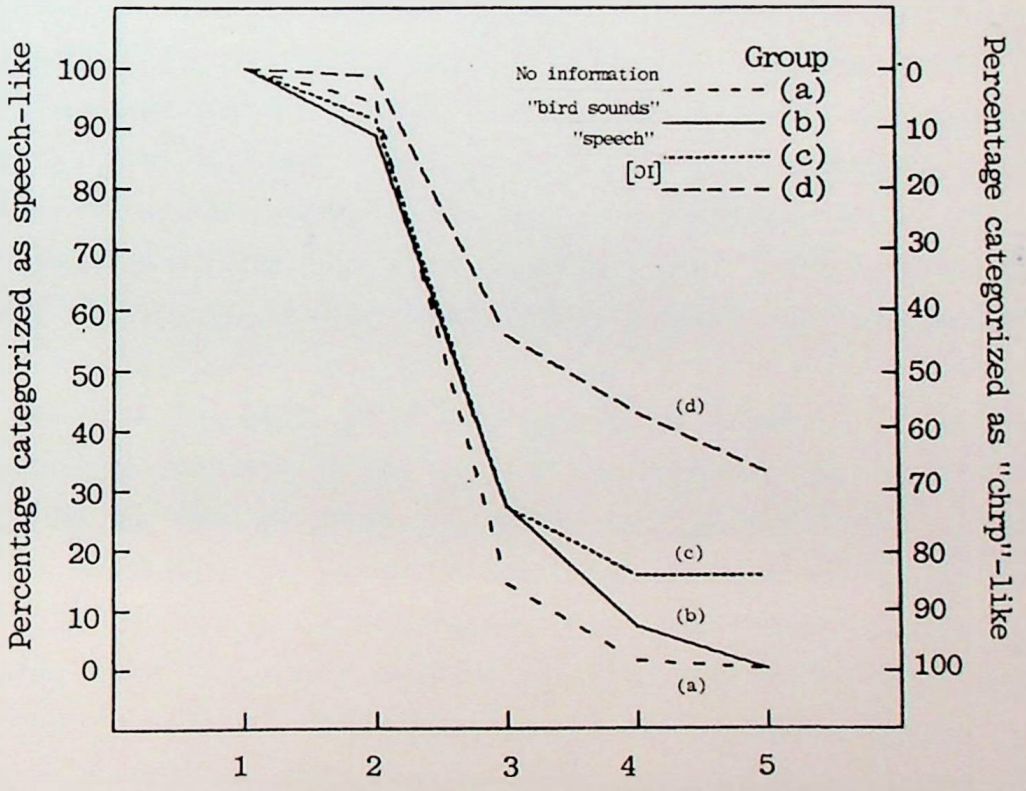
Categorization of the exemplar stimulus (numbered "1" on Plate 5.4) was, of course, 100 per cent correct for all four groups, and that of speech-like stimulus 2 - the stimulus most like the exemplar - was above 88 per cent correct for all groups. Categorization of speech-like stimuli 4 and 5 at the other end of the speech-like continuum was below 50 per cent for all four groups, showing that in the majority of instances listeners perceived these sounds as being more like the "chrp" exemplar than the [ɔɪ].

(The maintenance of the categorization curve's negative slope after the introduction of linguistic information indicates that the ordering of the five stimuli does indeed reflect a movement from a most-

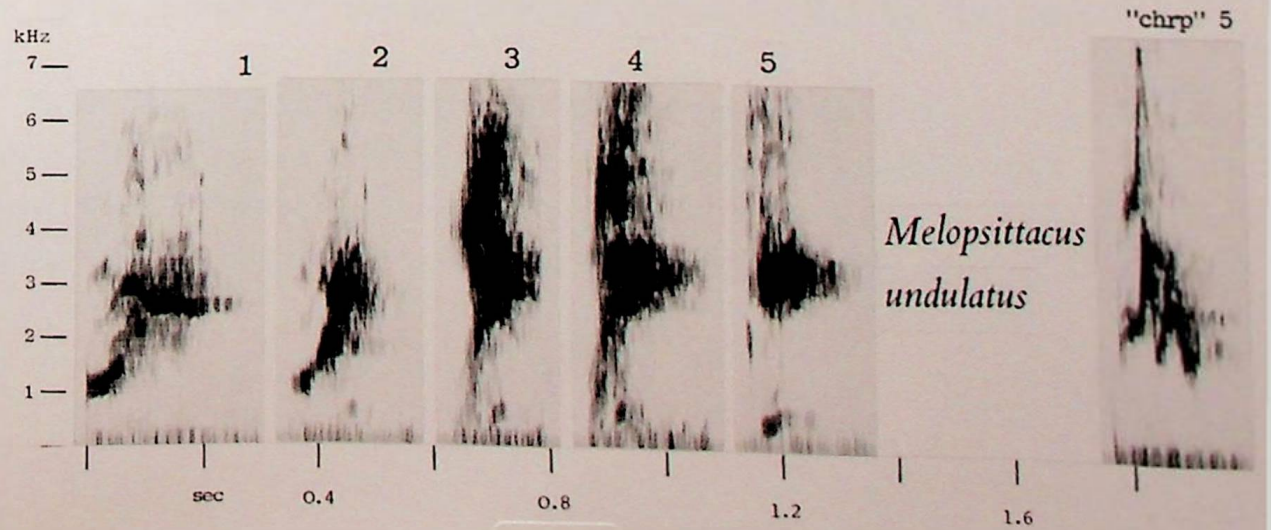
PLATE 5.4

Perception Test

Functions for the categorization of avian "speech" sounds by four differently instructed groups of subjects.



Speech-like stimuli - graded from "1" (most speech-like) to "5" (most "chrp"-like).



speech-like sound to a least-speech-like sound, and not just a progressive deviation from the particular acoustic qualities of speech-like stimulus 1.)

Statistical analysis of variance revealed a significant improvement in the number of correct responses to speech-like stimuli by Group (d) over the responses of Group (a) to the same stimuli ($F\ 1, 12 = 6.14$, $p<0.05$). On the other hand, there was no significant improvement in the categorization of speech-like stimuli by Group (c) over that by Group (a).

The flattened curve of the categorization function for Group (d) indicates a tendency in these subjects towards categorizing speech-like stimuli 4 and 5, as well as speech-like stimulus 3, together with the first two speech-like stimuli. A statistically significant difference between the interaction factors for Groups (a) and (d) ($F\ 9, 108 = 2.43$, $p<0.05$) confirms this picture of a changed response pattern - a change, that is, towards putting all five speech-like stimuli into the same category on the basis of a suggested "phonemic" similarity.

As can be seen from the graph on Plate 5.4, the most striking feature of Group (d)'s improvement over Group (a) is in the correct response to instances of speech-like stimulus 3, which rises from 14 per cent for Group (a) to 56 per cent for Group (d)[6]. Analysis of variance shows that this difference is statistically significant ($F\ 1, 12 = 6.71$, $p<0.05$). There is no significant improvement in response to this stimulus, however, by Group (c). In fact, the percentage of correct responses to this stimulus for Group (c) is exactly the same - 27 per cent - as the result for the control group, Group (b).

This re-categorization of speech-like stimulus 3 by Group (d) after the provision of phonemic information about the speech-like exemplar shows that it must have at least one acoustic feature in common with the exemplar

(speech-like stimulus 1) which could be described in terms of speech cues (see Chapter 4) necessary for the perception of [ɔɪ].

On the other hand, the failure of Group (c) subjects to re-categorize speech-like stimulus 3 after the provision of only the most general linguistic information about the exemplar emphasises that the bird is relying, in its imitation, on a particular correspondence between a "chrp" and a particular speech sound.

Thus the experiment has shown that the listener's expectation of what he is to hear can significantly modify the categorization patterns of isolated avian "speech" sounds.

The finding that sounds such as speech-like stimulus 3, which are categorized by uninformed subjects as being more like the budgerigar's "chrp" than like other [ɔɪ] imitations, are re-categorized when the subjects are told to listen for [ɔɪ] has important correlates in the study of human speech perception itself.

SUMMARY

This chapter has presented detailed evidence that a pet budgerigar (*Melopsittacus undulatus*) can mould some of its natural sounds into speech-imitation sounds. This modification process can be observed in the phenomenon of babbling (or "pre-performance warble"), which was discussed in relation to the sub-song of song birds.

In some speech-imitation sounds, such as those associated with the high front vowels, modification is often so slight that, in isolation, they are categorized by human listeners as "chrps" rather than speech-like sounds. This shows the importance of a verbal context in their perception as speech-like.

Experiments also indicate the importance of listeners' expectations in the perception as "speech" of such minimally-modified bird sounds.

The same sound can be categorized, according to the experimental conditions, as either an avian "chrp" or a particular "speech" sound. This draws attention to basic similarities between avian and human communication sounds, and helps to explain both the ability of some birds to copy speech, and the ability of human listeners to perceive those copies as speech-like.

NOTES TO CHAPTER 5

[1]An amusing example of this is an occurrence of the phrase "Here we go again" in a section of babble by one budgerigar between two sequences of his full repertoire. This was probably a phrase used by his trainer during the process of repeating his repertoire to him.

[2]Recordings obtained from the British Library of Wildlife Sounds.

[3]All four Melopsittacus subjects included rasping sounds in their babble.

[4]The low fundamental, in relation to the tokens of harmonic warble of wild birds I have analysed from the BLOWS recordings, is typical of these sounds on my recordings of pet budgerigars, and may in itself indicate an adaptation to the environment of human speech. (The fundamental frequency of analysed harmonic warble sounds produced by wild budgerigars ranged between 800 Hz and 1,800 Hz, compared with a range of 400 Hz to 1,000 Hz for tame birds.)

[5]They found that the speech-like sounds were perceived categorically according to certain critical values for the slope of the F2 transition, whereas the "chirps" were not perceived categorically. Liberman *et al.* (1967) had remarked that the F2 transition isolated from [di] sounded "like a rapidly rising whistle or glissando on high pitches", and that the F2 transition of [du] was "like a rapidly falling whistle on low pitches". However, both [d]s sounded the same.

- Another significant improvement in Group (d)'s performance over that of Group (a) concerns one of the "chrp" stimuli - "chrp" stimulus 5 (figured on Plate 5.4). Correct categorization of this stimulus rose from 36 per cent for Group (a) to 59 per cent for Group (d). This stimulus has the lowest perceived pitch of all the five "chrp" stimuli, and Group (a) subjects obviously associated it with the low-pitched speech-like stimuli. After the provision of specific linguistic information, however, the majority of Group (d) subjects were able to dissociate it from the speech-like stimuli. The introduction of linguistic information, then, led to the adoption of new categorization criteria in addition to that of perceived pitch - criteria based on the complex acoustic patterns associated with speech cues, of which low pitch is only a part.

CHAPTER 6:
"COMPARATIVE PHONETICS"

INTRODUCTION

Bremond, 1963, used the term "comparative phonetics" in presenting spectrograms of human and psittacine renditions of the same speech phrase. Bauk (1971) compared amplitude spectra of human and psittacine vowels.

The following discussion concerns "comparative phonetics" in this human/avian sense, as well as the acoustic comparison of lexically identical "speech" sounds produced by individuals of different avian species.

Figures 6.1A and 6.1B represent examples - relatively rare among my recordings - of avian imitations uttered immediately after their human models. (As explained in Chapter 1, avian "speech" behaviour is based on species-specific interactive vocal patterns. In most "talking" species the basic behaviour seems to be more commonly a duetting pattern than a call-matching pattern.) Figure 6.1B is of particular interest as it is an example of vocal copying in a species of Charadriiformes. (See Chapter 1 under "Vocal plasticity and neural complexity".)

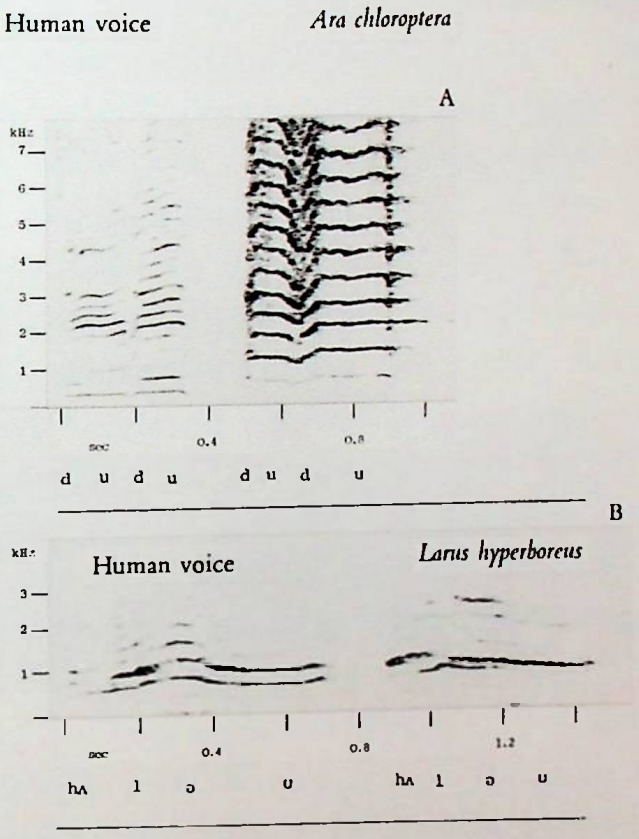
These two figures illustrate several important points which will be expanded on below.

Figure 6.1A illustrates the principle of rhythmic patterning in avian "speech". The basic rhythmic pattern of the model is adhered to, but within this basic pattern (in this case, a disyllabic pattern) there are both similarities and differences in temporal organization. For example, the bird's copy is longer than the human model, but the first syllable of the copy matches exactly, in duration (0.13 sec), the second syllable of the model.

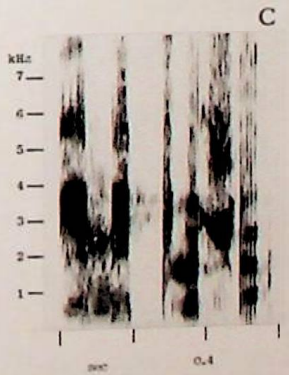
Another important aspect of Figure 6.1A is the

PLATE 6.1

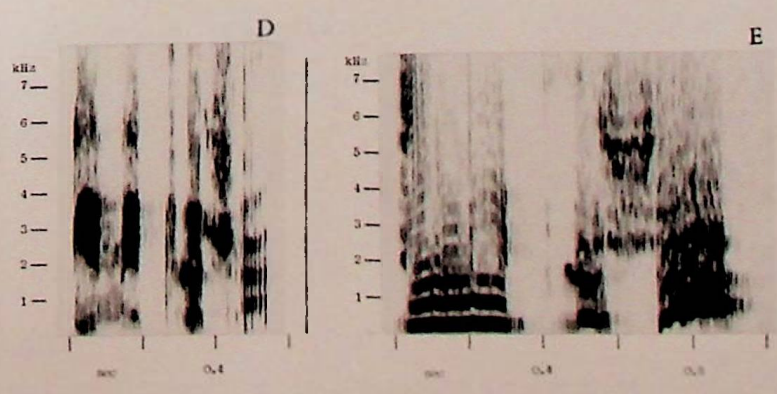
Comparison of bird and trainer



Psittacus erithacus



Human voice



contrast in fundamental frequencies between model and copy. Indeed, the bird's fundamental frequency, at about 700 Hz, matches the first formant frequency of the human model. This could reflect an aspect of avian hearing discussed in Chapter 4 and in this chapter, below, in which both fundamental and first formant may be perceived as modulations of a second-formant "carrier".

The most significant feature of Figure 6.1B is the emphasis on the second formant. It will be shown below that the effectivenes of much avian "speech" depends on the bird's imitative responsiveness to this essential feature of the speech signal.

These two figures (Figs 6.1A and 6.1B) indicate that an avian speech imitation, while incorporating elements of close correspondence with the model along both spectral and temporal parameters, is not necessarily a mere reproduction of speech, but may be more of a reinterpretation. This reinterpretation (or adaptation) is determined by a complex set of factors - anatomical and physiological (Chapters 2 and 3), perceptual (Chapter 4) and behavioural (Chapters 1 and 5).

That the differences between a human model and an avian imitation are the result of such an adaptation, and are not simply the result of random inaccuracies in reproduction, is demonstrated by the consistency of "speech" tokens produced by any individual bird.

Figures 6.1C and 6.1D are two tokens of the sentence [imit ɪə· ʃa] produced by a grey parrot (*Psittacus erithacus*). The remarkable similarity between these different tokens of a highly complex acoustic pattern is immediately apparent, and is typical of many avian speech imitations. (Klatt and Stefanski, 1974, present a similar figure to illustrate the consistency of a mynah's repetitions of a sentence. See below, this chapter, for comparative data on the consistency of acoustic features in repetitions of a sentence by a budgerigar and by his human trainer.)

Despite this remarkable similarity between tokens, the avian imitation differs considerably, in acoustic terms, from its human model (Fig. 6.1E). However, certain elements of the human pattern - most notably relative formant frequencies and formant transitions (for instance, in [lə·]) - are reproduced by the bird. Detailed discussions of such human/avian comparisons are presented in this chapter, below.

Such consistent differences from their human models show that birds are not merely "recording" and "playing back" the sounds of speech, but that they are filtering them through a non-human but intricately responsive pattern-processing nervous system. The consistency among tokens of a bird's speech-imitation sounds testifies to the ability of the avian nervous system to organize and reproduce highly complex acoustic patterns. Even when there are major differences in the reproduction of the same speech sound (Chapter 3), these differences can be shown to occur according to definite contextual rules. (See the Appendix to this chapter, below.)

Vowels and consonants. Plate 6.2, a comparative analysis of the sentence "Long live the Queen" [lɒŋ lɪv ðə· kwi:n], shows a budgerigar's attempt at imitating four different vowel sounds - [ɒ], [ɪ], [ə·] and [i] - as produced by his female trainer in a prescribed phonetic and prosodic context. (The brief comments which follow are intended as introductory to the more detailed discussion of another utterance by the same budgerigar figured on Plate 6.3.)

One of the most immediately striking features of this utterance by the budgerigar, and a feature which is common to all the budgerigar speech imitations discussed in this chapter, is the lack of most low-frequency (i.e., voicing fundamental and F1) energy in comparison with both human speech (here, and on Plate 6.3), and the speech imitation of some other avian species (e.g., Plate

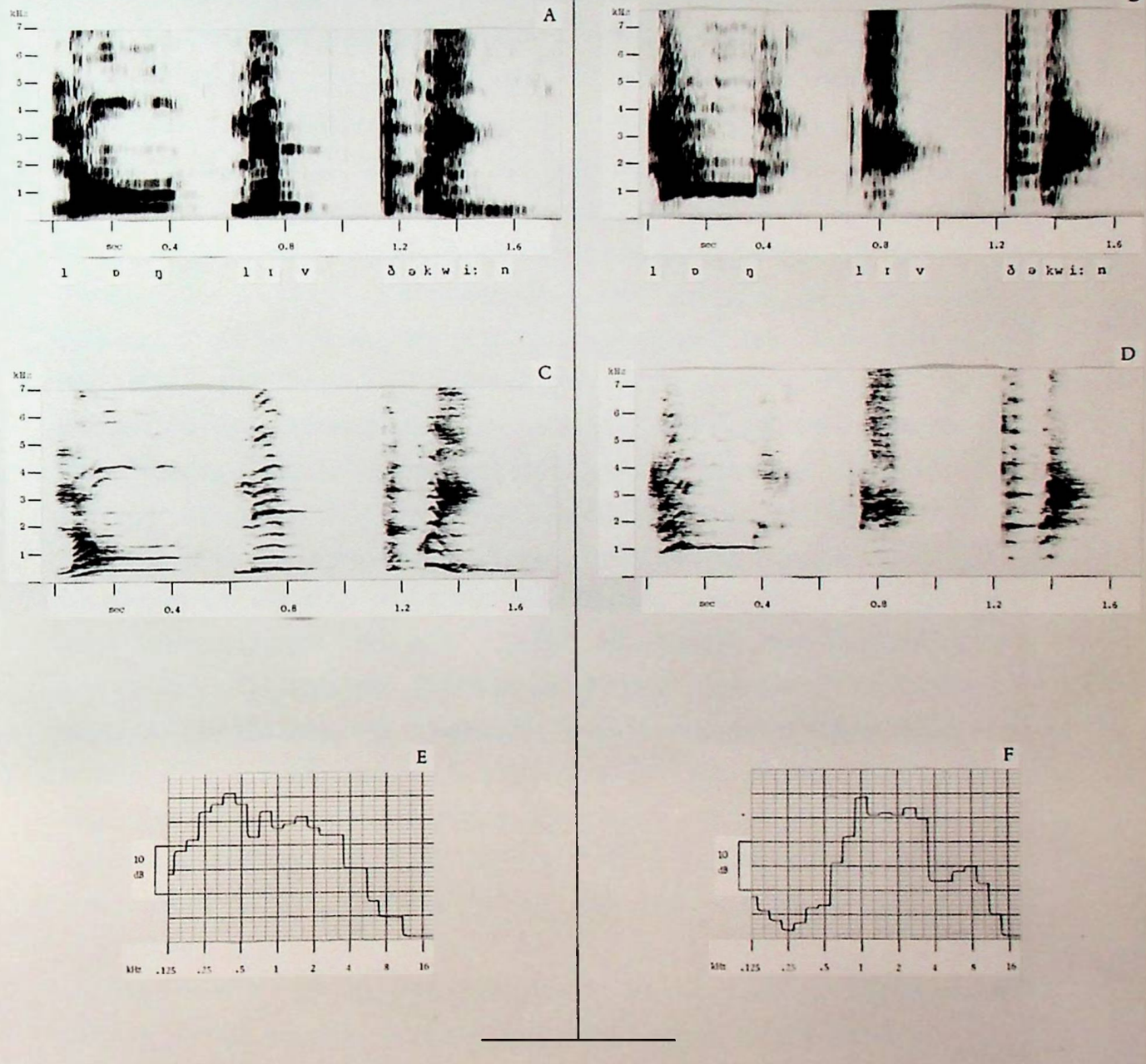
PLATE 6.2

Comparison of bird and trainer

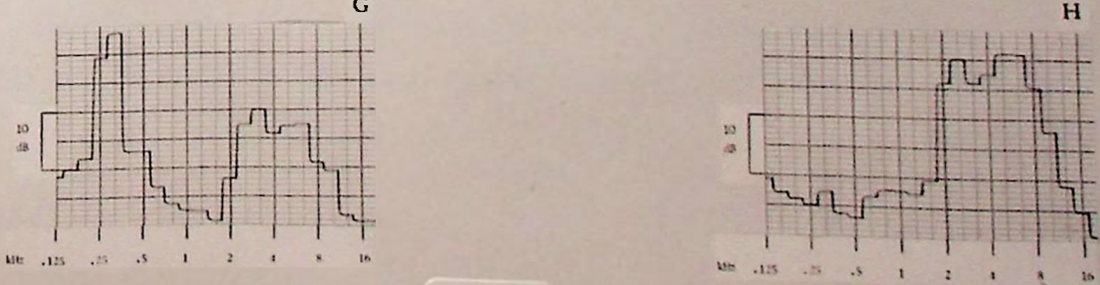
Human voice

THE PHRASE "LONG LIVE THE QUEEN" PRONOUNCED BY A BUDGERIGAR AND BY HIS TRAINER

Melopsittacus undulatus



THE WORD "CHEEKY" PRONOUNCED BY A BUDGERIGAR AND BY HIS TRAINER



6.8).

The budgerigar/human comparison in this respect is made clear in Figures 6.2E and 6.2F - long-term spectral analyses (using a filter bandwidth of one-third of an octave) of the complete human and Melopsittacus utterances. The major peak in the budgerigar's spectrum is at about 1,000 Hz, which is the main spectral peak in many budgerigar vocalizations. (See Chapter 3 for an explanation of this in terms of membrane resonance. Another good example of this is the imitation of [i] on Plate 6.3 - Fig. 6.3N.) The spectral band containing most acoustic energy is between 1,000 Hz and 4,000 Hz, which is also typical of budgerigar vocalizations (Dooling and Saunders, 1975a). In contrast, the region of greatest energy in the (female) trainer's spectrum is below about 500 Hz - the usual spectral pattern of speech, illustrating the dominance of the fundamental component. The budgerigar's spectrum exhibits a marked lack of energy in this region. The spectral band between about 700 Hz and 4,000 Hz contains relatively intense components in the woman's spectrum, however, and it is correspondences in this region on which the budgerigar bases his imitation. The impressive accuracy of the bird's imitation of these higher-frequency components is reflected in the similarity of envelope shape between the two spectra in this band (Figs 6.2E and 6.2F). This figure provides a good example of the combination of "realism" and "impressionism" in avian speech-pattern painting.

An even more impressive example of similarities and differences in Melopsittacus and human spectra is presented in Figures 6.2G and 6.2H. The budgerigar utterance is that featured on Plate 6.8, while Figure 6.2G represents the same word spoken by the bird's female trainer. (Neither bird nor trainer are the same as those featured in the rest of Plate 6.2.) Here the human spectrum consists of two main energy bands - a narrow

band of intense, low-frequency energy associated with the fundamental vibration of the vocal folds and the first vocalic formant, and a less intense band of energy between about 2 and 8 kHz associated with the second and higher formants and aspiration noise. In the budgerigar's imitation, the low-frequency band is entirely absent, while the 2-8 kHz band is quite precisely reproduced. Once again, the envelope shape within this band is remarkably similar in the two spectra, indicating rather precise imitation by the bird of frequency distribution within this band.

One of the most noticeable correspondences between the human and avian spectrograms on Plate 6.2 is the reproduction by the bird of several important formant transitions. The first is that from [l] to [ɒ], in which the human F2 transition from about 2,000 Hz to 1,000 Hz is beautifully reproduced by the budgerigar. In association with this falling F2 transition, and just as precisely imitated, is the rising contour of the lowest spectral components in the narrow-band spectrograms (Figs 6.2C and 6.2D). The bird's speech-like ability to simultaneously effect falling formant frequency and rising fundamental frequency lends great verisimilitude to the imitation at this point.

In contrast to the falling F2 transition of [lɒ], the budgerigar produces an appropriately rising F2 transition in [lɪ] (at about 0.8 sec on Figure 6.2B). He also successfully imitates another F2 transition - that of the glide [w] in [kwi:n] - rising from an initial 1,200 Hz to the high F2 (about 3,000 Hz) of [i]. These frequencies in the model are reproduced in the avian imitation.

The bird's [l] sound, while lacking the first formant of the human model, reproduces F2 and F3 at comparable frequencies (roughly 2,000 Hz and 3,500 Hz). The sharp onset of the succeeding vocalic sound [ɒ] is well imitated, while the second formant, centred on about 1,000 Hz in both the human and avian versions, is the

most impressive feature of the imitation of this vowel.

The two high front vowels in the sentence - [ɪ] and [i] - have the "chrp"-like structure typical of these vowels in budgerigar imitations. (See Chapter 5, and the analyses of similar sounds on Plate 6.3.) In both sounds, the main band of energy in the budgerigar's "chrp"-like imitation is centred on the second-formant frequency of the model - about 2,500 Hz for [ɪ] and 3,000 Hz for [i].

In reproducing the central vowel [ə], which has the first three formants fairly evenly spaced (Fig. 6.2A), the bird is able to employ a simple harmonic structure (Fig 6.2D), as it does for [æ] in Figures 6.3B and 6.3D. The fundamental frequency of about 700 Hz does double duty as F1 (see below, this chapter), while F2 and F3 are represented by the third and fifth harmonics.

The bird precisely reproduces, both in spectrum and relative timing, a noise burst which occurs about 16 msec before the onset of fricative noise in [ðə] (Figs 6.2A and 6.2B).

The overall duration of the two utterances figured here differs by only 24 msec (the budgerigar's being the shorter, as is usually the case in imitations by this species).

Detailed budgerigar-human comparison

Plate 6.3 presents a detailed acoustic analysis and comparison of the phrase "Pussy cat pussy cat" [pʊsi kæʔ pʊsi kæʔ] produced by the same woman trainer and budgerigar "pupil" featured on Plate 6.2.

Energy distribution. Perhaps the most striking difference between these two utterances, which show such a range of remarkable similarities, is in energy distribution. As on Plate 6.2, the budgerigar's output lacks most of the low-frequency energy (100-800 Hz) which comprises the fundamental and first formant of human

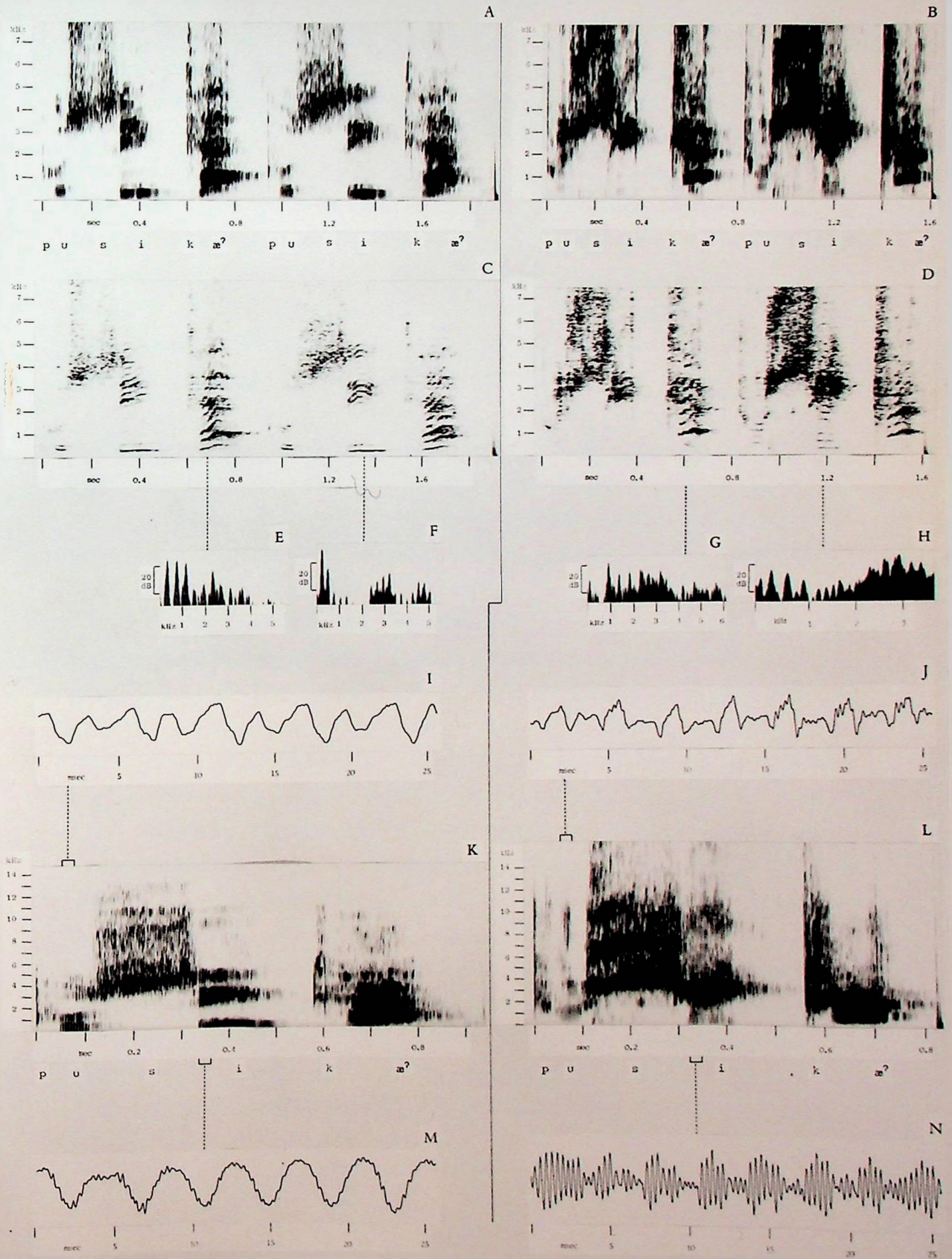
PLATE 6.3

Comparison of bird and trainer

Human voice

THE PHRASE "PUSSY CAT PUSSY CAT" PRONOUNCED BY A BUDGERIGAR AND BY HIS TRAINER

Melopsittacus undulatus



vowels. On the other hand high-frequency noise (4-11 kHz) in the fricative [s] and the aspiration after [k] is of greater relative intensity in the budgerigar sound than in the human (Figs 6.3K and 6.3L). Plosive bursts (e.g. the [p] sounds) are also more intense in the bird's imitation, with energy present at higher frequencies than in the human model.

High-frequency components are unusually pronounced in some of the bird's vowel imitations, too. Excitation between 7 kHz and 11 kHz in [u] and [i] (Fig. 6.3L), which corresponds with excitation at similar frequencies in the intervening fricative [s], is probably the result of acoustic resonance. (See Chapter 2.) The budgerigar's vocalic sounds [i] and [æ] contain more noise energy than their human models, particularly in the region of the second formant (Figs 6.3E, F, G and H).

Vibration patterns. Regularity of vibration during the vocalic sections is imitated by the bird, and is particularly obvious in the relatively low-frequency [u] sound. (Compare the patterns of vertical striations in this sound in Figures 6.3K and 6.3L.)

The waveform of the budgerigar's [u] sound (Fig. 6.3J) bears some remarkable similarities to - and some remarkable differences from - its human counterpart (Fig. 6.3I). Firstly, although the quasi-periodic pattern - with peaks within the fundamental period being of lower amplitude than those representing the fundamental itself - is similar to the typical acoustic resonance pattern exemplified by the corresponding human sound (Fig. 6.3I), there is strong evidence that this pattern is the result of mechanical rather than acoustic resonance.

The fundamental frequency of the bird's [u] sound, as derived from the waveform of Figure 6.3J, is about 270 Hz - which is considerably higher than the corresponding human fundamental of 180 Hz and approximates to the fundamental frequency of 300 Hz which is common in the

vocalizations of several avian species. (See Chapter 3.) The first formant of about 400 Hz in the human [u] is matched by a "formant" of 740 Hz in the budgerigar's [u] - a frequency close to the range of resonant vibrations of the lateral tympaniform membranes as discussed in Chapter 3. As it has been established (in Chapter 2) that acoustic resonances below 2,500 Hz are unlikely in the budgerigar's vocal tract, this first "formant" probably represents such a mechanical resonance of the membranes.

An aspect of the "second-formant" resonance pattern in the avian waveform (Fig. 6.3J) is diametrically opposed to that typical of human vowels. This component, of about 3,000 Hz, is most evident in that phase of the vibration immediately preceding closure of the membranes. (Closure is associated with the major peaks in the waveform.) If this component represented an acoustic resonance, it would be most evident - as often in human speech - just <u>after</u> the closure, as the snapping together of the membranes excited the formant resonance of the vocal tract. Its appearance just <u>before</u> the closure peaks is understandable, however, in terms of mechanical resonance: it would indicate a resonant excitation of the membranes during that phase of their vibration cycle (the closing phase) during which they are under least tension from myoelastic forces. (See Chapters 2 and 3.) The 3,000 Hz component is the 11th harmonic of the fundamental.

The waveform of the [i] sound produced by the bird (Fig. 6.3N) may well result from a somewhat different mechanism of production. Here the weak 3,000 Hz component of the spectrum of [u] has become the dominant component, and the 300 Hz fundamental in [u] has more the appearance, in the waveform of [i], of an amplitude modulation of the 3,000 Hz carrier. The possibility of a beating effect between the components of about 2,700 Hz and 3,000 Hz which appear in Figure 6.3H cannot be ruled out here, as this would produce a waveform similar to that in Figure 6.3N. This possibility appears less

likely, however, on analysis of some other budgerigar sounds (discussed below) which have comparable spectra to that in Figure 6.3N but whose waveforms do not exhibit a beat-like pattern.

The 3,000 Hz spectral peak (Fig. 6.3H) appears too intense, relative to the 300 Hz component, to be simply the result of acoustic-resonance reinforcement of the tenth harmonic of a 300 Hz fundamental. Also, the waveform (Fig. 6.3N) does not show the pattern of excitation and decay typical of an acoustic resonance. An important possibility is that this intense 3,000 Hz component - like the relatively weak 3,000 Hz component in Figure 6.3J - represents a resonant vibration of the membranes at the frequency of the tenth harmonic of a fundamental 300 Hz pulse. In this case, the far greater amplitude of the resonant vibration, and the great difference generally between the waveforms in Figures 6.3J and 6.3N, would suggest the operation of different mechanisms for exciting this resonance.

An answer could lie in the distinction, made in Chapter 3, between myoelastic and myogenic muscle-assisted vibrations. The waveform of Figure 6.3J resembles that produced by myoelastic vibration in the human larynx (Fig. 6.3I). The waveform in Figure 6.3N, on the other hand, may be consistent with a mechanism of myogenic vibration, in which periodic neuro-muscular activity causes periodic adduction of the membranes. The fundamental pulse (300 Hz) could excite the membrane resonance (3,000 Hz), while damping of the resonant vibration during the closure phase could result from complete membrane adduction in each "phonatory" cycle.

According to this model, the action of the vibrating membranes could be seen as analogous, in some respects, to that hypothesized in Chapter 3, in which the membranes periodically approach each other - without complete closure - and thus amplitude modulate the carrier by means of a "muffling" effect. Here, however, complete

adduction does occur, and it is this periodic "phonatory" pulse which both excites the membrane resonance and modulates it at the pulse frequency.

The relative weakness of the fundamental in relation to the 3,000 Hz component can be clearly seen in the spectrum (Fig. 6.3H), while in the human counterpart (Fig 6.3F) the typical speech pattern of strong fundamental and lower-amplitude formant resonances can be seen. (Note that Figures 6.3J, 6.3L and 6.3N are analyses of the second [pʊsi kæʔ] in Figure 6.3B. It can be seen that the [i] in the first [pʊsi kæʔ] has an even weaker fundamental vibration.)

This strong membrane vibration at 3,000 Hz in the imitation of high front vowels such as [i] can be related to some of the budgerigar's natural sounds and is discussed in detail in Chapter 5.

In human speech the vowel [æ] has a much higher-frequency first formant than the vowels [ʊ] or [i] (see below for an analysis of formant frequencies in this utterance), and it can be seen from the spectrograms that the budgerigar is able to produce an intense first formant for [æ] near one of its favoured frequencies of 1,000 Hz, whereas the first formants of the other two vowels are either of much lower intensity or nonexistent. The vowel [æ] has a formant pattern such that, if F1 is 1,000 Hz, adding a second formant of 2,000 Hz and a third formant of 3,000 Hz would produce a reasonable approximation to the sound. This means that a simple harmonic structure could be used in imitating [æ], and acoustic analysis reveals that this is what the bird does. The spectrum (Fig. 6.3G) shows such an harmonic pattern, with a very weak fundamental, and an intense second harmonic which serves as the first formant of the vowel. Higher harmonics represent the higher formants, the second-formant frequency corresponding with that of the fourth harmonic, and the sixth harmonic serving as the third formant. (For an example of the way in which a

budgerigar develops speech-imitation skills by adapting species-specific harmonic patterns, see Chapter 5, Plate 5.2.)

Timing. The budgerigar's utterance (1.6 sec) is shorter than its human model (1.9 sec). This reduced duration is typical of budgerigar imitations generally. All five tokens of [pʊsi kæʔ pʊsi kæʔ] produced by the budgerigar were shorter than any tokens of the same sequence produced by his trainer, the bird's average being 1.65 sec and the trainer's average 2.05 sec. The average figures for the duration of each single [pʊsi kæ'] were 0.99 sec for the trainer and 0.81 sec for the bird.

A comparison of imitations of voiced and unvoiced segments in the 11 tokens of the budgerigar's [pʊsi kæʔ] on tape revealed that voiced segments are much closer in duration to the corresponding segments in the human model than are unvoiced segments, and can even be longer than their human counterparts. For example, the average interval between the onset of voicing in [ʊ] and the beginning of fricative noise in the succeeding [s] sound is 61 msec and 68 msec for trainer and bird respectively, the bird's voiced segment being thus an average of 7 msec longer than the human. In the preceding aspirated segment however (measuring from the plosive burst of [p] to the onset of voicing in [ʊ]), the trainer's tokens are an average of 15 msec longer than the bird's, the average figures being 43.8 msec and 27.9 msec respectively. This greater temporal precision in the reproduction of vocalic sounds than in the reproduction of fricatives supports the theory, discussed in detail in Chapter 4, that the avian auditory system is able to resolve vowel sounds into a series of pulses, thus facilitating the perception and reproduction of temporal features in these sounds. Comparing the [ʊ] segments in Figures 6.3K and 6.3L, it can be seen that there are approximately the same number

of vertical striations - representing glottal or syringeal pulses - in each.

Comparative figures for the duration of fricative noise in [s] for all tokens of [pʊsi kæʔ] corroborate the view that the bird's reproduction of fricatives is less accurate in duration than his reproduction of vowels. The average length of this segment in the human model was 246 msec, while the budgerigar's average was 186 msec - once again considerably shorter than the model.

In considering the relative consistency of duration of all tokens recorded, it was found that the duration of the bird's utterance varied less than half as much as the duration of the trainer's. The same difference held when considering the relative duration of each [pʊsi kæʔ], where the trainer's tokens of this phrase varied within (i.e. between the extremes of) 0.92 sec and 1.1 sec, having a standard deviation of 0.07 sec, and the bird's varied within 0.75 sec and 0.85 sec, with a standard deviation of 0.03 sec.

The duration of vocalic [i] (measured from the onset of voicing to the plosive release of [ ]) was more constant for the bird than for the trainer, varying within 216 msec and 265 msec, with a standard deviation of 13.8 msec. The trainer's [i] sound varied within 244 msec and 330 msec, having a standard deviation of 30.8 msec.

There was an even more remarkable consistency of duration in the [æ] sound of the bird. Taking this sound in the first [pʊsi kæʔ] of each complete sequence, and measuring from the plosive burst of [k] to the plosive burst of the following [p], four of the bird's five segments were of exactly the same duration - 311 msec - the fifth being only 14 msec longer. The corresponding interval in the utterances of the trainer, on the other hand, varied within 345 msec and 471 msec. As this is the only vocalic sound in the phrase which, in the bird's reproduction, is both preceded and followed by a plosive

burst[1], the explanation of this temporal precision is probably contained in the bird's ability to perceive and reproduce precisely the interval between the two pulses, as well as in the capacity for temporal analysis of vocalic sounds discussed above.

Unlike the vocalic sounds, variation within the tokens of the fricative [s] was about the same for bird and trainer. The bird varied within 155 msec and 214 msec, with a standard deviation of 18 msec, while the trainer varied within 216 msec and 263 msec, with a standard deviation of 17 msec. Thus the poorer consistency of fricative sounds compared with vocalic sounds in the bird's imitation is exemplified.

Voice onset time (or VOT, discussed in Chapter 4) - the interval between the initial burst of a plosive consonant and the onset of voicing - is here precisely imitated by the bird. A comparison of Figures 6.3K and 6.3L reveals that the VOT of the trainer's [P] - 45 msec - is exactly reproduced. The following [k] of the trainer's utterance is idiosyncratic in that it has apparently two plosive bursts, the second occuring at an interval of 23 msec after the first. Measured from the second burst, the VOT of bird and trainer agree to within 2 msec, being about 49 msec and 47 msec respectively. (The situation here is complicated, however, by the appearance of a second burst in the low-frequency region of the bird's plosive too, perhaps in imitation of the model.)

Looking at the repertoires of bird and trainer as a whole, the bird is much more consistent in the duration of VOT, four out of nine tokens of avian [Pu] having exactly the same VOT - 41.7 msec. There is no such consistency within the human repertoire.

Formant structure. The three vowels represented in this utterance are of greatly differing formant structure in normal human speech. The vowel [ʊ] has a first-formant frequency near the middle of the range for F1 - about 470 Hz in a woman's pronunciation - and a relatively low-frequency second formant of about 1,100 Hz. The vowel [i] has the lowest first formant (about 300 Hz) and the highest second formant (about 2,400 Hz) of any vowel. The vowel [æ] has a relatively high first formant of about 700 Hz and a mid-range second formant of about 2,000 Hz.

It was shown in Chapter 3, and earlier in this chapter, that a budgerigar employs different physiological mechanisms to produce, in the syrinx, vowel sounds with differently patterned spectral envelopes. By these various means the bird is able to produce "formants" at frequencies remarkably close to those of the human model. For instance, the median value for F2 frequencies in all tokens of [i] pronounced by the trainer was 2,850 Hz, while the corresponding figure for the bird was 2,700 Hz. The bird's formant frequency was rather more variable than the trainer's, however, as it varied within 2,600 Hz and 3,100 Hz while the trainer's varied within 2,700 Hz and 3,000 Hz. Moreover, six out of eight of the trainer's [i] sounds had a second-formant frequency of 2,800 Hz, while such consistency was not shown by the bird. This greater consistency of the human formant frequency is probably because it is a function, largely, of the relatively stable dimensions of resonant cavities in the vocal tract, while the bird's "formant", in this case, is a membrane vibration, and is thus subject to considerable variation.

Greater variability of formant frequencies is usual in budgerigar speech imitations, the only exception in the sequence under discussion here being F1 of the vowel [æ]. As argued above, [æ] appears to consist of a source-generated harmonic spectrum. The fundamental is

relatively weak, while the intense second harmonic is at a frequency - about 1,000 Hz - which has been shown (in Chapter 3) to be a common mechanical resonance in the vibratory pattern of a budgerigar's syringeal membranes. This reinforced second harmonic serves as the first formant of the vowel, while the second and third formants are the results of similar reinforcement of the fourth and sixth harmonics.

In this technique of producing the vowel sound [æ], the fundamental must be around 500 Hz for the resonance-enhanced harmonics to coincide with appropriate formant frequencies. The first "formant" will, then, also be fairly constant in frequency, and the higher "formants" more variable as each variation in the fundamental is multiplied in them by their harmonic number. This is the case in the budgerigar's [æ], the standard deviation of F1 being 71.5 Hz, that for F2 157 Hz, and that for F3 230 Hz. (Note that these figures, as would be expected in the situation hypothesized, are close to the ratio 1:2:3.) The human model, on the other hand, was most constant in F2 (a standard deviation of 103.5 Hz), the most important formant for vowel recognition, and more variable in F1 and F3 (standard deviations of 156.8 Hz and 174.1 Hz respectively).

The ratios among the bird's formant frequencies are roughly those predicted by the hypothesized theory of sound production. The mean values of the formant frequencies are 1,070 Hz, 2,005 Hz and 3,045 Hz - a pattern close to that of 1:2:3 predicted by the theory. Ratios among the formant frequencies of the trainer's utterance, however, differ considerably from this pattern, her mean formant frequencies being 956 Hz, 2,225 Hz and 3,406 Hz.

Formant transitions. Formant transitions are among those aspects of the speech signal which birds reproduce best. This is probably because the formant transitions of speech resemble, in frequency range and rate of change, elements in the vocal repertoires of many avian species. (See Chapter 5 for an experimental investigation of this resemblance and its importance in speech imitation.) A pattern typical of the syllable [kæ] can be seen in the trainer's utterance (Fig. 6.3A). The first formant does not appear on the spectrogram until the onset of voicing; there is then a rising first-formant transition (of about 32 msec duration in the first token of the sound) before the steady state of the formant is reached. (See Liberman *et al.*, 1958, for a discussion of these patterns.)

The second-formant falling transition characteristic of this syllable is discernible in the aspiration noise between the plosive burst of [k] and the onset of voicing. (See particularly the second token of [kæ].) After voicing begins, the falling transition continues briefly until the steady state of the formant is reached.

These features are very precisely imitated by the bird. The rising F1 transition, in both instances of his [kæ], has the same duration (32 msec) as that of the first token of the trainer. (Nabelek and Hirsh, 1969, found that transition durations of around 30 msec produced the best discriminability among vowel sounds. See also Liberman *et al.*, 1956. These figures agree with the results of Pollack, 1968, discussed in Chapter 4.) The falling F2 transition of the budgerigar's first [kæ] is remarkably similar in frequency range and rate of change to that of the trainer. In both cases, the starting point of the F2 transition is about 2,600 Hz, which is within the range typical of a syllable beginning with a velar consonant (Liberman *et al.*, 1967).

The noise energy in the bird's transition is, however, much more intense than that in the trainer's. This

difference is again in evidence in the rising F2 transition at the end of [pu]. For the trainer, although aiming at the band of energy in [s] which, typically in the human voice, is centred on F3 frequencies of the adjacent vowels (Stevens, 1980), it is not apparent on the spectrograms after the voicing of [u] has ceased. In the bird's imitation, however, this transition is not produced by an acoustic resonance dependent on a sound source for its excitation, but is produced directly by membrane vibration. In this case, therefore, the transition can continue to its end point at the frequency of the energy band in [s].

The rising F2 transition at the beginning of [i] is well imitated by the bird. (See Chapter 5 for a discussion of the way in which some [i] imitations are based on the budgerigar "chrp", which often has an initial rising glide.)

Relationship of perceived pitch and F2 frequency

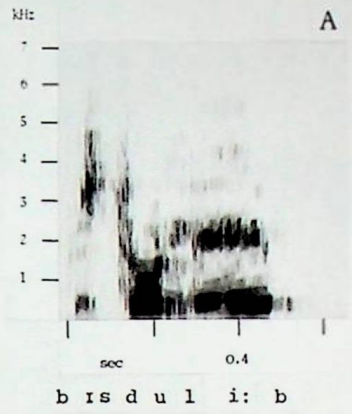
Nymphicus hollandicus. There have been numerous examples (e.g. Fig. 6.3N; see Chapter 3 for discussion) of the budgerigar's ability to produce the impression of a low-frequency fundamental when there is no spectral component at such a frequency. It was shown that he is able to reproduce the intonation pattern of the human model even when, at the physiological level, there is no fundamental vibration analogous to that of the human vocal folds.

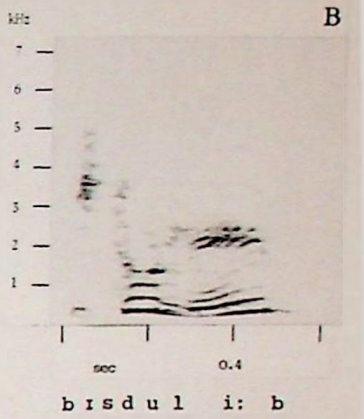
A speech-imitation recording of another small parrot - the cockatiel (*Nymphicus hollandicus*) - shows that in this case there is little or no attempt to reproduce fundamental frequency, and thus the intonation pattern of the model. This bird, no doubt from a combination of perceptual and physiological constraints (see Chapters 3 and 4), is forced to sacrifice prosody in its imitation and to simply rely on the reproduction of higher-frequency components of the model - usually features related to the second formant. This means that the "intonation pattern" of the imitation follows the contour of this high-frequency energy, and thus bears no relation to the intonation pattern of the model. Indeed, the two could not be more different. Figures 6.4A and 6.4B are wide-band and narrow-band spectrographic analyses of the human utterance. Following the curve of the human fundamental (and its harmonics) in Figure 6.4B, it can be seen that the fundamental frequency of the first syllable, [bɪs] (ending at about 0.1 sec), is much lower than that of the two succeeding syllables. (Note, correspondingly, the more widely-spaced vertical striations in this first syllable in the wide-band

PLATE 6.4

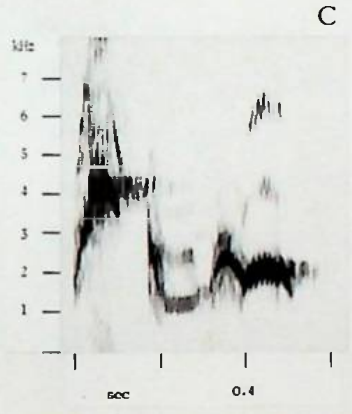
Relationship of pitch and F2 frequency

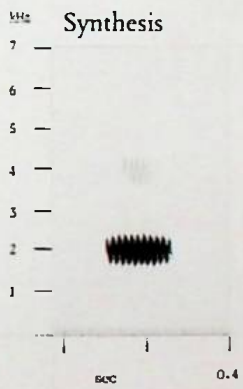
Human voice

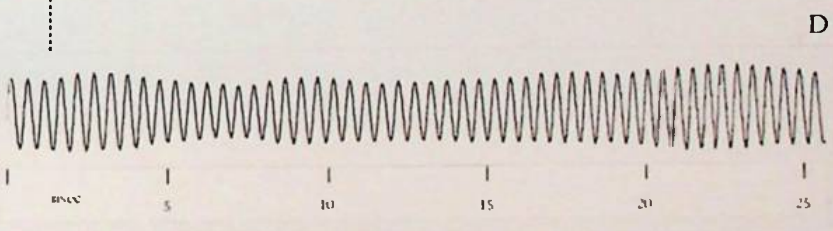


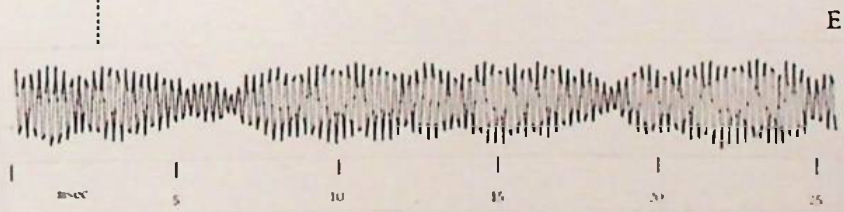


Nymphicus hollandicus

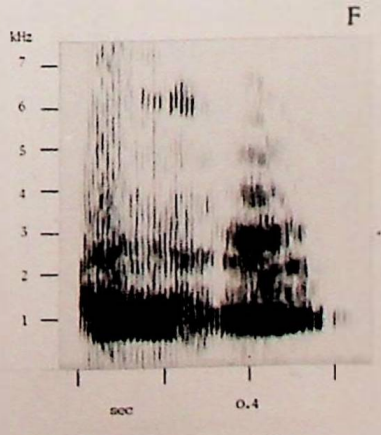


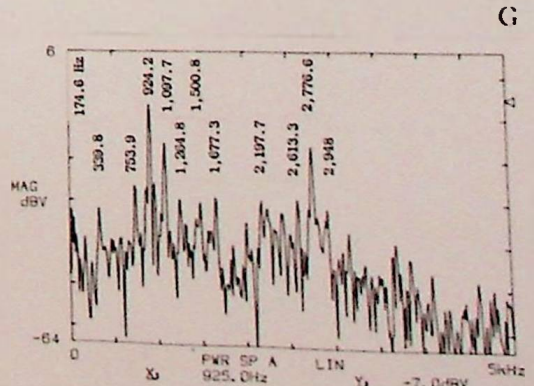






Psittacus erithacus





analysis of Figure 6.4A.) Figure 6.4C is a wide-band analysis of the bird's imitation. In this avian version the perceived pitch of the first syllable is much higher than that of the other two syllables. This is because the bird is simply reproducing the intense concentration of energy which, in the model, rises from about 3,000 Hz to 3,400 Hz and is associated with the succeeding fricative. The bird's steep, "chrp"-like upward glide at the beginning of the syllable is an exaggeration of this rising pattern in the human model, while part of the downward glide of [du] (Fig. 6.4C) closely resembles the model's F2 transition in this syllable (Fig. 6.4A).

In the second and third syllables of the phrase the bird concentrates on reproducing the strong second-formant pattern of the human model while, as in the budgerigar's [ɔ] (Fig. 3.1C), producing no output in the region of F1. The trainer pitches the first of these syllables, [du], higher than the second, [li:b]. (Note, in Figure 6.4B, the falling contour of the harmonics between [du] and [li:b] - followed by a more gradual rise towards the end of the final syllable.) The second formant of [u], however, is of lower frequency than that of [i] (Fig. 6.4A), and it is this relationship which determines the pitch contour of the bird's imitation (Fig 6.4C).

The human F2 of about 1,000 Hz in [u] is faithfully reproduced by the bird, and the considerably higher F2 of [i:] (about 2,000 Hz - see Fig. 6.4A) is reproduced as a couple of "chrp"-like elements centred on the appropriate frequency. This "chrp"-like appearance, together with the 2,000 Hz frequency, which is near the most sensitive audible frequency of many avian species (Dooling et al., 1979) including that of the closely-related budgerigar (Dooling and Saunders, 1975a), support the view that the cockatiel is here using species-specific sounds as part of its speech-imitation strategy. (See Chapter 5.)

The waveform of the first of the two elements of [i:]

(Fig. 6.4D) shows the dominant 2,000 Hz component modified only very slightly by other spectral components and a weak amplitude modulation. Compare this with the waveform of the same vowel sound as produced by a budgerigar (Fig. 6.3N). Low-frequency periodicity in the waveform of the budgerigar is much more marked, and at a frequency corresponding to the fundamental of the human speech signal.

The second "chrp"-like element in [i:] (Fig. 6.4C), however, shows strong FM at a modulation frequency of about 70 Hz. This produces the perceptual effect of a low-frequency periodicity, like a pulse train, which affects the timbre of the sound, while the perceived pitch is still that of the 2,000 Hz carrier. In order to confirm this analysis, a synthetic sound was produced consisting of a 2,000 Hz carrier frequency-modulated at 70 Hz. (A spectrogram of this sound, labelled "Synthesis", is presented on Plate 6.4.) The synthetic sound matches the bird sound well in terms of both the spectrogram and the auditory percept, while the spectrum of the bird sound (Fig. 6.5A) presents a biological version of the electronically-produced sound (Fig. 6.5B).

The waveform of [ɪ] in the first syllable (Fig. 6.4E) shows a more pronounced modulation of the 4,500 Hz carrier. The 85 Hz cycle evident in this section of the waveform is repeated several times in succession, and could impart a certain speech-like quality to the high-pitched whistle of the bird. It is much weaker and less regular, however, than low-frequency envelope periodicity in the waveforms of the best budgerigar speech imitation, and the imitation therefore lacks the budgerigar's illusion of voicing. Rather, this *Nymphicus* imitation resembles certain speech-like whistles in the budgerigar's "pre-performance warble" (see Chapter 5) illustrated in Figure 5.2B.

Although generalizations about *Nymphicus* speech imitations cannot be made on the basis of this one

example, it is significant that the token illustrated here was presented by the trainer as a complete and polished performance by the bird, and not an example of "pre-performance warble".

Psittacus erithacus. Figure 6.4F shows how in a larger parrot - in this case an African grey - second-formant frequency is independent of perceived pitch. The spectrogram represents the word "no" [nəʊ] pronounced by the parrot with rising pitch. The considerable rise in fundamental frequency - from about 80 Hz to 160 Hz - can be seen in the spacing of the vertical striations (i.e., wider spacing in the first syllable). The range of greatest spectral energy, however, remains roughly the same throughout the utterance, and the spectral peak representing the second formant actually falls in frequency, following the pattern typical for the diphthong [əʊ], at the same time as the fundamental frequency is rising (between about 0.2 and 0.3 sec on the spectrogram, Figure 6.4F). The high first formant of [ə], at about 800 Hz, is almost indistinguishable on the spectrogram from the low second formant because of the bandwidth of the analysing filter. The medial transition section, between 0.2 sec. and 0.35 sec. on the spectrogram, clarifies this ambiguity, however, as it shows the second formant of [ə], at about 1,200 Hz, falling to about 800 Hz - an appropriate F2 frequency for [ʊ]. The first formant of [ə] is left without a counterpart in [ʊ], presumably because the bird is not able to produce a resonant vibration - either mechanical or acoustic - at the low frequency necessary (400-500 Hz).

The spectrum (Fig. 6.4G) of the second half of the word shows the major peak, representing a resonant vibration of the lateral tympaniform membranes, at about 925 Hz, with a prominent third harmonic at about 2,775 Hz. This complex vibration is amplitude modulated by

periodic adduction of the membranes at a frequency - about 170 Hz - unrelated to the membrane resonance. The AM pattern is revealed in the appearance of peaks about 170 Hz above and below both the 925 Hz peak and its third harmonic. (The adduction of the membranes may be complete enough to produce a relatively weak quasi-periodic pulse at 170 Hz, represented in the spectrum by a low-amplitude peak at this frequency.) The pattern of vertical striations on the spectrogram is directly related to the periodic peaks in the modulated waveform. (See the discussion of Figure 6.6B below for an explanation of the voice-like quality of these striations.) The complexity of the spectrum is the result of the interaction of two complex vibrations, neither carrier nor modulator being simply sinusoidal.

Relationship of perceived pitch and F1 frequency

Ara chloroptera. Figure 6.5D is a call by a green-winged macaw (*Ara chloroptera*) having a frequency modulation pattern similar in some respects to the cockatiel sound discussed above (Fig. 6.4C), but with a much lower carrier frequency (550 Hz), a lower modulation frequency (50 Hz), and a more complex harmonic structure. The spectrum of this sound (Fig. 6.5C) illustrates the harmonic structure more precisely, showing the dominant second harmonic of the carrier at 1,100 Hz. The spectral pattern created by this second harmonic and its associated sidebands can be compared with the spectrum of the cockatiel's FM sound in [i:] (Fig. 6.5A). The greater compression of the macaw's spectral envelope in the frequency dimension around this peak is due to the lower modulation frequency.

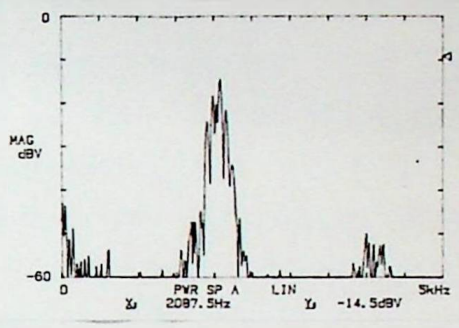
The macaw's vocalization sounds like a species-specific call[2] slightly modified - mainly by resonant reinforcement of the second harmonic - to bear some slight resemblance to the word "hello" [hʌləʊ]. The

PLATE 6.5

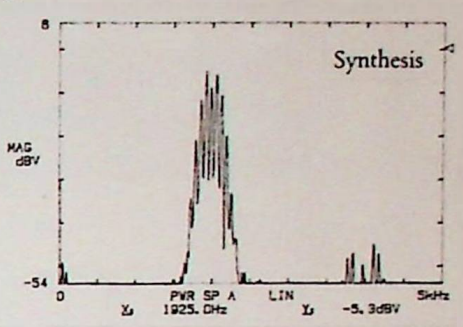
How bird sounds become speech sounds

Nymphicus hollandicus

A

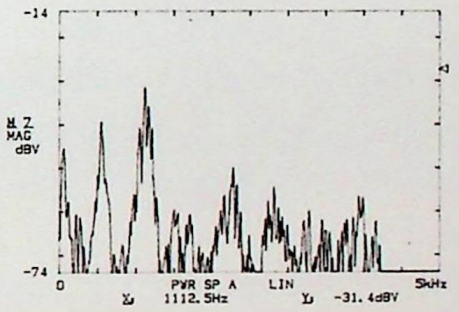


B



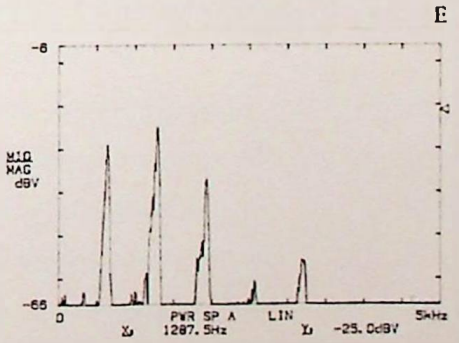
Ara chloroptera

C

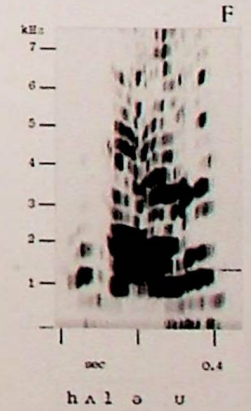


D

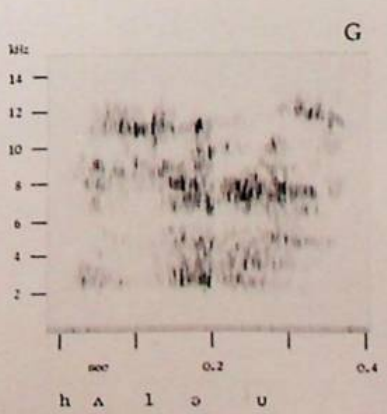
E



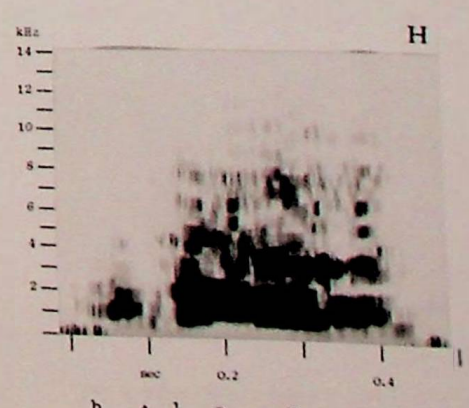
F



G



H



spectral energy peak around 1,100 Hz approximates a frequency on which most of the second-formant energy in a woman's pronunciation of the word [hʌləʊ] could be centred. There is a slight fall in the centre frequency of this "formant" towards the end of the vocalization, typical of F2 in [əʊ]. The fundamental is of an appropriate frequency (550 Hz) to serve as a first formant in all the vocalic sounds included in the word. (See numerous other examples - e.g. the [æ] sound on Plate 6.3 - of the use of a high-frequency fundamental as F1 and its second harmonic as F2.) A concentration of energy near the fourth harmonic (2,200 Hz) of this fundamental (see Fig. 6.5C) could serve as a third formant. The perceived pitch is that of the 50 Hz modulation frequency.

Figure 6.5F is a clearly perceptible imitation of "hello" by the same bird. Note, first of all, the similar overall timing (about 0.4 sec) to the utterance in Figure 6.5D, and the similar concentration of energy - representing F2 - at a little over 1,000 Hz. Aspiration energy at the beginning of the utterance (centred on about 1,000 Hz) is appropriately low for [h]. (There is a weak burst of energy extending over much the same frequency range at the beginning of the utterance in Figure 6.5D. This is followed by a period of silence of about the same duration (70 msec) as the [ʌ] sound in Figure 6.5F.)

It is immediately obvious, particularly from a comparison of the spectra of these two sounds, that frequency modulation is greatly reduced in the more speech-like utterance (Fig. 6.5E). Here the spectrum is reduced in complexity to resemble the simple harmonic structure typical of human vowels. The second harmonic is still dominant and serves, probably with the aid of acoustic resonance (see below), as a second "formant". F3 in this case is represented by the third harmonic. The fundamental of about 640 Hz does double duty,

perceptually, as both the perceived pitch (in the virtual absence of frequency modulation) and F1 frequency. (See Mol, 1971, discussed above in Chapter 4, for similar observations on the use of the fundamental as F1 in grey parrot speech imitation.)

Remez *et al.* (1981) studied the perception of synthetic signals consisting simply of three sinusoids varying independently in time to follow the centre frequencies of the first three formants in a human sentence[3]. They found that this sinusoidal stimulation was sufficient for phonetic perception of the sentence. Listeners were able to report the linguistic content of the signal, while being aware that they were not listening to human speech. The authors concluded that "the process of speech perception makes use of acoustic properties that are more abstract than the spectral templates and speech cues typically studied in speech research". Spectra of these three-component syntheses (Remez and Rubin, 1984) resemble closely the spectrum of the macaw's vowel imitation in Figure 6.5E.

Remez and Rubin (1984) found that a listener's impression of the pitch of such three-component sounds corresponded to the frequency of the component representing the first formant. This was so even when a fourth, low-frequency component, representing the fundamental of the human model, was added. Their work extended into the realm of speech some earlier studies on the perception of complex sounds. Plomp (1967) had shown that, for complex sounds with a fundamental frequency lower than 700 Hz, pitch perception depended on the periodicity of the second and higher harmonics rather than on the frequency of the fundamental. Ritsma (1967) had conducted experiments which showed that the perception of the pitch of complex sounds depended on a relatively narrow band or "dominance region" of the spectrum. "In the majority of cases, a spectral region comprising only two frequency components was as effective

[for pitch perception] as the entire low-frequency band" (Ritsma, *ibid.*). This "dominance region" comprised the third, fourth and fifth harmonics of sounds with a fundamental frequency between 100 Hz and 400 Hz[4].

Remez and Rubin (1984) concluded that "a person listening to speech normally uses the region of the spectrum associated with the first formant to obtain periodicity information as well as to detect the frequency of the first formant itself[5]. Ordinarily, the periodicity of the stimulus in this region and the frequency of the first formant will differ, although in the present case they happen to be identical."

Whispering. This identity of fundamental and first-formant frequency is suggested in the spectrogram (Fig. 6.5G) of an imitation of whispered "hello" by the same macaw. (Figure 6.5H represents the utterance in Figure 6.5F analysed over the same frequency range as Figure 6.5G.)

In acoustic terms, human whispering involves the elimination of the fundamental and its harmonics from the vocalic sounds of speech. (See Chapter 4, under "Ambiguities in interpretation: fine structure".) This is effected from the glottis by turbulent rather than periodic excitation of the formant resonances. Therefore, in wide-bandwidth spectrograms of human whispering, the vertical striations representing periodic closure at the glottis are absent, and are replaced by the densely distributed random markings typical of noise energy. All the formants of the corresponding voiced sound are excited in this random way. Although the macaw's whisper (Fig. 6.5G) shows such noisy excitation of higher formants, F1 is entirely absent. (Compare Figures 6.5G and 6.5H.) The absence of periodicity (i.e., fundamental frequency) in Figure 6.5G, attended by this absence of first-formant energy, is an indication of the identity of fundamental and F1.

Despite this absence of F1 in the macaw's imitation, the timbre of the whispered vowels is retained, demonstrating a certain independence of timbre (related to spectral envelope shape) and pitch (related to fundamental frequency)[6]. In the cockatiel imitation discussed above, fundamental frequency represents the second formant of the speech sound. As the second formant and its transitions are the most important cues in vowel discrimination, it follows that in cockatiel imitations timbre and perceived pitch are interdependent. Second-formant frequency has more autonomy in the speech imitation of larger parrots. Figure 6.5G shows that green-winged macaw "speech" can have energy present in the region of the second and third formants (2,500 Hz - 5,500 Hz) in the absence of voicing. Thus in this species these formants appear to be influenced by a mechanism (e.g., acoustic resonance) quite independent from the syringeal vibration which produces the fundamental. While, as in so many psittacine vowels, the fine structure of the spectrum consists of a fundamental and two harmonics near the frequencies of first, second and third formants respectively (see Figure 6.5E), the speech-like shape of the spectral envelope is enhanced by acoustic resonance - particularly around the second-formant frequency.

Whispering was observed only in macaws, which are the largest birds included in this study. Indeed, all three macaws recorded included whispering in their repertoires, which indicates that it is something they do with peculiar facility. Because of their relatively long vocal tracts, they are more likely than smaller birds to make use of separately controllable acoustic resonance in their vocalizations. (See Chapter 2 for a detailed discussion of this subject.) It would appear, then, from this consideration as well as from the results of acoustic analysis, that acoustic resonance is employed in the vocalizations of these birds.

(As in human whispering, high-frequency resonances are emphasised in the bird's whisper more than they are in voiced sounds; this is because the turbulent excitation contains more high-frequency energy than a periodic, voice-like excitation. In Figure 6.5G the main spectral peak is centred on about 8,000 Hz. This could represent the resonance of a relatively small buccal cavity which is excited much less strongly by periodic acoustic energy from the syrinx.)

AVIAN INTER-SPECIES COMPARISONS

Plate 6.6 is a comparison of a short word containing a diphthong - the word "nice" [nɑɪs] produced by a grey parrot and by a mynah. (The grey parrot here is a different individual from the subject of Figure 6.4F.) An immediate impression of the analyses presented is one of a vowel-like periodicity in the utterences of both birds, particularly noticeable on the spectrograms in the apparent harmonic frequency pattern of the mynah (Fig. 6.6D) and the pulse-like temporal pattern of the parrot (Fig. 6.6B). It will be shown, however, that in neither case can these patterns be attributable simply to a source-generated harmonic spectrum.

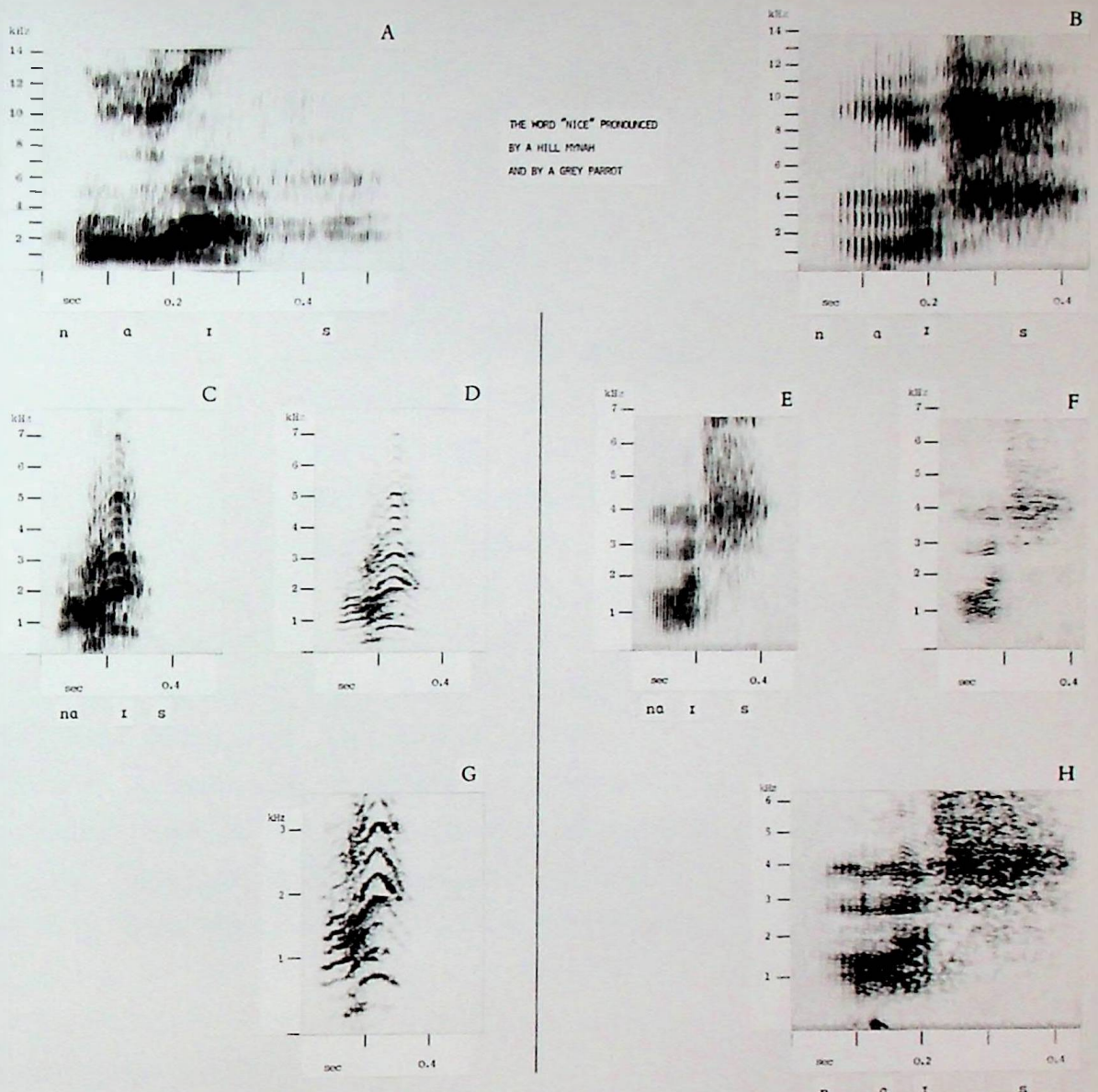
Concentrating initially on the speech-like qualities of these imitations, one of the most striking features of both is the accuracy of their representation of formant frequencies and formant transitions. Looking at Figures 6.6C and 6.6E, remarkable similarities can be seen in their representation of F1 and F2 in the initial [ɑ] half of the diphthong, and the transitions of both formants to the following [ɪ] half - i.e., that part of each spectrogram bounded by 0.2 sec along the time axis and 2 kHz along the frequency axis. In both cases the human pattern of a band of energy representing closely-spaced F1 and F2 centred on about 1,000 Hz and the subsequent upward transition of F2 (to about 2,000 Hz) and downward transition of F1 is clearly discernible.

The similarity between the two ends at this point, however, as the acoustic structure of the mynah's [ɪ] sound is very different from that of the parrot. (See Plate 6.8 for further contrasting examples of high front vowels produced by a mynah and by a parrot.) Disregarding differences in prosodic structure, the mynah's [ɪ] sound consists of an extensive array of apparently harmonic

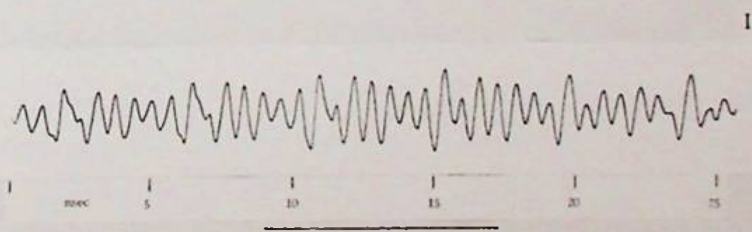
PLATE 6.6

Inter-species comparison

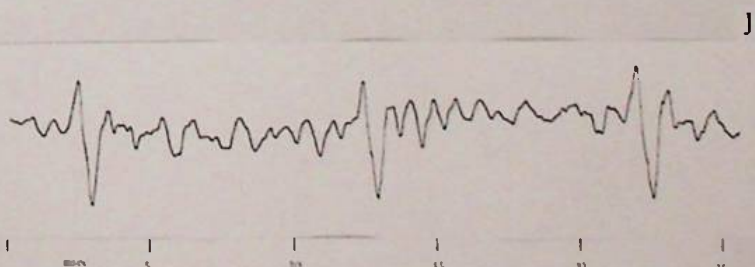
Gracula religiosa *Psittacus erithacus*



Gracula religiosa



Psittacus erithacus



elements (Figs 6.6D and 6.6G) with a formant-like amplitude peak at about 2,000 Hz, while the parrot brings briefly into operation a resonance, either mechanical or acoustic, at about the 2,000 Hz second-formant frequency (Figs 6.6F and 6.6H). The mynah's [ɪ] sound has a well-defined first formant at the appropriately low human frequency of about 400 Hz, whereas the parrot's [ɪ] sound has no first formant.

The parrot's vocal mechanism

This parrot's reproduction of the diphthong resembles its conspecific's version of [əʊ] in Figure 6.4F, in that the high first formant of the first half of the diphthong, [ɑ] is present, together with a representation of the downward transition between the two vocalic segments, while the low first formant of the second vocalic segment [ɪ] is absent.

Figure 6.6H, a narrow-band analysis, made with increased gain, of the lower frequency range of the parrot's utterance, confirms this picture of low energy content below about 400 Hz. The waveform (Fig. 6.6J) of the [ɑ] vocalic section of the parrot's utterance, however, has dominant periodic peaks representing a frequency of about 100 Hz corresponding with the spacing of the vertical striations in this section of the spectrogram (Fig. 6.6B).

By examining the spectrum of this sound (Fig. 6.7B) it can be confirmed that this 100 Hz periodicity of the waveform - a frequency unrepresented in the narrow-band spectrogram - represents a modulation frequency rather than a voice-like pulse. There is no spectral component at 100 Hz; and, more importantly, there are no harmonic components at multiples of 100 Hz. Components are spaced at 100 Hz intervals, however, as a result of amplitude modulation of a complex carrier. It is this periodic intensification of the amplitude spectrum, the components

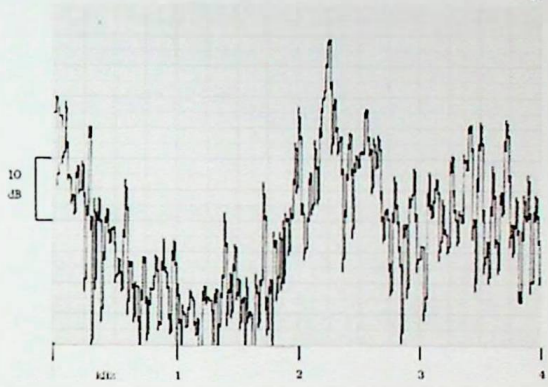
PLATE 6.7

Inter-species comparison

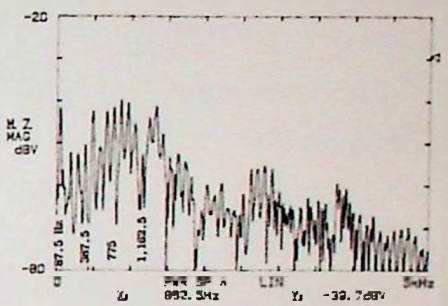
Melopsittacus undulatus

Psittacus erithacus

A

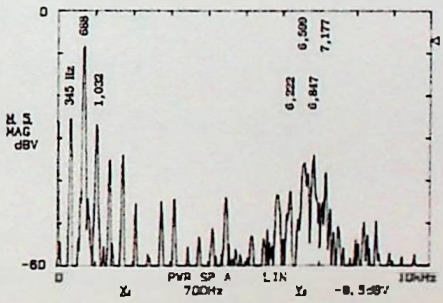


B

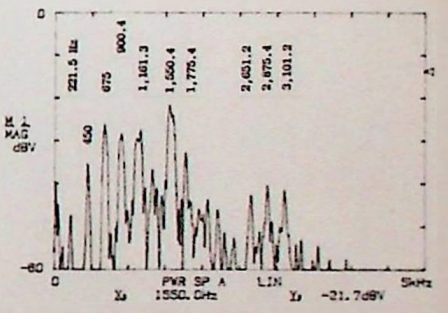


Gracula religiosa

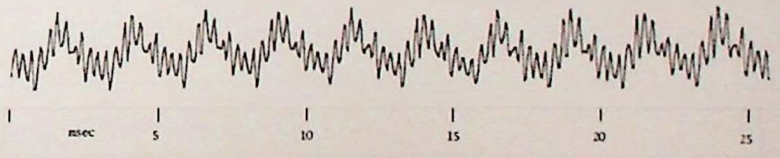
C



D

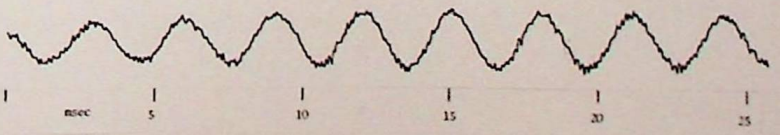


E

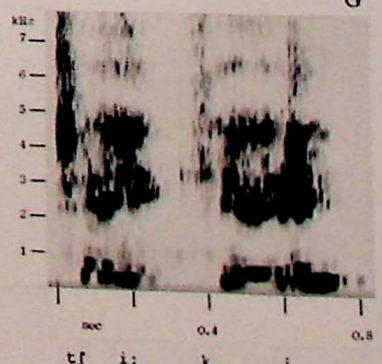


Human voice

F



G



of which are too closely spaced to be resolved by the 300 Hz bandwidth analysing filter of the spectrograph, which produces the vertical striations on the wide bandwidth (i.e., highly damped) filter analysis.

The lowest frequency component of the spectrum (disregarding the 50 Hz mains humm) is at 87.5 Hz. This is not a fundamental, however, as there are no components at multiples of this frequency. The 100 Hz spacing of the components can be most simply explained as an amplitude modulation of a complex sound consisting of a fairly low-frequency fundamental and a series of harmonics. The spectral component at 387.5 Hz is the most likely representative of the fundamental carrier frequency, as it has a second harmonic (775 Hz) and a third harmonic (1,162.5 Hz). If such a complex acoustic vibration were modulated by another harmonic complex, and if the fundamental frequency of the modulator were 100 Hz, the resulting sound would have a spectrum like that in Figure 6.7B: that is, a spectrum consisting of a number of harmonically related components based on a fundamental of 387.5 Hz, and sidebands at intervals of 100 Hz (and harmonic multiples of 100 Hz) above and below those components. Occasional small irregularities in the basically 100 Hz spacing of the spectral components, such as appear in Figure 6.7B, would be a result of the harmonic unrelatedness of the fundamental carrier frequency (387.5 Hz) and the fundamental modulation frequency (100 Hz).

The voice-like perceived pitch of 100 Hz can thus be related to the non-speech-like agency of an amplitude modulation frequency. The resemblance of the spectrogram (Fig. 6.6B) to one of human voicing is a visual correlate of this aural impression.

The mynah's vocal mechanism

Spectrograms of speech imitation by the passerine mynah also exhibit a voice-like pattern of vertical striations (Figs 6.4A, 6.8A). In all mynah recordings analysed, however, the apparent pitch of speech imitations is considerably higher than in the example of grey parrot "speech" discussed above. This apparent pitch is related to a spectral component usually between about 200 Hz and 500 Hz - roughly the range of fundamental frequencies for a female speaking voice. Compare Figure 6.7E - the waveform of a 25 msec section of the sound [ɹu] in "Where are you?" as imitated by a mynah - with Figure 6.7F - the same sound produced by the female trainer. The fundamental frequencies of both model and imitation fall within the 200 Hz - 500 Hz range, the mynah's fundamental being 407 Hz, while that of the trainer is 332 Hz. (Note the characteristic intensity of a higher-frequency component in the mynah's sound, and compare this pair of waveforms with Figures 6.3M and 6.3N. This comparison illustrates the very different acoustic pattern of a budgerigar's attempt at a similar human sound.)

That this lowest component in the vowel spectrum of a mynah, unlike the lowest component in vocalic sounds (such as that illustrated in Figure 6.7B) produced by a grey parrot, represents a fundamental frequency is confirmed by the presence in the mynah sounds of harmonic spectra (Figs 6.8G, 6.8I). Figure 6.7D is the spectrum of a 40 msec section of the [ɑɪ] diphthong in the mynah's imitation of [nɑɪs]. (See Figures 6.6D and 6.6G for narrow-band spectrograms of this utterance.) Comparing this with the spectrum of the parrot's version of the same sound in Figure 6.7B, the harmonic structure of the mynah's utterance, based on the 221.5 Hz fundamental, is in obvious contrast to the structure of the parrot's

spectrum, with its 100 Hz spacing of components but no component at 100 Hz.

More detailed study of the mynah spectrum, however, reveals anomalies which indicate that the sound source involved is very different from the human larynx.

For instance, while most of the peaks in this spectrum (Fig. 6.7D) are harmonics of the 221.5 Hz fundamental, there is an intense peak at 1,161.3 Hz which does not fit into this harmonic pattern.

The simplest way of explaining this spectral pattern would be to invoke the "two-voice" mechanism of songbirds - i.e., their ability to control each half of the syrinx independently (Greenewalt, 1968; Klatt and Stefanski, 1974).

In this case, the harmonic spectrum could be the result of quasi-periodic vibration of <u>one</u> of the lateral labia in a manner analogous to the vibration of the human vocal folds. Simultaneously, the component of about 1,160 Hz could be a membrane resonance excited in the <u>other</u> half of the syrinx - perhaps by aerodynamic perturbation.

This model preserves the distinction (discussed in Chapter 3 in relation to the budgerigar) between muscle-assisted and resonant vibrations, the simultaneous occurrence of which is fundamental to complex acoustic patterns in many bird sounds. In songbirds, however, the "double" syrinx provides the possibility of greater independence of these two vibrations than in parrots. In the present instance, the muscle-assisted vibration (the fundamental) could be produced in one half of the syrinx, and the resonant vibration (representing a speech formant) could be produced in the other half.

Figure 6.7C is the spectrum of another sound produced by the same mynah. The spectral pattern here shows many similarities to that in Figure 6.7D. There is an harmonic spectrum (based, in this case, on a fundamental of 345 Hz). In this case there is a major peak - unrelated to the harmonic spectrum - at 6,847 Hz, which could

represent a membrane resonance.

This vocalization of the mynah (Fig. 6.7C) seems to be analogous to some of the "pre-performance warble" vocalizations of the budgerigar; i.e., it is a preparation for true speech imitation. It is unusual in that both the low-frequency fundamental at 345 Hz, and a high-frequency whistle-like component can be heard simultaneously. This high-pitched component in the perceived sound is probably related to the spectral peak at 6,807 Hz. In true speech imitations, such as that represented by Figure 6.7D, the resonant frequency of the membrane is lowered, by relaxation of tension in the syringeal muscles, to a level appropriate for the second formant of human vowel sounds.

While the high-frequency resonance in the sound represented by Figure 6.7C is heard distinctly, the lower resonance in the speech imitation sound (Fig. 6.7D) is not heard as a distinct tone, but rather contributes significantly to the timbre of the sound. The close association of the 1,161.3 Hz resonance with a component of the harmonic spectrum - the fifth harmonic at 1,100.4 Hz - may contribute to the typical, rather grating quality of mynah "speech".

"Speech" patterns in mynah and parrot

Plate 6.8 is designed to illustrate some differences between the speech imitation of *Gracula religiosa* (a passerine species), and *Melopsittacus undulatus* (a psittaciform species), in the light of the above discussion of possible vocal mechanisms in parrots and songbirds.

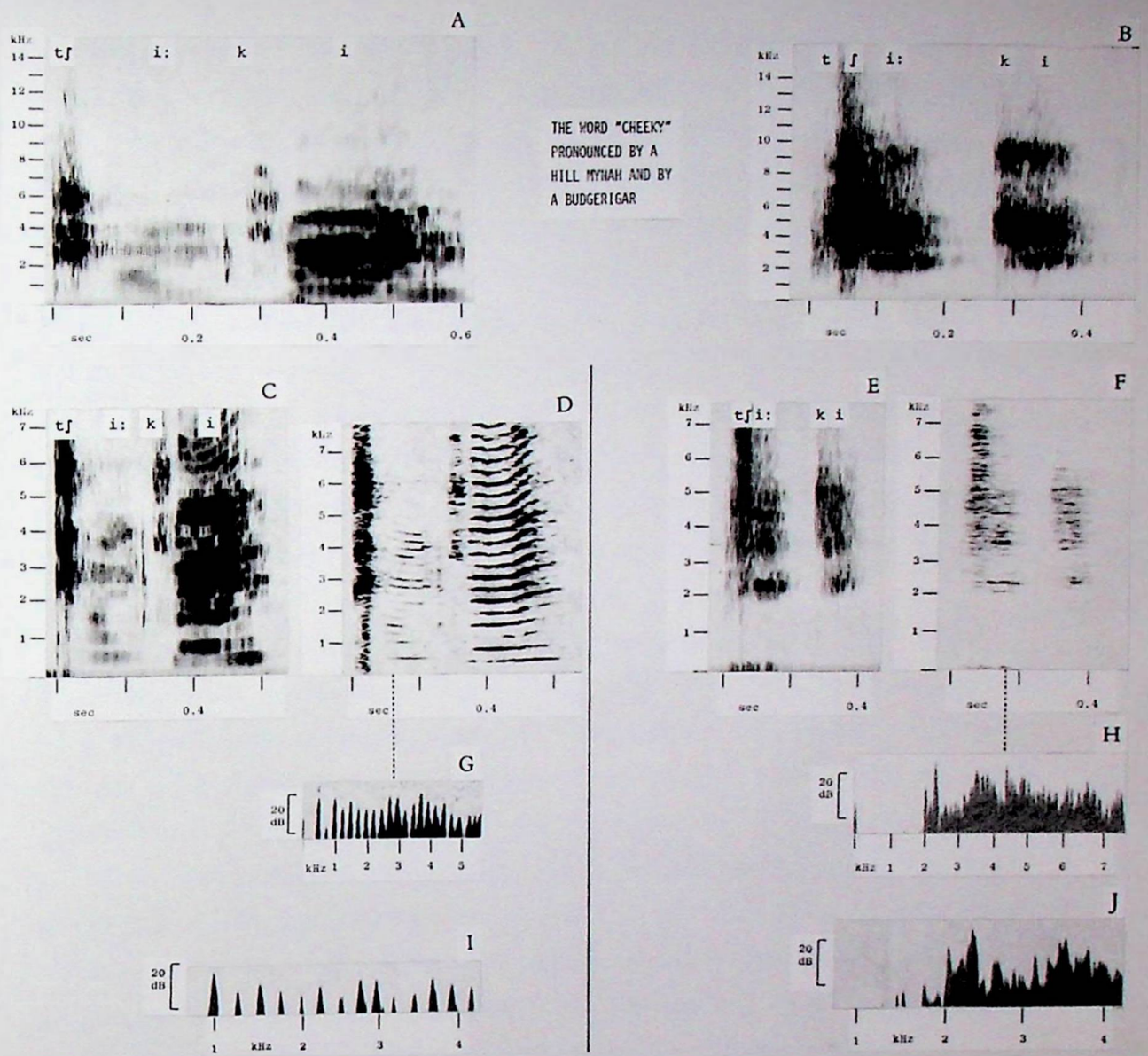
The two birds featured on Plate 6.8, both of which were trained by women, are pronouncing the same word - "cheeky" [tʃi:ki].

Leaving aside differences in overall timing, which may be largely dependent on differences in training, it is

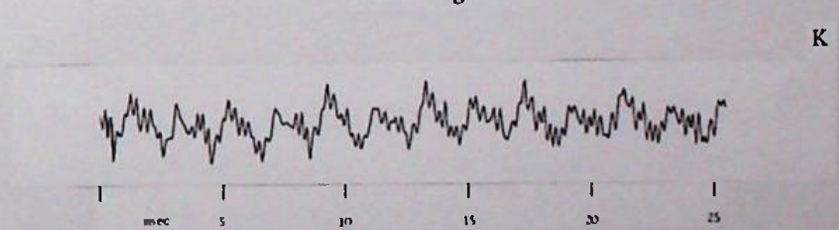
PLATE 6.8

Inter-species comparison

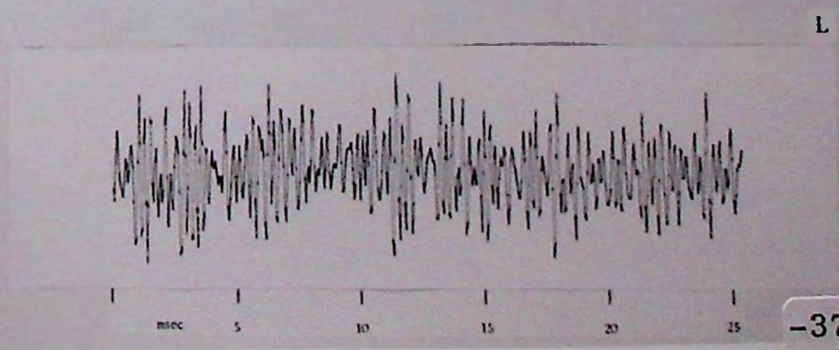
Gracula religiosa — *Melopsittacus undulatus*



Gracula religiosa



Melopsittacus undulatus



obvious from the figures that there are important differences in the finer temporal structure, in segmentation, in frequency distribution, and in amplitude range.

Frequency/time analyses (Figs 6.8A and 6.8B) and amplitude/time analyses (Figs 6.8K and 6.8L) show much more obvious periodicity in the vocalic segments of the mynah's utterance than in the corresponding segments of the budgerigar sound. (Figures 6.8C and 6.8D, 6.8E and 6.8F are wide and narrow bandwidth spectrograms of the same utterances as in Figs 6.8A and 6.8B, only analysed at a faster speed.) These temporal features are reflected in the spectra (Figs 6.8G and 6.8H, and magnified portions of those spectra - Figs 6.8I and 6.8J)., which reveal the apparently harmonic structure of the mynah's first vowel and the much more complex structure of the same vowel in the budgerigar's utterance.

This first [i] sound in the mynah's "cheeky" has a fundamental frequency of about 250 Hz, while the second [i] sound has a fundamental frequency of about 370 Hz. These are frequencies within the range (200-500 Hz, see above) of vocal fold vibrations during a woman's speech.

The comparable analyses of the budgerigar's "cheeky" show no such speech-like regularity. In the first vowel, there is a marked peak at about 2,330 Hz, and relatively intense components about 330 Hz above and below this peak (Fig. 6.8J). This suggests, at first sight, a mechanism similar to that hypothesized in Chapter 3. However, the near harmonic-relatedness of these components, and the presence - in a more detailed analysis of the same sound (Fig. 6.7A) - of a weak harmonic spectrum based on a fundamental of about 300 Hz, indicate rather the (closely related) mechanism of myogenic "phonation" proposed in relation to Figure 6.3N, above. The waveform representing 25 msec of this first [i] sound in the budgerigar's vocalization (Fig. 6.8L) shows an envelope similar to, but much less regular in

its periodicity than that in Figure 6.3N, representing a different budgerigar's realization of the same phoneme. It certainly bears little resemblance to the same budgerigar's attempt at a more voice-like waveform, which may be related to a myoelastic mechanism (Fig. 6.3J). Detectable periodicity in the envelope of Figure 6.8L appears to be related to the spectral spacing of 300 Hz. The degree of frequency modulation apparent around the major peak in the spectrum (Fig. 6.7A) could contribute to the irregularity of the waveform.

The positive correlation between overall intensity and fundamental frequency in the two [i] sounds of the mynah's utterance - i.e., the greater acoustic energy in the second [i], with its higher fundamental, in relation to the first - may be the result of physiological constraints. Recordings of mynahs in the wild[5] contain many tokens of a whistled - physically sinusoidal - glide, descending gradually from about 3,000 Hz to 1,000 Hz, in which intensity always diminishes along with frequency.

Compared with that of the mynah, the budgerigar's vocalization shows much less variation in intensity, although the second syllable (particularly in the F2 region) is generally less intense than the first.

Another important difference between the speech imitation of these two birds is the very precise segmentation of the mynah's word in comparison with that of the budgerigar. The mynah's two plosive bursts are well-defined, and the sections of fricative and aspiration noise are clearly distinct from their ensuing vocalic segments. These are not characteristics of the budgerigar's utterance.

In formant structure, the low F1 and high F2 of human [i] are reproduced by the mynah, as well as higher "formants". (Note, however, the simultaneously rising and falling formants between 0.1 sec and 0.2 sec on Figure 6.8A - a pattern more characteristic of a diphthong such

as [ai].) Figure 6.7G represents a token of "cheeky" as spoken by the mynah's woman trainer. The mynah's first three formants (of the first [i]) at about 500 Hz, 2,800 Hz and 3,700 Hz (Fig. 6.8G) can be compared with the formants of the trainer's corresponding vowel at about 250 Hz, 2,200 Hz and 3,100 Hz. Thus the mynah is seen to be reproducing formant frequency ratios rather than absolute formant frequencies. This example corroborates the data of Klatt and Stefanski (1974), who wrote: "The results clearly show that the mynah does not make exact formant frequency matches to the different vowels, but tends to mimic relative formant changes instead".

SUMMARY

Avian "speech" demonstrates the flexibility of a bird's sound-production mechanism: for example, a budgerigar employs a number of vocal techniques in producing the different vowel sounds. The particular technique employed for a particular vowel depends on the formant structure of that vowel.

In some respects, there is more consistency among tokens of a "speech" sound produced by a bird than among tokens produced by a human speaker. This is particularly so in the time domain. A bird's imitation of the temporal feature of voice onset time, for example, can be remarkably consistent. As well as some exact correspondences, temporal patterning includes some differences from the model. (A budgerigar phrase is often shorter, overall, than the human model.) The overall rhythm, however, is usually well preserved.

An example of "speech" by a cockatiel (*Nymphicus hollandicus*) shows the intonation pattern following the F2 contour, and in no way resembling the intonation pattern of the human model. This extreme example illustrates the overriding significance of the second formant - in terms of both perception and production - in much avian speech-copying behaviour. Throughout the acoustic data, spectral details associated with the second formant - particularly second-formant transitions - are the aspects of avian "speech" which most closely resemble the human models.

The fundamental frequency and first formant are often represented, in psittacine "speech", by a single spectral component at the first-formant frequency.

Measurements of formant frequencies confirm earlier reports (Klatt and Stefanski, 1974; Thorpe, 1969) that a mynah reproduces the formant-frequency *ratios* rather than

the absolute formant frequencies of the model.

On the basis of the acoustic data, hypothetical production mechanisms in mynahs and parrots are compared.

NOTES TO CHAPTER 6

[1]In none of its tokens of [pʊsi kæʔ] does the bird reproduce final [t]. Final [t] is also not apparent in the spectrograms of the trainer's utterance on Plate 6.3. While this is unrepresentative of her pronunciation in the citation-form examples on my recording, it is probably typical of her normal pronunciation to the bird. Glottal closure, marked by irregularity of glottal vibration at the end of [æ] in this example (Fig. 6.3A), is well imitated by the bird in all his tokens of the phrase. Note that the thick vertical line at the right-hand end of each spectrogram is an artifact of the spectrograph and does not represent a plosive burst.

[2]Forshaw (1973) describes this bird's call (in captivity) as "a raucous, deep raw - awk . . . raw - awk".

[3]Bouman and Kucharski (1929) synthesized vowels using only two sinusoidal components. Whistled speech (reviewed in Busnel and Classe, 1976) employs a quasi-sinusoidal signal varying within the range of F2 frequencies to convey all the essential phonemic and grammatical information of articulated speech. (See Chapter 3.)

[4]"This spectral region also covers that part of the existence region of the tonal residue for an amplitude-modulated sinusoid that is most sensitive to modulation depth in producing a tonal residue" (Ritsma, 1967). See Ritsma (1962) for his investigation of the "tonal residue", which is discussed in Chapter 3, above, in connexion with amplitude modulation in avian speech-like sounds.

[5]Helmholtz (1885), in discussing the ability of the human ear to detect harmonics in speech sounds, wrote: "Seiler of Leipzig relates that while listening to the chant of the watchman during a sleepless night, he occasionally heard at first, when the watchman was at a distance, the twelfth of the melody, and afterwards the prime tone" (p. 104). Smith (1947) investigated the audibility of harmonics in vowel sounds, and described F1 in terms of the intensification of a particular harmonic (which differed in number for different groups of vowels).

[6]Meyer-Eppler (1957) showed that, in the perception of whispered speech, pitch changes in certain vowels can be related to changes in frequency of one or more of the formants - especially F2.

APPENDIX TO CHAPTER 6

One of the longest sentences recorded was the nursery-rhyme couplet "Georgy Porgy puddin' and pie/ kissed the girls and made them cry" in the repertoire of the budgerigar Lucky. Seven sequences involving this couplet were recorded, and it was noticed, first of all, that slight verbal and morphemic (i.e., linguistically significant) differences among and within sequences followed consistent patterns. The seven sequences are as follows:

1. (i) Georgy Porgy puddin' and pie
   Kissed the girls and - and made them cry
   Kissed the girls and
   (ii) (G)orgy Porgy Porgy puddin' and pie
2. Georgy Porgy puddin' and pie
3. Georgy Porgy puddin' and pie
   Kissed the
4. Georgy Porgy puddin' and pie
   Kissed the girl made them cry
   Kissed the girls and made them cry
5. (i) Georgy Porgy Porgy puddin' and pie
   Kissed the girl made them cry
   Kissed the girls and - and made them cry
   (ii) Georgy Porgy Porgy puddin' and pie
   Kissed the girls and
6. (i) Georgy Porgy Porgy puddin' and pie
   Kissed the girls and made them
   (ii) Georgy Porgy puddin' and pie
   Kissed the girls and - and made them cry
   Kissed the girl made them
7. Georgy Porgy Porgy puddin' and pie
   Kissed the girls and - and made them cry

It can be seen, for example, that when the final [z] of [gɜlz] is omitted, the following [æn] is always omitted. In four cases out of five, when the [z] of [gɜlz] is retained, the following [æn] is repeated.

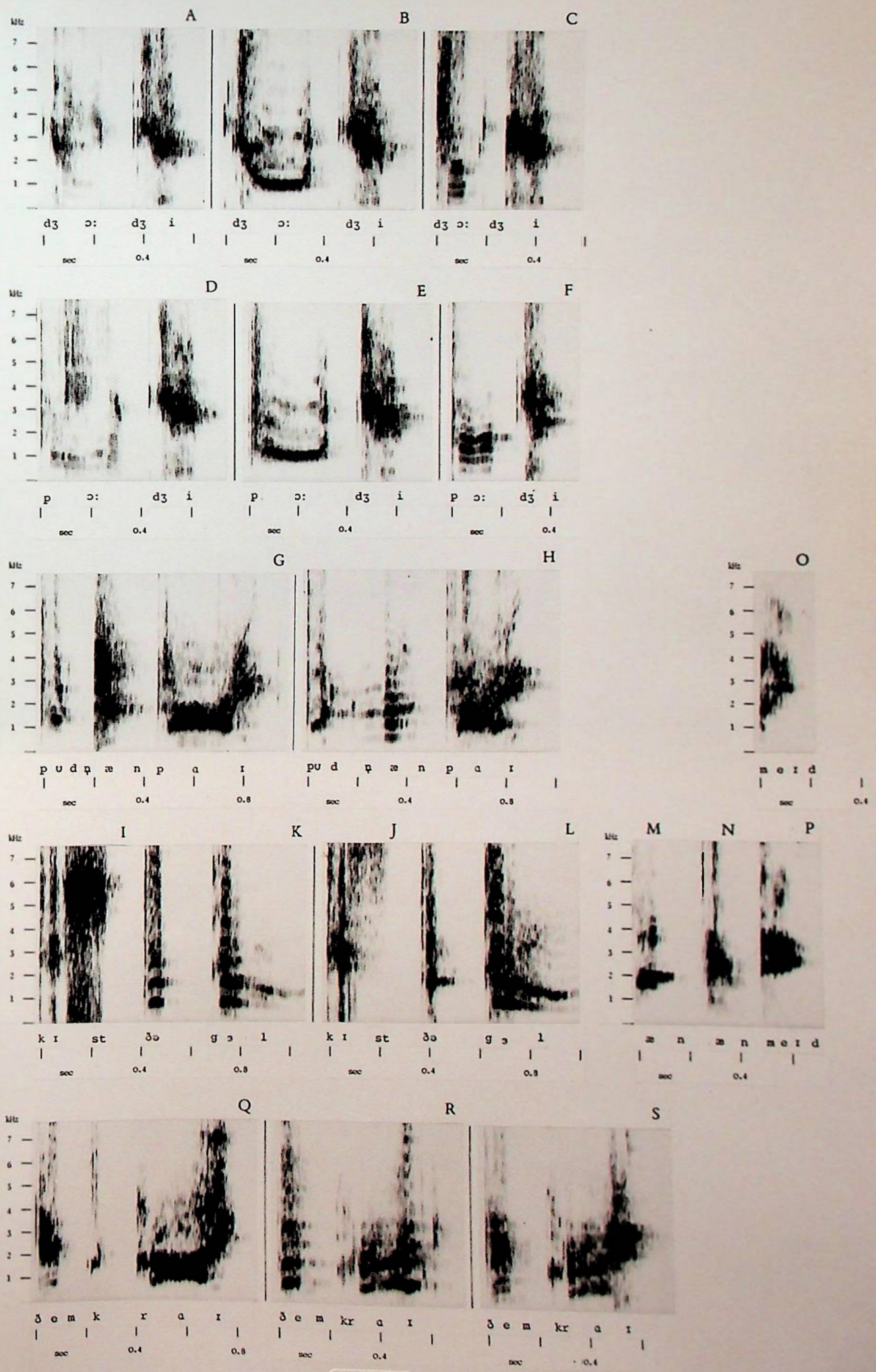
Chapter 3 dealt at some length with a difference on the acoustic level between two versions of the word [pɔdʒi] as it occurs in this sequence. It was pointed out that this acoustic (non-phonetic) difference was associated with a verbal difference - the repetition or non-repetition of the word [pɔdʒi]. (See the transcription of sequences in the preceding paragraph, and the discussion in Chapter 3.) Thus the obvious <u>verbal</u> contrast (either one or two "Porgy"s) can be seen as signalling an underlying non-speech <u>acoustic</u> contrast.

The spectrographic analysis reported on below was, therefore, undertaken in order to see whether such contrastive patterns occurred throughout the bird's rendition of the couplet. Such patterns could provide evidence of species-specific organization principles underlying avian sentence imitations in the absence of the phonological, semantic and syntactic principles governing sentence organization for a human speaker.

The major difference between the two versions of [pɔdʒi] described in Chapter 3 involves alternative renditions ("allophones") of the speech phoneme /ɔ/. The same alternatives are available for imitation of the same phoneme in [dʒɔdʒi], and there are indeed two versions of [dʒɔdʒi] (Figs 6.9A and 6.9B) which are analogous to the two versions of [pɔdʒi] (Figs 6.9D and 6.9E) described in Chapter 3. There is a third, much less common, version of each word (Figs 6.9C and 6.9F), in which the spectrum of the [ɔ] sound contains a greater number of intense harmonic components than the spectrum of [ɔ] in Figures 6.9A and 6.9D. There is considerable low-frequency energy, with a peak at about 1,800 Hz. This sound is similar to one described in connexion with "pre-performance warble" and depicted in Figure 5.2D.

Underlying "sentence" structure

Melopsittacus undulatus



(From now on, the variations represented by the figures on Plate 6.9 will be referred to simply as "A", "B", "C", etc.)

Analysis shows that all renditions of the phrase [pʊdn̩ æn pɑɪ] are, acoustically, of two distinct kinds (G and H). G differs most noticeably from H in having predominantly noisy rather than periodic energy in [æn]. The nasal [n] in [pʊdn̩] is more pronounced in H, and its harmonic pattern is continuous with that of the following [æn]. The first section of the diphthong [ɑɪ] is shorter in H than in G, while the formant pattern in the final section of the diphthong differs between the two versions.

The organization of the first half of the couplet, in all seven sequences, can now be represented as follows using the letters A to H: -

Georgy Porgy (Porgy) puddin' and pie

1. (i)	B	E		G
(ii)	A	D	E	G
2.	A	D		G
3.	A	E		G
4.	B	F		H
5. (i)	A	D	E	G
(ii)	A	D	E	G
6. (i)	A	D	E	G
(ii)	C	F		H
7.	A	D	E	G

There is a dominant pattern - ADEG. The less common arrangements also follow patterns: for example, an initial B is always followed by E or F rather than D; E is always followed by G; F is always followed by H.

At this underlying level of acoustic patterning, then, the bird can be seen to be organizing contrastive units into consistent patterns which bear no relationship to

the phonological, semantic or syntactic structure of the human sentence. These units (usually between 0.6 sec and 1 sec in duration) do correspond, however, to certain prosodic elements in the human utterance, and may indicate an inter-specific basis for the temporal organization of bio-communication sounds.

Turning to the second half of the couplet, it was found that the bird uses two acoustic patterns - labelled I and J - for the word [kɪst]. The first pattern, I, in contrast to the second pattern, J, has an intense concentration of fricative noise energy around 6,000 Hz in [s].

There are two versions, too, of the phrase [ðə gɜl] (not considering, for the moment, the presence or absence of the final [z]). The first of them, K, has noticeably more energy in harmonic components, and less noise energy, than the second version, L. There are only two patterns (out of a possible four) which the bird generally uses in combining I, J, K and L: they are IK and JL. Only once in the recorded sequences is there an exception - the use of JK in sequence 7.

Another level of patterning can be discerned here, however; it concerns the presence or absence of the final [z] in [gɜlz]. In every example in which the [z] is omitted, the pattern is JL. When the [z] is included, the pattern is IK, except in two instances when the sequence breaks off after [ænd]. (See the list of sequences below.) Here, then, the underlying acoustic patterning comes to the surface in a linguistically significant contrast - presence or absence of the plural-signifying [z].

The word [ænd] in this final half of the couplet has two acoustic patterns (M and N), the second of them (N) being reserved for instances in which the word is repeated. The first version (M) consists, acoustically, of an harmonic spectrum with very little energy in the fundamental (at about 340 Hz) and the lower harmonics.

The main amplitude peak in its spectral envelope is centred on the fifth harmonic at about 1,700 Hz, which is an appropriate frequency for F2 of [æ]. There is a secondary peak centred on the tenth harmonic. The second version (N) of the word has a lower fundamental (at about 200 Hz), and energy distributed over a wider range of harmonics, including those at the lower end of the spectrum. The major peak in the spectrum of N, as in that of M, is at about 1,700 Hz. The comparative spectrograms give the impression that N has a much more speech-like acoustic structure than M.

The acoustic structure of the word [meɪd] depends on that of the [ænd] which precedes it. Thus there are two versions of [meɪd] - O, which follows M, and P, which follows N. When there is no preceding [ænd], the O version of [meɪd] is used. O has an initial section in which the formant structure of [e] is clearly represented (at about 800 Hz and 1,800 Hz) in a steady state for about a twentieth of a second. This section is virtually absent in P. In both O and P there is the pattern of a rising second-formant transition typical of the diphthong [eɪ], and a concentration of energy between 2,000 Hz and 3,000 Hz representing the second and third formants of the [ɪ] part of the diphthong. This is more intense in P, however, and has the "chrp"-like appearance of a similar sound in Figure 5.3G. There is very little energy below 2,000 Hz in P.

The last two words of the sequence - [ðem kraɪ] - can be regarded as a unit because, in six of the seven sequences, they are combined in a consistent pattern labelled Q. The main acoustic feature of [ðem] in Q is an intense concentration of energy over a broad band between about 1,500 Hz and 4,000, while the diphthong in [kraɪ] follows the same acoustic pattern as that in [paɪ]. Two alternatives to Q, labelled R and S, occur only once each - both in the fourth sequence. R is dominated by harmonics both in [ðem] and in [kraɪ]; in the diphthong

[ɑɪ], the first and second formants follow the harmonic pattern rather than that of the human diphthong. In S, [ðem] is the ame as in Q, while the [ɑ] of [krɑɪ] is like that in R and the [ɪ] like that in Q.

The patterning of the second half of the couplet, then, looks as follows:

Kissed the girl(s) (and) (and) made them cry

1. (i)	I	K	M	N	P	Q
	J	L	M			
3.	J					
4.	J	L			O	S
	I	K	M		O	R
5. (i)	J	L			O	Q
	I	K	M	N	P	Q
(ii)	J	L	M			
6. (i)	I	K	M		O	
(ii)	I	K	M	N	P	Q
	J	L			O	
7.	J	K	M	N	P	Q

It can be seen that the dominant overall pattern here is IKMNPQ. If, however, in the first element of this pattern, the bird replaces I with J, then in five instances out of six the J is followed by L.

Although there are intricately patterned consistencies in each separate half of the couplet, as detailed above, there does not appear to be a pattern which consistently links both halves. (This could indicate a limitation on the length of repeated sequences in the natural vocalizations of*Melopsittacus undulatus* - see below.)

The foregoing analysis has revealed a subtle level of acoustic patterning - one which has no <u>acoustic</u> counterpart in the human model - in an avian speech imitation. These underlying patterns are interpreted as reflecting some aspect of <u>sequential organization</u> in the

bird's species-specific vocalizations. The ability of the bird to construct a recognizable speech imitation on the basis of species-specific organizational principles suggests the presence of a common substratum at this level of sequential organization in both avian vocalization and human speech. (The use of such terms as "phrase" and "syllable" in describing the sequential patterning of bird song is symptomatic of this common substratum. See Armstrong, 1963, for definition and discussion of these terms in relation to bird song.)

It can be seen that the acoustically contrastive units (named by letters of the alphabet) in the analyses of avian imitations figured on Plate 6.9 correspond generally to rhythmic (stress-pattern) units in the human utterance. Thus, for example, units such as A, B and C [dʒɔdʒi], D, E and F [pɔdʒi], K and L [ðə gəl], and Q, R and S [ðem kraɪ] all correspond to rhythmic units with a two-stress pattern in speech. It seems, then, that in the absence of the phonological, semantic and syntactic structure of human speech, the chief organizational principle of an avian "sentence" is one of prosody. Many studies (e.g., Brackenbury, 1978a; Emlen, 1972; Guttinger, 1978, 1981) have demonstrated the importance of rhythmic elements in avian vocal communication sequences.

Although beyond the scope of the present work, detailed analysis of such "species-specific syntagmatic" patterns in the natural sounds of "talking" species would be an important contribution to this subject.

GENERAL CONCLUSION

The common factor which emerges most persistently in cross-species analysis of avian "speech" is an apparent concentration on second-formant spectral detail. Even more than in human speech, a spectral-envelope peak in the second-formant region can be seen as the "hub" (Potter, Kopp and Green, 1947) of the avian "speech" signal.

A consideration of avian vocal anatomy and physiology (Chapter 2) in relation to acoustic data (Chapters 3 and 6) has indicated that many bird vocalizations - including much avian "speech" - is based on vibrations of the syringeal membranes within the second-formant range of frequencies.

In human speech, the second formant is part of a spectral-envelope pattern imposed by acoustic resonances of the vocal tract on a source-generated fine structure (a laryngeal pulse train). In much avian "speech", on the other hand, the "second formant" may be related to the primary source vibration, while waveform periodicity at lower frequencies (representing, for example, the fundamental and first formant of speech) may result from modulations of this "second-formant" carrier.

Data on avian hearing (reviewed in Chapter 4) show that audibility thresholds are lowest in the second/third-formant region (1-3 kHz), and that frequency analysis within this range is more highly selective than in man. Temporal-resolution data suggest that lower-frequency components of the speech signal - i.e., fundamental and first-formant frequencies - may be perceived as amplitude modulations of a second-formant "carrier". This perceptual possibility lends further weight to the interpretation of some "speech" data - in terms of production - as indicating amplitude modulation

of a second-formant carrier.

Acoustic cues associated with the second formant are of great importance in human perception and discrimination of speech sounds. The avian auditory system is selectively responsive to spectral information in the second-formant range; complementarily, the sound-production system allows direct, precise control over "second formant" generation. Therefore, "talking" birds are well equipped, at the levels of sound reception and generation, for the production of "speech" sounds which are recognizable to human ears.

Finally, following the tripartite scheme outlined in the "Aims" at the beginning of Chapter 1, it remains to draw some conclusions as to the central audio-vocal processing of speech sounds by "talking" birds.

An early observation in the course of this research was that budgerigar copies of certain second-formant speech patterns appear to be little more than slight modifications of species-specific vocal patterns. This impression was confirmed by a more thorough investigation of the budgerigar data (Chapter 5).

"Talking" birds are able to capitalize on acoustic features which bird sounds and speech have in common. The fact that these common features involve important acoustic cues in human speech perception suggests that they may have, as well, an analogous significance in avian vocal communication.

This "bio-communication cue" hypothesis is complementary to the new definition of "vocal imitation" advocated and explained in Chapter 1 and adhered to throughout this work. "Vocal imitation" (as distinct from "vocal mimicry") is defined as the copying of sounds which are perceived and processed by the copier according to audio-vocal behaviour patterns appropriate for species-specific vocalizations. Some of these behaviour patterns - such as a "babbling" phase (related to subsong), and the importance of a social bond with the

human model/trainer - are observable in the ontogeny of avian "speech". The "bio-acoustic cue" hypothesis states that the processing of speech by "talking" birds according to audio-vocal patterns appropriate for bird sounds is facilitated by (1) acoustic cues which are common to bird sounds and speech, and (2) feature detectors in both bird and man which are responsive to those cues.

Thus acoustic data such as those presented in Chapter 5, interpreted in the light of a general theory of vocal imitation, may help to explain the central processing involved in avian "speech".

REFERENCES

Abe, I. (1969): "Linguistics of the vocalizations of some mimics". *Bull. Tokyo Inst. Technol.* 93, 21-30.

Abramson, A. S. and Lisker, L. (1970): "Discriminability along the voicing continuum: cross-language tests". *Proc. 6th Int. Congr. Phonet. Sci.*, Prague, Academia.

Abramson, A. S. and Lisker, L. (1973): "Voice-timing perception in Spanish word-initial stops". *J. Phonet.* 1, 1-8.

Ames, P. L. (1971): "The morphology of the syrinx in passerine birds". *Bull. Peabody Mus. Nat. Hist.* 37.

Andrew, R. J. (1969): "The effects of testosterone on avian vocalizations". In *Bird Vocalizations"* (R. A. Hinde, ed.), Cambridge University Press.

Andrew, R. J. (1973): "The evocation of calls by diencephalic stimulation in the conscious chick". *Brain, Behav. Evol.* 7, 424-446.

Armstrong, E. A. (1963): *A Study of Bird Song*, Oxford University Press.

Arnold, A. P. (1980): "Sexual differences in the brain". *Am. Sci.* 68, 165-173.

Arnold, A. P. (1982): "Neural control of passerine song". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds) Vol 1, Academic Press, New York and London.

Baer, T. (1981a): "Investigation of the phonatory mechanism". In *Proceedings of the Conference on the Assessment of Vocal Pathology*, Bethesda, Maryland, 1979 (C. L. Ludlow and M. O'Connell, eds), Am. Speech-Language -Hearing Assn., Maryland.

Baer, T. (1981b): "Observation of vocal fold vibration: measurement of excised larynges". In *Vocal Fold Physiology* (K. N. Stevens and M. Hirano, eds), Tokyo

University Press.

Bailey, E. D. and Baker, J. A. (1982): "Recognition characteristics in covey dialects of bobwhite quail". *Condor* 84, 317-320.

Bailey, P. J. (1979): "Aspects of categorical processing of speech sounds in man: on the determinants of phonetic category boundaries". In *Hearing Mechanisms and Speech* (*Exp. Brain Res.* Suppl. 2) (O. Creutzfeld, H. Scheich and Chr. Schreiner, eds), Springer-Verlag, Berlin and Heidelberg.

Baker, M. C. (1974): "Genetic structure of two populations of white-crowned sparrows with different song dialects". *Condor* 76, 351-356.

Baker, M. C. (1975): "Song dialects and genetic differences in white-crowned sparrows (*Zonotrichia leucophrys*)". *Evolution* 29, 226-241.

Baker, M. C. (1982): "Genetic popultion structure and vocal dialects in *Zonotrichia* (Emberizidae)". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds) Vol. 2, Academic Press, New York and London.

Baker, M. C., Bottjer, S. W. and Arnold, A. P. (1984): "Sexual dimorphism and lack of seasonal changes in vocal control regions of the white-crowned sparrow brain". *Brain Res.* 295, 85-89.

Baldwin, J. M. (1914): "Deferred imitation in West African grey parrots". *IXth Int. Congr. Zool.*, Monaco, 536-537.

Bandura, A. (1969): *Principles of Behaviour Modification*, Holt, Rinehart and Winston, New York.

Bandura, A. (1971): "Analysis of modelling processes". In *Psychological Modelling* (A. Bandura, ed.), Aldine-Atherton, Chicago.

Bandura, A. (1977): *Social Modelling Theory*, Aldine-Atherton, Chicago.

Baptista, L. F. and Morton, M. L. (1981): "Interspecific song acquisition by a white-crowned sparrow". *Auk* 98,

383-385.

Baptista, L. F. and Petrinovich, L. (1984): "Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow". *Anim. Behav.* 32, 172-181.

Baru, A. V. (1975): "Discrimination of synthesized vowels /a/ and /i/ with varying parameters in dog". In *Auditory Analysis and Perception of Speech* (G. Fant and M. A. A. Tatham, eds), Academic Press, New York and London.

Bauk, V. (1971): "Analyse au spectroscope de la voix de la perruche". *Acta Otolaryng.* 71, 181-185.

Baylis, J. R. (1982): "Avian vocal mimicry: its function and evolution". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds) Vol. 2, Academic Press, New York and London.

Baylis, J. R. and McConnell, P. B. (1983): "Interspecific acoustic communication between cooperative herders *H. sapiens* and *C. lupus*". *Proceedings of the 18th International Ethological Conference*, Brisbane.

Beddard, F. E. B. (1898): *The Structure and Classification of Birds*, Longmans, Green and Co., London.

Beebe, W. (1925): "The variegated tinamou (*Crypturus variegatus variegatus* Gmelin)". *Zoologica* 6, 195-227.

Beecher, M. D., Petersen, M. R., Zoloth, S. R., Moody, D. B. and Stebbins, W. C. (1979): "Perception of conspecific vocalizations by Japanese macaques". *Brain Behav. Evol.* 16, 443-460.

Békésy, G. von (1960): *Experiments in Hearing*, McGraw-Hill, New York.

Bertram, B. (1970): "The vocal behaviour of the Indian hill mynah, *Gracula religiosa*". *Anim. Behav. Monogr.* 3, 79-192.

Bever, T. G. and Chiarello, R. J. (1974): "Cerebral dominance in musicians and nonmusicians". *Science* 185, 537-539.

Biederman-Thorson, M. (1970): "Auditory responses of units in the ovoid nucleus and cerebrum (field L) of the ring dove", *Brain Res.* 24, 247-256.

Bierens de Haan, J. A. (1929): "Animal language and its relation to that of man". *Biological Reviews* 4, 249-268.

Bilger, R. C. (1976): "A revised critical band hypothesis". In *Hearing and Davis: Essays Honouring Hallowell Davis* (S. K. Hirsh *et al.*, eds), Washington University Press, St Louis, Mo.

Blevins, R. D. (1977): *Flow-induced Vibration*, Van Nostrand Reinhold, New York.

Blumstein, S. (1974): "The use and theoretical implications of the dichotic listening technique for investigating distinctive features". *Brain Lang.* 1, 337-350.

Bock, W. J. (1978): "Morphology of the larynx of *Corvus brachyrhynchos*". *Wilson Bull.* 90, 553-565.

Boer, E. de (1976): "On the 'residue' and auditory pitch perception". In *Handbook of Sensory Physiology* Vol. V/3 (W. D. Keidel and W. D. Neff, eds), Springer-Verlag, Berlin.

Bohne, B. A. and Dooling, R. J. (1974): "Morphological changes in the ears of noise exposed parakeets", *J. Acoust. Soc. Am.* 55, S77(A).

Bonke, B. A., Bonke, D. and Scheich, H. (1979): "Connectivity of the auditory forebrain nuclei in the Guinea fowl (*Numida meleagris*)", *Cell Tissue Res.* 200, 101-121.

Bonke, D., Scheich, H. and 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", *J. Comp. Physiol.* 132, 243-255.

Boord, R. L. (1965): "Efferent projections of cochlear nuclei of the pigeon", *Am. Zool.* 5, 669.

Boord, R. L. (1968): "Ascending projections of the primary cochlear nuclei and *nucleus laminaris* in the pigeon", *J. Comp. Neur.* 133, 523-542.

Boord, R. L. (1969): "The anatomy of the avian auditory system". *Ann. N.Y. Acad. Sci.* 167, 186-198.

Boord, R. L. and Rasmussen, G. L. (1963): "Projection of the cochlear and lagenar nerves on the cochlear nuclei of the pigeon". *J. Comp. Neurol.* 120, 463-471.

Borror, D. J. and Reese, C. R. (1956): "Vocal gymnastics in wood-thrush songs". *Ohio J. Sci.* 56, 177-182.

Bouman, H. D. and Kucharski, P. (1929): "Synthèses de voyelles a moyen de deux sons simples". *Arch. Néerl. de Phonét. Exp.* 4, 90-111.

Brackenbury, J. H. (1977): "Physiological energetics of cock-crow". *Nature* 270, 433-435.

Brackenbury, J. H. (1978a): "A comparison of the origin and temporal arrangement of pulsed sounds in the songs of the grasshopper and sedge warblers, *Locustella naevia* and *Acrocephalus schoenobaenus*". *J. Zool. Lond.* 184, 187-206.

Brackenbury, J. H. (1978b): "Respiratory mechanics of sound production in chickens and geese". *J. Exp. Biol.* 72, 229-250.

Brackenbury, J. H. (1979a): "Aeroacoustics of the vocal organ of birds". *J. Theor. Biol.* 81, 341-349.

Brackenbury. J. H. (1979b): "Power capabilities of the avian sound-producing system". *J. Exp. Biol.* 78, 163-166.

Brackenbury, J. H. (1980a): "Control of sound production in the syrinx of the fowl *Gallus gallus*". *J. Exp. Biol.* 85, 239-251.

Brackenbury, J. H. (1980b): "Respiration and production of sounds by birds". *Biol. Rev.* 55, 363-378.

Brackenbury, J. H. (1982): "The structural basis of voice production and its relationship to sound characteristics". In *Acoustic Communication in Birds* (Vol. 1), Academic Press, New York and London.

Breazile, J. E. and Yasuda, M. (1979): "Systema Nervosum Peripheriale". In *Nomina Anatomica Avium* (J. J. Baumel *et al.*, eds), Academic Press, London and New York.

Bremond, J. C. (1963): "Acoustic behaviour of birds". In *Acoustic Behaviour of Animals* (R.-G. Busnel, ed.), Elsevier, Amsterdam, London, New York.

Broadbent, D. E. and Gregory, M. (1964): "Accuracy of recognition for speech presented to the right and left ears". *Q. J. Exp. Psych.* 16, 359-360.

Broca, P. (1865): "Du siège de la faculté du langage articulé". *Bull. Soc. Anthrop.* 6, 377-393.

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 *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Brooks, R. J. and Falls, J. B. (1975): "Individual recognition by song in white-throated sparrows: III song features used in individual recognition". *Can. J. Zool.* 53, 1749-1761.

Burdick, C. K. and Miller, J. D. (1975): "Speech perception by the chinchilla". *J. Acoust. Soc. Am.* 58, 415-427.

Burton, P. J. K. (1974): "Jaw and tongue features in Psittaciformes and other orders, with special reference to the anatomy of the tooth-billed pigeon (*Didunculus strigirostris*)". *J. Zool. Lond.* 174, 255-276.

Busnel, R.-G. (1968): "Acoustic communication". In *Animal Communication* (T. A. Sebeok, ed.), Indiana University Press.

Busnel, R.-G. and Classe, A. (1976): *Whistled Languages*, Springer-Verlag, Berlin, Heidelberg, New York.

Busnel, R.-G., Leroy, C. and Moles, A. (1970): "Etude sur la langue sifflée de Kusköy en Turquie". *Revue de*

Phonétique Appliquée, Centre Universitaire de l'Etat de Mons (Belgium), 1-166.

Calder, W. A. (1970): "Respiration during song in the canary (*Serinus canarius*)". *Comp. Biochem. Physiol.* 32, 251-258.

Carlson, R., Fant, G. and Granstrom, B. (1975): "Two-formant models, pitch and vowel perception". In *Auditory Analysis and Perception of Speech* (G. Fant and M. A. A. Tatham, eds), Academic Press, New York and London.

Casey, R. M. (1981): *Theoretical analysis of tympanic membranes in avian syrinx*. M.S. thesis, Ohio State Univ., Columbus, Ohio.

Catchpole, C. K. (1979): *Vocal Communication in Birds*, Edward Arnold, London.

Chamberlain, D. R., Gross, W. B., Cornwell, G. W. and Mosby, H. S. (1968): "Syringeal anatomy in the common crow". *Auk* 85, 244-252.

Chanaud, R. C. (1970): "Aerodynamic whistles". *Sci. Am.* 222, 40-46.

Chauvin, R. and Muckensturm-Chauvin, B. (1980): *Behavioural Complexities*, International Universities Press, New York.

Chomsky, N. (1965): *Aspects of the Theory of Syntax*, M.I.T. Press.

Classe, A. (1957): "Phonetics of the Silbo Gomero". *Archivum Linguisticum* 9, 44-61.

Clayton, N. S. (1987): "Song learning in cross-fostered zebra finches: a re-examination of the sensitive phase". *Behaviour* 102, 67-81.

Cobb, S. (1964): "A comparison of the size of an auditory nucleus (*n. mesencephalicus lateralis, pars dorsalis*) with the size of the optic lobe in twenty-seven species of birds", *J. Comp. Neur.* 122, 271-279.

Cole, R. A. and Scott, B. (1974): "Toward a theory of speech perception". *Psychol. Rev.* 81, 348-374.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst,

J. M. and Gerstman, L. J. (1952): "Some experiments on the perception of synthetic speech sounds". *J. Acoust. Soc. Am.* 24, 597-606.

Cutting, J. E. (1974): "Two left-hemisphere mechanisms in speech perception". *Perc. & Psych.* 16, 601-612.

Cutting, J. E. and Eimas, P. D. (1975): "Phonetic feature analyzers and the processing of speech in infants". In *The Role of Speech in Language* (J. F. Kavanagh and J. E. Cutting, eds), M.I.T. Press.

Cuvier, G. (1805): *Lecons d'Anatomie comparée*, Vol. 6.

Darwin, C. (1871): *The Descent of Man, and Selection in Relation to Sex*, John Murray, London.

Darwin, C. (1980): *Metaphysics, Materialism, and the Evolution of Mind: Early Writings of Charles Darwin* (transcribed and annotated by P. H. Barrett, commentary by H. E. Gruber), University of Chicago Press.

Darwin, C. J. (1971): "Ear differences in the recall of fricatives and vowels". *Q. J. Exp. Psychol.* 23, 46-62.

Davis, J. M. (1973): "Imitation: a review and critique". In *Perspectives in Ethology* (P. P. G. Bateson and P. H. Klopfer, eds), Plenum Press, New York.

Delattre, P. C., Liberman, A. M. and Cooper, F. S. (1955): "Acoustic loci and transitional cues for consonants". *J. Acoust. Soc. Am.* 27, 769-773.

Delattre, P. C., Liberman, A. M., Cooper, F. S. and Gerstman, L. J. (1952): "An experimental study of the acoustic determinants of vowel colour; observations of one- and two-formant vowels synthesized from spectrographic patterns". *Word* 8, 195-210.

Denker, A. (1907): "Das Gehörorgan und die Sprechwerkzeuge der Papageien", J. F. Bergmann, Wiesbaden.

DeVoogd, T. J., Nixdorf, B. and Nottebohm, F. (1982): "Recruitment of additional synapses into a brain network takes extra brain space". *Soc. Neur. Abstr.* 8, 140.

Dewson, J. H. (1964): "Speech sound discrimination by cats". *Science* 144, 555-556.

Dewson, J. H. (1975): "Recall of auditory patterns by monkeys". *J. Acoust. Soc. Am.* 57, S11.

Dewson, J. H. (1977): "Preliminary evidence of hemispheric asymmetry of auditory function in monkeys". In *Lateralization in the Nervous System* (S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes and G. Krauthamer, eds), Academic Press, New York and London.

Dewson, J. H., Pribram, K. H. and Lynch, J. C. (1969): "Effects of ablations of temporal cortex upon speech sound discrimination in the monkey". *Exp. Neurol.* 24, 579-591.

Dittus, W. P. J. and Lemon, R. E. (1969): "Effects of song tutoring and acoustic isolation on the song repertoires of cardinals". *Anim. Behav.* 17, 523-533.

Dittus, W. P. J. and Lemon, R. E. (1970): "Auditory feedback in the singing of cardinals". *Ibis* 112, 544-548.

Divenyi, P. L. and Sachs, R. M. (1978): "Discrimination of time intervals bounded by tone bursts". *Perc. Psych.* 24, 429-436.

Dobkin, D. S. (1979): "Functional and evolutionary relationships of vocal copying phenomena in birds". *Z. Tierpsychol.* 50, 348-363.

Dooling, R. J. (1973): "Behavioural audiometry with the parakeet (*Melopsittacus undulatus*)". *J. Acoust. Soc. Am.* 53, 1,757-1,758.

Dooling, R. J. (1979): "Temporal summation of pure tones in birds". *J. Acoust. Soc. Am.* 65, 1,058-1,060.

Dooling, R. J. (1980): "Behaviour and psychophysics of hearing in birds", in *Comparative Studies of Hearing in Vertebrates* (A. N. Popper and R. R. Fay, eds), Springer-Verlag, New York.

Dooling, R. J. (1982): "Auditory perception in birds", in *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds), Academic Press, New York and London.

Dooling, R. J. (1985): "Perception of vocal signals by the budgerigar", *Abstracts of the Eighth Midwinter Reserch Meeting, Association for Reserch in Otolaryngology*, Clearwater Beach, Florida.

Dooling, R. J. and Haskell, R. J. (1978): "Auditory duration discrimination in the parakeet (*Melopsittacus undulatus*)". *J. Acoust. Soc. Am.* 63, 1640-1642.

Dooling, R. J., Mulligan, J. A. and Miller, J. D. (1971): "Auditory sensitivity and song spectrum of the common canary (*Serinus canarius*)". *J. Acoust. Soc. Am.* 50, 700-709.

Dooling, R. J., Peters, S. S. and Searcy, M. H. (1979): "Auditory sensitivity and vocalizations of the field sparrow (*Spizella pusilla*)". *Bull. Psycho. Soc.* 14, 106-108.

Dooling, R. J. and Saunders, J. C. (1975a): "Hearing in the parakeet (*Melopsittacus undulatus*)". *J. Comp. Physiol. Psychol.* 88, 1-20.

Dooling, R. J. and Saunders, J. C. (1975b): "Auditory intensity discrimination in the parakeet (*Melopsittacus undulatus*)". *J. Acoust. Soc. Am.* 58, 1,308-1,310.

Dooling, R. J. and Searcy, M. H. (1979): "The relation among critical ratios, critical bands, and intensity difference limens in the parakeet (*Melopsittacus undulatus*)". *Bull. Psycho. Soc.* 13, 300-302.

Dooling, R. J. and Searcy, M. H. (1980): "Forward and backward auditory masking in the parakeet (*Melopsittacus undulatus*)", *Hearing Res.* 3, 279-284.

Dooling, R. J. and Searcy, M. H. (1981): "Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*)". *J. Comp. Physiol.* 143, 383-388.

Dooling, R. J. and Searcy, M. H. (1985a): "Temporal integration of acoustic signals by the budgerigar (*Melopsittacus undulatus*)". *J. Acoust. Soc. Am.* 77, 1917-1920.

Dooling, R. J. and Searcy, M. H. (1985b):

"Nonsimultaneous auditory masking in the budgerigar (*Melopsittacus undulatus*)", *J. Comp. Psychol.* 99, 226-230.

Dooling, R. J., Searcy, M. H. and Cohen, M. S. (1982): "Nonsimultaneous masking and temporal summation in the parakeet (*Melopsittacus undulatus*)". *J. Acoust. Soc. Am.* 71, S49.

Dooling, R. J., Zoloth, S. R. and Baylis, J. R. (1978): "Auditory sensitivity, equal loudness, temporal resolving power and vocalizations in the house finch (*Carpodacus mexicanus*)", *J. Comp. Physiol. Psychol.* 92, 867-876.

Dorman, M. F. (1974): "Auditory evoked potential correlates of speech and discrimination". *Perc. & Psych.* 15, 215-220.

DuBrul, E. L. (1958): *Evolution of the Speech Apparatus*, Charles C. Thomas, Springfield, Illinois.

Eales, L. A. (1987): "Song learning in female-raised zebra finches: another look at the sensitive phase". *Anim. Behav.* 35, 1356-1365.

Ehret, G. (1987a): "Categorical perception of sound signals: facts and hypotheses from animal studies". In *Categorical Perception* (S. Harnad, ed.), Cambridge University Press.

Ehret, G. (1987): "Left hemisphere advantage in the mouse brain for recognizing ultrasonic communiction calls". *Nature* 325, 249-251.

Eilers, R. E., Gavin, W. and Wilson, W. R. (1979): "Linguistic experience and phonemic perception in infancy: a cross-linguistic study". *Child Development* 50, 14-18.

Eimas, P. D. and Corbit, J. (1973): "Selective adaptation of linguistic feature detectors". *Cognitive Psychol.* 4, 99-109.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. and Vigorito, J. (1971): "Speech perception in infants". *Science* 171, 303-306.

Ellis, C. J. (1973): "Syringeal histology I: Japanese quail (*Coturnix coturnix japonica*)". *Iowa State J. Res.* 48, 7-46.

Emlen, S. T. (1972): "An experimental analysis of the parameter of bird song eliciting species recognition". *Behaviour* 41, 130-171.

Evans, E. F. and Whitfield, I. C. (1964): "Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat", *J. Physiol.* 171, 476-493.

Evans, H. E. (1969): "Anatomy of the budgerigar". In *Diseases of Cage and Aviary Birds* (M. L. Petrak, ed.), Lea and Febiger, Philadelphia.

Falls, J. B. (1969): "Functions of territorial song in the white-throated sparrow". In *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Fant, G. (1960): *Acoustic Theory of Speech Production*, Mouton, The Hague.

Fant, G. (1982): "The voice source: acoustic modelling". *STL - QPSR* 4, 28-48.

Farabaugh, S. M. (1982): "The ecological and social significance of duetting". In *Acoustic Communication in Birds* Vol. 2 (D. E. Kroodsma and E. H. Miller, eds), Academic Press, New York and London.

Fentress, J. C. (1976): "Behavioural networks and the simpler systems approach". In *Simpler Networks and Behaviour* (J. C. Fentress, ed.), Sinauer Associates, Sunderland, Mass.

Fletcher, H. (1940): "Auditory patterns". *Reviews of Modern Physics* 12, 47-65.

Ford, J. (1987): "Did Australia's birds sing the first songs?". *New Scientist* No. 1562, 33.

Forshaw, J. M. (1973): *Parrots of the World*, Lansdowne Press, Melbourne.

Foss, B. M. (1964): "Mimicry in mynahs". *Br. J. Psychol.* 55, 85-88.

Fourcin, A. J. (1968): "Speech source inference". *IEEE*

Transactions on Audio and Electroacoustics 16, 65-67.

Fourcin, A. J. (1980): *Speech Pattern Audiometry* (Oxford Medical Engineering Series), Clarendon Press, Oxford.

Frishkopf, L. S. and Goldstein, M. H. (1963): "Responses to acoustic stimuli from single units in the eighth nerve of the bullfrog". *J. Acoust. Soc. Am.* 35, 1,219-1,228.

Fukuda, H., Saito, S., Kitahara, S., Isogai, Y., Makino, K., Tsuzuki, T., Kogawa, N. and Ono, H. (1983): "Vocal fold vibration in excised larynges viewed with an x-ray stroboscope and an ultra-high-speed camera". In *Vocal Fold Physiology: contemporary research and clinical issues* (D. M. Bless and J. H. Abbs, eds), College Hill Press, San Diego.

Gabor, D. (1946): "Theory of communication". *J. Inst. Elec. Eng.* 93, 429-457.

Gadow, H. (1896): "Syrinx". In *A Dictionary of Birds* (A. Newton, ed.) Part IV, Adam and Charles Black, London.

Galaburda, A. M., LeMay, M. Kemper, T. L. and Geschwind, N. (1978): "Right-left asymmetries in the brain". *Science* 199, 852-856.

Garrod, A. H. (1873): "On some points in the anatomy of *Steatornis*". *Proc. Zool. Soc. Lond.*, 526-533.

Garrod, A. H. (1874): "On some anatomical peculiarities which bear upon the major divisions of the passerine birds" (Part 1). *Proc. Zool. Soc. Lond.*, 586.

Gaunt, A. S. (1983): "An hypothesis concerning the relationship of syringeal structure to vocal abilities". *Auk* 100, 853-862.

Gaunt, A. S. and Gaunt, S. L. L. (1977): "Mechanics of the syrinx in *Gallus gallus* II: electromyographic studies of *ad libitum* vocalizations". *J. Morph.* 152, 1-19.

Gaunt, A. S. and Gaunt, S. L. L. (1980): "Phonation in the ring dove: implications and variants". *Am. Zool.* 20, 758.

Gaunt, A. S. and Gaunt, S. L. L. (1982):

"Electromyography of the syringeal muscles in parrots". *Am. Zool.* 22, 943.

Gaunt, A. S. and Gaunt, S. L. L. (1985): "Electromyographic studies of the syrinx in parrots (Aves, Psittacidae)". *Zoomorphology* 105, 1-11.

Gaunt, A. S., Gaunt, S. L. L. and Casey, R. M. (1982): "Syringeal mechanics reassessed: evidence from *Streptopelia*". *Auk* 99, 474-494.

Gaunt, A. S., Gaunt, S. L. L. and Hector, D. H. (1976): "Mechanics of the syrinx in *Gallus gallus* I: a comparison of pressure events in chickens to those in oscines". *Condor* 78, 208-223.

Gaunt, A. S., Stein, R. C. and Gaunt, S. L. L. (1973): "Pressure and air flow during distress calls of the starling". *J. Exp. Zool.* 183, 241-262.

Gaunt, A. S. and Wells, M. K. (1973): "Models of syringeal mechanisms". *Am. Zool.* 13, 1,227-1,247.

Gaunt, S. L. L. and Gaunt, A. S. (1980): "Phonation of the ring dove: the basic mechanism". *Am. Zool.* 20, 757.

Gautier, J.-P. and Gautier, A. (1977): "Communication in Old World monkeys". In *How Animals Communicate* (T. A. Sebeok, ed.), Indiana University Press.

Ginsburg, N. (1960): "Conditioned vocalizations in the budgerigar". *J. Comp. Physiol. Psychol.* 53, 183-186.

Ginsburg, N. (1963): "Conditioned talking in the mynah bird". *J. Comp. Physiol. Psychol.* 56, 1,061-1,063.

Goldstein, J. L. (1967): "Auditory spectral filtering and monaural phase perception". *J. Acoust. Soc. Am.* 41, 458-479.

Gordon, H. W. (1975): "Hemispheric asymmetry and musical performance". *Science* 189, 68-69.

Gottlieb, G. and Vandenbergh, J. G. (1968): "Ontogeny of vocalization in duck and chick embryos". *J. Exp. Zool.* 168, 307-326.

Gould, S. J. (1977): *Ontogeny and Phylogeny*, Harvard University Press, Cambridge, Mass.

Gramza, A. F. (1970): "Vocal mimicry in captive budgerigars". *Z. Tierpsychol.* 27, 971-983.

Gramza, A. F. (1972): "Avian vocal mimicry: the phenomenon and its analysis". *Z. Tierpsychol.* 30, 259-265.

Green, S. M. (1975): "Variation of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): a field study". In *Primate Behaviour* Vol. 4 (L. A. Rosenblum, ed.), Academic Press, New York.

Green, S. M. *et al.* (1977): "Comparative aspects of vocal signals including speech". In *Recognition of Complex Acoustic Signals* (T. H. Bullock, ed.), Dahlem Konferenzen, Berlin.

Greenewalt, C. H. (1968): *Bird Song: Acoustics and Physiology*, Smithsonian Institution Press, Washington D.C.

Gregory, R. and Hopkins, P. (1974): "Pupils of a talking parrot". *Nature* 252, 637-638.

Griffin, D. R. (1953): "Acoustic orientation in the oil bird, *Steatornis*". *Proc. Nat. Acad. Sci. U.S.A.* 39, 884-893.

Griffin, D. R. and Suthers, R. A. (1970): "Sensitivity of echolocation in cave swiftlets". *Biol. Bull.* 139, 495-501.

Gross, N. B. and Anderson, D. J. (1976): "Single unit responses recorded from the first order neuron of the pigeon auditory system", *Brain Res.* 101, 209-222.

Gross, W. B. (1964a): "Voice production by the chicken". *Poultry Sci.* 43, 1005-1008.

Gross, W. B. (1964b): "Devoicing the chicken". *Poultry Sci.* 43, 1143-1144.

Grosslight, J. H., Harrison, P. C. and Weiser, C. M. (1962): "Reinforcement control of vocal responses in the mynah bird (*Gracula religiosa*)". *Psychol. Rec.* 12, 193-201.

Grosslight, J. H. and Zaynor, W. C. (1967): "Verbal behaviour in the mynah bird". In *Research in Verbal*

Behaviour and some Neurophysiological Implications, Academic Press, New York and London.

Grosslight, J. H., Zaynor, W. C. and Lively, B. L. (1964): "Speech as a stimulus for differential vocal behaviour in the mynah bird". *Psychon. Sci.* 1, 7-8.

Gruber, H. E. (1974): *Darwin on Man*, Wildwood House, London.

Gwinner, E. and Kneutgen, J. (1962): "Über die biologische Bedeutung der "zweckdienlichen" Anwendung erlernter Laute bei Vögeln". *Z. Tierpsychol.* 19, 692-696.

Häcker, V. (1900): *Der Gesang der Vögel, seine anatomischen und biologischen Grundlagen*, Gustav Fischer, Jena.

Haggard, M. P. (1971): "Encoding and the REA for speech signals". *Q. J. Exp. Psychol.* 23, 34-45.

Hall, J. L. and Goldstein, M. H. (1968): "Representation of binaural stimuli by single units in primary auditory cortex of unanesthetized cats". *J. Acoust. Soc. Am.* 3, 456-461.

Hall-Craggs, J. (1962): "The development of song in the blackbird (*Turdus merula*)". *Ibis* 104, 277-300.

Halperin, Y., Nachshon, I. and Carmon, A. (1973): "Shift of ear superiority in dichotic listening to temporally patterned nonverbal stimuli". *J. Acoust. Soc. Am.* 53, 46-50.

Handford, P. and Nottebohm, F. (1976): "Allozymic and morphological variation in population samples of rufous-collared sparrow (*Zonotrichia capensis*) in relation to vocal dialects". *Evolution* 30, 802-817.

Harnad, S., ed. (1987): *Categorical Perception*, Cambridge University Press.

Harris, C. L., Gross, W. B. and Robeson, A. (1968): "Vocal acoustics of the chicken". *Poultry Sci.* 47, 107-112.

Harris, M. A. and Lemon, R. E. (1972): "Songs of song sparrows (*Melospiza melodia*): individual variation and

dialects". *Can. J. Zool.* 50, 301-309.

Heinz, R. D., Sinnott, J. M. and Sachs, M. B. (1977): "Auditory sensitivity of the redwing blackbird (*Agelaius phoeniceus*) and brown-headed cowbird (*Molothrus ater*)", *J. Comp. Physiol. Psychol.* 91, 1365-1376.

Helmholtz, H. L. F. (1877): *Die Lehre von den Tonempfindungen* (4th ed.). Trans. A. J. Ellis (1885): *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, Longmans, London.

Henschel, G. (1903): "Bullfinch and canary". *Nature* 67, 609-610.

Hérissant, M. (1753): "Recherches sur les organes de la voix des quadrupèdes et celles des oiseaux". *Mem. Acad. Roy. Sci*, Paris. (See also *Collection Academique*, Partie francoise, 1786: "Sur les organes de la voix des quadrupèdes et des oiseaux".)

Hersh, G. L. (1966): *Bird Voices and Resonant Tuning in Helium-air Mixtures*. Ph.D. thesis, Univ. of California, Berkeley.

Hewes, G. W. (1977): "Language origin theories". In *Language Learning by a Chimpanzee: the Lana Project*, Academic Press, New York.

Hirsh, I. J. (1959): "Auditory perception of temporal order". *J. Acoust. Soc. Am.* 31, 759-767.

Homberger, D. G. (1979): "Functional morphology of the larynx in the parrot *Psittacus erithacus*". *Am. Zool.* 19, 988.

Homberger, D. G. (1986): "The lingual apparatus of the African grey parrot, *Psittacus erithacus* L. (Aves: Psittacidae): description and theoretical mechanical analysis". *Ornith. Monogr.* 39.

Hose, B., Langner, G. and Scheich, H. (1983): "Linear phoneme boundaries for German synthetic two-formant vowels". *Hearing Res.* 9, 13-25.

Hulse, S. H. (1987): "Comprative aspects of complex acoustic perception". In *Auditory Processing of*

Complex Sounds (W. A. Yost and C. S. Watson, eds), Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Hulse, S. H., Cynx, J. and Humpal, J. (1984): "Cognitive processing of pitch and rhythm structures in birds". In *Animal Cognition* (H. L. Roitblat, T. G. Bever and H. S. Terrace, eds), Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Hutchison, R. E., Stevenson, J. G. and Thorpe, W. H. (1968): "The basis for individual recognition by voice in the sandwich tern". *Behaviour* 32, 150-157.

Huxley, T. H. (1871): *A Manual of the Anatomy of Vertebrated Animals*, J. and A. Churchill, London.

Immelmann, K. (1969): "Song development in the zebra finch and other estrildid finches". In *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Jakobson, R., Fant, G. M. and Halle, M. (1963): *Preliminaries to Speech Analysis*, MIT Press.

James, A. E., Hutchins, G., Bush, M., Natarajan, T. K. and Burns, B. (1976): "How birds breathe: correlation of radiographic with anatomical and pathological studies". *J. Am. Vet. Radiol. Soc.* 17, 77-86.

Jenkins, P. F. (1977): "Cultural transmission of song patterns and dialect development in a free-living bird population". *Anim. Behav.* 25, 50-78.

Joos, M. (1948): "Acoustic phonetics", *Language* 24 (Suppl.), 1-136.

Jouventrin, P., Guillotin, M. and Cornet, A. (1979): "Le chant du Manchot empereur et sa signification adaptative". *Behaviour* 70, 231-250.

Jusczyk, P. W., Pisoni, D. B., Walley, A. and Murray, J. (1980): "Discrimination of relative onset time of two-component tones by infants". *J. Acoust. Soc. Am.* 67, 262-270.

Kaneko, T., Uchida, K., Suzuki, H., Komatsu, K., Kanesaka, T., Kobayashi, N. and Naito, J. (1981): "Ultrasonic observations of vocal fold vibration". In

Vocal Fold Physiology (K. N. Stevens and M. Hirano, eds), Tokyo University Press.

Karten, H. J. (1967): "The organization of the ascending auditory pathway in the pigeon (*Columba livia*) I: diencephalic projections of the inferior colliculus (*nucleus mesencephalicus lateralis, pars dorsalis*)", *Brain Res.* 6, 409-427.

Karten, H. J. (1968): "The ascending auditory pathway in the pigeon (*Columba livia*) II: telencephalic projections of the *nucleus ovoidalis thalami*", *Brain Res.* 11, 134-153.

Kelly, D. B. and Nottebohm, F. (1979): "Projections of a telencephalic auditory nucleus - field L - in the canary". *J. Comp. Neurol.* 183, 455-470.

Kimura, D. (1961a): "Some effects of temporal-lobe damage on auditory perception". *Can. J. Psychol.* 15, 156-165.

Kimura, D. (1961b): "Cerebral dominance and the perception of verbal stimuli". *Can. J. Psychol.* 15, 166-171.

Kimura, D. (1964): "Left-right differences in the perception of melodies". *Q. J. Exp. Psychol.* 16, 355-358.

King, A. S. (1975): "Aves respiratory system". In Sisson and Grossman's *The Anatomy of the Domestic Animals* (R. Getty, ed.), W. B. Saunders Co., Philadelphia.

King, A. S. (1979): "Systema Respiratorium". In *Nomina Anatomica Avium* (J. J. Baumel *et al.*, eds), Academic Press, London and New York.

King, A. S. and McLelland, J. (1984): *Birds, their Structure and Function*, Baillière Tindall, London.

Klatt, D. H. and Stefanski, R. A. (1974): "How does a mynah bird imitate human speech?". *J. Acoust. Soc. Am.* 55, 822-832.

Klingholz, F. (1979): "Supplement to 'How does a mynah bird imitate human speech?'". *J. Acoust. Soc. Am.* 65, 537.

Koehler, O. (1972): "The ability to imitate human speech

in parrots". In *Grzimek's Animal Life Encyclopedia* (B. Grzimek, ed.), Van Nostrand Reinhold, New York.

Konishi, M. (1963): "The role of auditory feedback in the vocal behaviour of the domestic fowl". *Z. Tierpsychol.* 20, 349-367.

Konishi, M. (1964): "Effects of deafening on song development in two species of juncos". *Condor* 66, 85-102.

Konishi, M. (1965a): "Effects of deafening on song development in American robins and black-headed grosbeaks". *Z. Tierpsychol.* 22, 584-599.

Konishi, M. (1965b): "The role of auditory feedback in the control of vocalization in the white-crowned sparrow". *Z. Tierpsychol.* 22, 770-783.

Konishi, M. (1969a): "Time resolution by single auditory neurones in birds". *Nature* 222, 566-567.

Konishi, M. (1969b): "Hearing, single-unit analysis, and vocalizations in songbirds". *Science* 166, 1178-1181.

Konishi, M. (1970): "Comparative neurophysiological studies of hearing and vocalizations in songbirds". *Z. Vergl. Physiologie* 66, 257-272.

Konishi, M. (1973): "How the owl tracks its prey", *Am. Scient.* 61, 414-424.

Konishi, M. (1974): "Hearing and vocalization in songbirds". In *Birds: Brain and Behaviour* (I. J. Goodman and M. W. Schein, eds), Academic Press, New York.

Konishi, M. and Nottebohm, F. (1969): "Experimental studies in the ontogeny of avian vocalizations". In *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Kroodsma, D. E. (1977): "A re-evaluation of song development in the song sparrow". *Anim. Behav.* 25, 390-399.

Kroodsma, D. E. (1978): "Aspects of learning in the ontogeny of bird song: where, from whom, when, how many, which, and how accurately?". In *The Development*

of Behaviour: Comparative and Evolutionary Aspects (G. M. Burghardt and M. Bekoff, eds), Garland STPM Press, New York and London.

Kroodsma, D. E. (1979): "Vocal duelling among male marsh wrens: evidence for ritualized expressions of dominance/subdominance". *Auk* 96, 506-515.

Kroodsma, D. E. (1981): "Ontogeny of bird song". In *Behavioural Development* (K. Immelmann, G. W. Barlow, L. Petrinovich and M. Main, eds), Cambridge University Press.

Kroodsma, D. E. (1982): "Vocal learning in birds". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds) Vol. 2, Academic Press, New York and London.

Kroodsma, D. E. and Baylis, J. R. (1982): "A world survey of evidence for vocal learning in birds". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller eds) Vol. 2, Academic Press, New York and London.

Kroodsma, D. E. and Pickert, R. (1984): "Sensitive phases for song learning: effects of social interaction and individual variation". *Anim. Behav.* 32, 389-394.

Kroodsma, D. E. and Verner, J. (1978): "Complex singing behaviours among *Cistothorus* wrens". *Auk* 95, 703-716.

Kuhl, P. K. (1976): "Speech perception by the chinchilla: categorical perception of synthetic alveolar plosive consonants". *J. Acoust. Soc. Am.* 60, S81(A).

Kuhl, P. K. (1979): "Models and mechanisms in speech perception: species comparisons provide further contributions". *Brain Behav. Evol.* 16, 374-408.

Kuhl, P. K. (1981): "Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound ctegories". *J. Acoust. Soc. Am.* 70, 340-349.

Kuhl, P. K. (1987): "The special-mechanisms debate in speech research: categorization tests on animals and infants". In *Categorical Perception* (S. Harnad, ed.),

Cambridge University Press.

Kuhl, P. K. and Miller, J. D. (1975a): "Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants". *Science* 190, 69-72.

Kuhl, P. K. and Miller, J. D. (1975b): "Speech perception by the chinchilla: phonetic boundaries for synthetic VOT stimuli". *J. Acoust. Soc. Am.* 57, S549(A).

Kuhl, P. K. and Miller, J. D. (1975c): "Speech perception in early infancy: discrimination of speech-sound categories". *J. Acoust. Soc. Am.* 58, S56.

Kuhl, P. K. and Miller, J. D. (1978): "Speech perception by the chinchilla: identification functions for synthetic VOT stimuli". *J. Acoust. Soc. Am.* 63, 905-917.

Kuhl, P. K. and Padden, D. M. (1982): "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques". *Perc. Psych.* 32, 542-550.

Kuhl, P. K. and Padden, D. M. (1983): "Enhanced discriminability at the phonetic boundaries for the place feature in macaques". *J. Acoust. Soc. Am.* 73, 1003-1010.

Kuhne, R. and Lewis, B. (1985): "External and middle ears", in *Form and Function in Birds*" Vol. 3 (A. S. King and J. McLelland, eds), Academic Press, London and New York.

Lane, H. (1965): "Motor theory of speech perception: a critical review". *Psychol. Rev.* 72, 275-309.

Langner, G. (1981): "Neuronal mechanisms for pitch analysis in the time domain", *Exp. Brain Res.* 44, 450-454.

Langner, G., Bonke, D. and Scheich, H. (1979): "Selective responses of neurons to vowel sounds in the auditory neostriatum of the mynah bird", in *Hearing Mechanisms and Speech* (*Exp. Brain Res.* Suppl. 2) (O. Creutzfeld, H. Scheich and Chr. Schreiner, eds), Springer-Verlag, Berlin and Heidelberg.

Langner, G., Bonke, D and Scheich, H. (1981): "Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds". *Exp. Brain Res.* 43, 11-24.

Lanyon, W. E. (1960): "The ontogeny of vocalization in birds". In *Animal Sounds and Communication*, Am. Inst. Biol. Sci., Washington.

Lashley, K. S. (1913): "Reproduction of inarticulate sounds in the parrot". *J. Anim. Behav.* 3, 361-366.

Lasky, R. E., Syrdal-Lasky, A. and Klein, R. E. (1975): "VOT discrimination by four to six month and a half old infants from Spanish environments". *J. Exp. Child Psychol.* 20, 215-225.

Lemon, R. E. (1973): "Nervous control of the syrinx in white-throated sparrows (*Zonotrichia albicollis*)". *J. Zool. Lond.* 171, 131-140.

Lemon, R. E. and Scott, D. M. (1966): "On the development of song in young cardinals". *Can. J. Zool.* 44, 191-197.

Leppelsack, H.-J. (1979): "The increase of response selectivity within the avian auditory pathway". In *Hearing Mechanisms and Speech* (*Exp. Brain Res.* Sup. 2) (O. Creutzfeldt, H. Scheich and Chr. Schreiner, eds), Springer-Verlag, Berlin and Heidelberg.

Leppelsack, H. J. and Vogt, M. (1976): "Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds". *J. Comp. Physiol.* 107, 263-274.

Lewis, M. M. (1963): *Language, Thought and Personality in Infancy and Childhood*, Harrap, London.

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. and Studdert-Kennedy, M. (1967): "Perception of the speech code". *Psychol. Rev.* 74, 431-461.

Liberman, A. M., Delattre, P. C. and Cooper, F. S.

(1952): "The role of selected stimulus vriables in the perception of the unvoiced stop consonants". *Am. J. Psychol.* 65, 497-516.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1958): "Some cues for the distinction between voiced and voiceless stops in initial position". *Lang. & Speech* 1, 153-167.

Liberman, A. M., Delattre, P. C., Cooper, F. S. and Gerstman, L. J. (1954): "The role of consonant-vowel transitions in the perception of the stop and nasal consonants". *Psychol. Monogr.* 68(8), Whole No. 379.

Liberman, A. M., Delattre, P. C., Gerstman, L. J. and Cooper, F. S. (1956): "Tempo of frequency change as a cue for distinguishing classes of speech sounds". *J. Exp. Psychol.* 52, 127-137.

Liberman, A. M., Harris, K. S., Hoffman, H. S. and Griffith, B. C. (1957): "The discrimination of speech sounds within and across phoneme boundaries". *J. Exp. Psychol.* 54, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A. and Lane, H. (1961): "The discrimination of relative onset-time of the components of certain speech and nonspeech patterns". *J. Exp. Psychol.* 61, 379-388.

Lieberman, P. (1977): *Speech Physiology and Acoustic Phonetics*, Macmillan, New York.

Lill, A. (1985): "Lyrebird". In *A Dictionary of Birds* (B. Campbell and E. Lack, eds), T. and A. D. Poyser, Calton.

Lillehei, R. A. and Snowdon, C. T. (1978): "Individual and situational differences in the vocalizations of young stump-tail macaques (*Macaca arctoides*)". *Behaviour* 65, 270-281.

Lindblom, B. and Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw and larynx movement". *J. Acoust. Soc. Am.* 50, 1,166-1,179.

Lisker, L. (1975): "Is it VOT or a first-formant transition detector?". *J. Acoust. Soc. Am.* 57, 1547-

1551.

Lisker, L. and Abramson, A. S. (1964): "A cross-language study of voicing in initial stops: acoustical measurements". *Word* 20, 384-422.

Lockner, F. R. and Murrish, D. E. (1975): "Interclavicular air sac pressures and vocalization in mallard ducks (*Anas platyrhynchos*)". *Comp. Biochem. Physiol.* 52A, 183-187.

Lockner, F. R. and Youngren, O. M. (1976): "Functional syringeal anatomy of the mallard I: *in situ* electromyograms during ESB elicited calling". *Auk* 93, 324-342.

Lorenz, K. (1950): "The comparative method in studying innate behaviour patterns". *Symp. Soc. Exp. Biol.* 4, 221-268.

Lorenz, K. (1952): *King Solomon's Ring*, Methuen, London.

Mack, M. and Blumstein, S. E. (1983): "Further evidence of acoustic invariance in speech production: the stop-glide contrast". *J. Acoust. Soc. Am.* 73, 1739-1750.

McLelland, J. (1965): "The anatomy of the rings and muscles of the trachea of *G. domesticus*". *J. Anat.* 99, 651-656.

McLelland, J. (1975): "Aves digestive system". In Sisson and Grossman's *The Anatomy of the Domestic Animals* (R. Getty, ed.), W. B. Saunders Co., Philadelphia.

Maier, V. (1982): "Acoustic communication in the Guinea fowl (*Numida meleagris*): structure and use of vocalizations, and the principles of message coding". *Z. Tierpsychol.* 59, 29-83.

Manley, G. A. (1971): "Some aspects of the evolution of hearing in vertebrates". *Nature* 320, 506-509.

Manogue, K. R. and Nottebohm, F. (1982): "Relation of medullary motor nuclei to nerves supplying the vocal tract of the budgerigar (*Melopsittacus undulatus*)". *J. Comp. Neur.* 204, 384-391.

Marler, P. (1956): "The voice of the chaffinch and its function as a language". *Ibis* 98, 231-261.

Marler, P. (1959): "Developments in the study of animal communication". In *Darwin's Biological Work: some aspects reconsidered* (P. R. Bell, ed.), Cambridge University Press.

Marler, P. (1967): "Comparative study of song development in sparrows". In *Proceedings of the XIV International Ornithological Congress* (D. W. Snow, ed.), Blackwell, Oxford.

Marler, P. (1969): "Tonal quality of bird sounds". In *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Marler, P. (1970a): "Birdsong and speech development: could there be parallels?". *Am. Sci.* 58, 669-673.

Marler, P. (1970b): "A comparative approach to vocal learning: song development in white-crowned sparrows". *J. Comp. Physiol. Psychol.* 71, monograph 2(2), 1-25.

Marler, P. (1975a): "On strategies of behavioural development". In *Function and Evolution in Behaviour* (G. Baerends, C. Beer and A. Manning, eds), Clarendon Press, Oxford.

Marler, P. (1975b): "On the origin of speech from animal sounds". In *The Role of Speech in Language* (J. F. Kavanagh and J. E. Cutting, eds), M.I.T. Press.

Marler, P. (1976): "Sensory templates in species-specific behaviour". In *Simpler Networks and Behaviour* (J. C. Fentress, ed.), Sinauer Associates.

Marler, P. (1977a): "The structure of animal communication sounds". In *Dahlem Workshop on Recognition of Complex Acoustic Signals* (T. H. Bullock, ed.), Dahlem-Konfererzen, Berlin.

Marler, P. (1977b): "The evolution of communication". In *How Animals Communicate* (T. A. Sebeok, ed.), Indiana University Press.

Marler, P. and Mundinger, P. (1971): "Vocal learning in birds". In *The Ontogeny of Vertebrate Behaviour* (H. Moltz, ed.), Academic Press, New York and London.

Marler, P., Mundinger, P., Waser, M. S. and Lutjen, A.

(1972): "Effects of acoustical stimulation and deprivation on song development in red-winged blackbirds". *Anim. Behav.* 20, 586-606.

Marler, P. and Peters, S. (1977): "Selective vocal learning in the sparrow". *Science* 198, 519-521.

Marler, P. and Peters, S. (1981): "Birdsong and speech: evidence for special processing". In *Perspectives on the Study of Speech* (P. D. Eimas and J. L. Miller, eds), Lawrence Erlbaum Associates, Hillsdale, NJ.

Marler, P. and Tamura, M. (1962): "Song 'dialects' in three populations of white-crowned sparrows". *Condor* 64, 368-377.

Marler, P. and Tamura, M. (1964): "Culturally transmitted patterns of vocal behaviour in sparrows". *Science* 146, 1483-1486.

Marler, P. and Waser, M. S. (1977): "Role of auditory feedback in canary song development". *J. Comp. Physiol. Psychol.* 91, 8-16.

Marshall, A. J. (1950): "The function of vocal mimicry in birds". *Emu* 50, 5-16.

Mathes, R. C. and Miller, R. L. (1947): "Phase effects in monaural perception". *J. Acoust. Soc. Am.* 19, 780-797.

Mattingley, I. (1972): "Speech cues and sign stimuli". *Am. Scient.* 60, 327-337.

Mattingley, I., Liberman, A., Syrdal, A. and Halwes, T. (1971): "Discrimination in speech and nonspeech modes". *Cog. Psychol.* 2, 131-157.

Maynard, C. J. (1928): *Vocal organs of talking birds and some other species*, C. J. Maynard, West Newton, Mass.

Mayr, E. (1958): "The sequence of the songbird families". *Condor* 60, 194-195.

Medway, Lord and Pye, J. D. (1977): "Echolocation and the systematics of swiftlets". In *Evolutionary Ecology* (B. Stonehouse and C. Perrins, eds), Macmillan, London.

Menyuk, P. and Menn, L. (1979): "Early strategies for the perception and production of words and sounds". In *Language Acquisition* (P. Fletcher and M. Garman, eds),

Cambridge University Press.

Messmer, E. and Messmer, I. (1956): "Die Entwicklung der lautäusserungen und einiger Verhaltenweisen der Amsel (*Turdus merula merula* L.) unter natürlichen Bedingungen und nach Einzelaufzucht in schalldichten Räumen". *Z. Tierpsychol.* 13, 341-441.

Meyer-Eppler, W. (1957): "Realization of prosodic features in whispered speech". *J. Acoust. Soc. Am.* 29, 104-106.

Miller, A. H. (1934): "The vocal apparatus of some North American owls". *Condor* 36, 204-213.

Miller, D. B. (1977): "Two-voice phenomenon in birds: further evidence". *Auk* 94, 567-572.

Miller, J. D. (1970): "Audibility curve of the chinchilla". *J. Acoust. Soc. Am.* 48, 513-523.

Miller, J. D. (1977): "Perception of speech sounds in animals: evidence for speech processing by mammalian auditory mechanisms". In *Recognition of Complex Acoustic Signals* (T. H. Bullock, ed.), Dahlem Konferenzen, Berlin.

Miller, J. D. and Kuhl, P. K. (1976): "Speech perception by the chinchilla: a progress report on syllable-initial voiced-plosive consonants". *J. Acoust. Soc. Am.* 59, S554(A).

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J. and Dooling, R. (1976): Discrimination and labelling of noise-buzz sequences with varying noise-lead times: an example of categorical perception". *J. Acoust. Soc. Am.* 60, 410-422.

Miller, J. M., Sutton, D., Pfingst, B., Ryan, A. and Beaton, R. (1972): "Single cell activity in the auditory cortex of rhesus monkeys: behavioural dependency". *Science* 177, 449-451.

Milner, B., Taylor, L. and Sperry, R. W. (1968): "Lateralized suppression of dichotically presented digits after commissural section in man". *Science* 161, 184-185.

Miskimen, M. (1951): "Sound production in passerine birds". *Auk* 68, 493-504.

Mivart, St G. (1895): "On the hyoid bone of certain parrots". *Proc. Zool. Soc. Lond.*, 162-174.

Mol, H. (1971): "Some notes on talking birds". In *Form and Substance: phonetic and linguistic papers presented to Eli Fischer-Jorgensen* (L. L. Hammerich, R. Jakobson and E. Zwirner, eds), Akademisk Forlag.

Mol, H. (1972): "Acoustic theory of the parrot". *Proc. 7th Int. Congr. Phonet. Sci.*, Mouton.

Molfese, D. L. (1978a): "Left and right hemisphere involvement in speech perception: electrophysiological correltes". *Perc. Psych.* 23, 237-243.

Molfese, D. L. (1978b): "Neuroelectrical correlates of categorical speech perception in adults". *Brain Lang.* 5, 25-35.

Molfese, D. L. (1980a): "Hemispheric specialization for temporal information: implications for the perception of voicing cues during speech perception". *Brain Lang.* 11, 285-299.

Molfese, D. L. (1980b): "The phoneme and the engram: electrophysiological evidence for the acoustic invariant in stop consonants". *Brain Lang.* 9, 372-376.

Molfese, D. L. (1984): "Left hemisphere sensitivity to consonant sounds not displayed by the right hemisphere: electrophysiological correlates". *Brain Lang.* 26, 49-62.

Molfese, D. L., Buhrke, R. A. and Wang, S. L. (1985): "The right hemisphere and temporal processing of consonant transition durations: electrophysiological correlates". *Brain Lang.* 26, 49-62.

Molfese, D. L. and Erwin, R. J. (1981): "Intrahemispheric differentiation of vowels: principal component analysis of auditory evoked responses to computer-synthesized vowel sounds". *Brain Lang.* 13, 333-344.

Molfese, D. L. and Schmidt, A. (1983): "An auditory evoked potential study of consonant perception in

different vowel environments". *Brain Lang.* 18, 57-70.

Molfese, V. J., Molfese, D. L. and Parsons, C. (1983): "Hemispheric processing of phonological information". In *Language Functions and Brain Organization* (S. J. Segalowitz, ed.), Academic Press, New York and London.

Morgan, C. L. (1891): *Animal Life and Intelligence*, Edward Arnold, London.

Morse, P. A., Molfese, D., Laughlin, N. K., Linnville, S. and Wetzel, F. (1987): "Categorical perception for voicing contrasts in normal and lead-treated rhesus monkeys: electrophysiological indices". *Brain Lang.* 30, 63-80.

Morse, P. A. and Snowdon, C. T. (1975): "An investigation of categorical speech discrimination by rhesus monkeys". *Perc. & Psych.* 17, 9-16.

Morse, P. M. and Ingard, K. U. (1968): *Theoretical Acoustics*, McGraw-Hill, New York.

Morton, E. S. (1975): "Ecological sources of selection on avian sounds". *Am. Nat.* 109, 17-34.

Morton, E. S. (1977): "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds". *Am. Nat.* 111, 855-869.

Morton, E. S. (1982): "Grading, discreteness, redundancy, and motivation-structural rules". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds) Vol. 1, Academic Press, New York and London.

Mowrer, O. H. (1950): "On the psychology of 'talking birds'". In *Learning Theory and Personality Dynamics* (O. H. Mowrer, ed.), Ronald Press, New York.

Mowrer, O. H. (1952): "The autism theory of speech development and some clinical applications". *J. Sp. Hear. Disord.* 17, 263-268.

Mowrer, O. H. (1954): "The psychologist looks at language". *Am. Psychol.* 9, 660-694.

Mowrer, O. H. (1958): "Hearing and speaking: an analysis of language learning". *J. Sp. Hear. Disord.* 23, 143-

152.

Mowrer, O. H. (1980): *Psychology of Language and Learning*, Plenum Press, New York.

Mudge, G. P. (1902): "On the myology of the tongue of parrots, with a classification of the order based upon the structure of the tongue". *Trans. Zool. Soc. Lond.* 16, 211-278.

Müller, J. (1847): "Über die bisher unbekannten Typischen Verschiedenheiten der Stimmorgane der Passerinen". *Abh. Konnig. Akad. Wiss.*, Berlin.

Müller, J. (1878): *On certain variations in the Vocal Organs of the Passeres that have hitherto escaped notice* (F. J. Bell, trans.), Clarendon Press, Oxford.

Mulligan, J. A. (1966): "Singing behaviour and its development in the song sparrow, *Melospiza melodia*". *Univ. Calif. Publ. Zool.* 81, 1-76.

Mulligan, J. A. and Olsen, K. C. (1969): "Communication in canary courtship calls". In *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Mundinger, P. C. (1970): "Vocal imitation and individual recognition of finch calls". *Science* 168, 480-482.

Mundinger, P. C. (1982): "Microgeographic and macrogeographic variation in the acquired vocalizations of birds". In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds) Vol. 2, Academic Press, New York and London.

Myers, J. A. (1917): "Studies on the syrinx of *Gallus domesticus*". *J. Morph.* 29, 165-215.

Nabelek, I. and Hirsh, I. J. (1969): "On the discrimination of frequency transitions". *J. Acoust. Soc. Am.* 45, 1,510-1,519.

Narins, P. M. and Gleich, O. (1986): "Phase response of low-frequency cochlear ganglion cells in the starling". In *Auditory Frequency Selectivity* (B. C. J. Moore and R. D. Patterson, eds), Plenum Press, New York and London.

Newman, J. D. *et al.* (1977): "Biological filtering and

neural mechanisms". In *Recognition of Complex Acoustic Signals* (T. H. Bullock, ed.), Dahlem Konferenzen, Berlin.

Newton, A., ed. (1893-1896): *A Dictionary of Birds*, Adam and Charles Black, London.

Nicolai, J. (1959): "Familientradition in der Gesangsentwicklung des Gimpels (*Pyrrhula pyrrhula* L.)". *J. Ornith.* 100, 39-46.

Nicolai, J. (1964): "Der Brutparasitismus der Viduinae als ethologisches Problem". *Z. Tierpsychol.* 21, 129-204.

Nitzsch, C. L. (1829): *Observationes de Avium Arteria carotide communi*, Halle.

Noscoe, N. J., Fourcin, A. J., Brown, N. J. and Berry, R. J. (1983): "Examination of vocal fold movement by ultra-short pulse X radiography". *Br. J. Radiol.* 56, 641-645.

Nottebohm, F. (1967): "The role of sensory feedback in the development of avian vocalizations". In *Proc. XIVth Int. Ornith. Congr.* (D. W. Snow, ed.), Blackwell, Oxford.

Nottebohm, F. (1968): "Auditory experience and song development in the chaffinch (*Fringilla coelebs*)". *Ibis* 110, 549-568.

Nottebohm, F. (1969a): "The 'critical period' for song learning in birds". *Ibis* 3, 386-387.

Nottebohm, F. (1969b): "The song of the chingolo, *Zonotrichia capensis*, in Argentina: description and evaluation of a system of dialects". *Condor* 71, 299-315.

Nottebohm, F. (1970): "Ontogeny of bird song". *Science* 167, 950-956.

Nottebohm, F. (1971): "Neural lateralization of vocal control in a passerine bird I: song". *J. Exp. Zool.* 177, 229-262.

Nottebohm, F. (1972a): "Neural lateralization of vocal control in a passerine bird II: subsong, calls, and a

theory of vocal learning". *J. Exp. Zool.* 179, 35-50.

Nottebohm, F. (1972b): "The origins of vocal learning". *Am. Nat.* 106, 116-140.

Nottebohm, F. (1975): "Vocal behaviour in birds". In *Avian Biology* (D. S. Farner and J. R. King, eds), Academic Press, New York and London.

Nottebohm, F. (1976): "Phonation in the orange-winged Amazon parrot (*Amazona amazonica*)". *J. Comp. Physiol.* 108, 157-170.

Nottebohm, F. (1977) "Asymmetries in neural control of vocalization in the canary". In *Lateralization in the Nervous System* (S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes and G. Krauthamer, eds), Academic Press, New York and London.

Nottebohm, F. (1980a): "Testosterone triggers growth of brain vocal control nuclei in adult female canaries". *Brain Res.* 189, 429-436.

Nottebohm, F. (1980b): "Brain pathways for vocal learning in birds: a review of the first ten years". *Prog. Psychobiol. Physiol. Psychol.* 9, 85-124.

Nottebohm, F. (1981): "A brain for all seasons: cyclical anatomical changes in song control nuclei of the canary brain". *Science* 214, 1368-1370.

Nottebohm, F. (1984): "Birdsong as a model in which to study brain processes related to learning". *Condor* 86, 227-236.

Nottebohm, F., Kasparian, S. and Pandazis, C. (1981): "Brain space for a learnt task". *Brain Res.* 213, 99-109.

Nottebohm, F. and Nottebohm, M. (1971): "Vocalizations and breeding behaviour of surgically deafened ring doves (*Streptopelia risoria*)". *Anim. Behav.* 19, 313-327.

Nottebohm, F. and Nottebohm, M. (1976): "Left hypoglossal dominance in the control of canary and white-crowned sparrow song". *J. Comp. Physiol.* 108, 171-192.

Nottebohm, F. and Nottebohm, M. (1978): "Relationship

between song repertoire and age in the canary, *Serinus canarius*". *Z. Tierpsychol.* 46, 298-305.

Nottebohm, F., Nottebohm, M. E. and Crane, L. (1986): "Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei". *Behav. Neur. Biol.* 46, 445-471.

Nottebohm, F. and Selander, R. K. (1972): "Vocal dialects and gene frequencies in the chingolo sparrow (*Zonotrichia capensis*)". *Condor* 74, 137-143.

Nottebohm, F., Stokes, T. M. and Leonard, C. M. (1976): "Central control of song in the canary (*Serinus canarius*)". *J. Comp. Neur.* 165, 457-486.

Nowicki, S. (1987): "Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere". *Nature* 325, 53-55.

Olson, S. L. (1985): "The fossil record of birds". In *Avian Biology* Vol. 8 (D. S. Farner, J. R. King and K. C. Parkes, eds), Academic Press, New York and London.

Oppenheimer, J. R. (1977): "Communication in New World monkeys". In *How Animals Communicate* (T. A. Sebeok, ed.), Indiana University Press.

Owen, R. (1866): *On the Anatomy of Vertebrates* Vol. 2, Longmans Green, London.

Papcun, G., Krashen, S., Terbeek, D., Remington, R and Harshman, R. (1974): "Is the left hemisphere specialized for speech, language and/or something else?". *J. Acoust. Soc. Am.* 55, 319-327.

Parks, T. N. and Rubel, E. W. (1975): "Organization and development of brain stem auditory nuclei of the chicken: organization of projections from *n. magnocellularis* to *n. laminaris*", *J. Comp. Neur.* 164, 435-448.

Paulsen, K. (1967): *Das Prinzip der Stimmbildung in der Wirbeltierreihe und beim Menschen*, Akademische Verlagsgesellschaft, Frankfurt am Main.

Payne, R. B. (1981): "Song learning and social interaction in indigo buntings". *Anim. Behav.* 29, 688-

697.

Payne, R. B. (1982): "Ecological consequences of song matching: breeding success and intraspecific song mimicry in indigo buntings". *Ecology* 63, 401-411.

Payne, R. B. (1983): "The social context of song mimicry: song-matching dialects in indigo buntings (*Passerina cyanea*)". *Anim. Behav.* 31, 788-805.

Payne, R. B., Payne, L. L. and Doehlert, S. M. (1984): "Interspecific song learning in a wild chestnut-sided warbler". *Wilson Bull.* 96, 292-294.

Peek, F. W. and Phillips, R. E. (1971): "Repetitive vocalizations evoked by local electrical stimulation of avian brains: II anaesthetized chickens (*Gallus gallus*)". *Brain, Behav. Evol.* 4, 417-438.

Pepperberg, I. M. (1981): "Functional vocalizations by an African grey parrot". *Z. Tierpsychol.* 55, 139-160.

Pepperberg, I. M. (1983): "Cognition in the African grey parrot: preliminary evidence for auditory/vocal comprehension of the class concept". *Anim. Learn. Behav.* 11, 179-185.

Pepperberg, I. M. (1985a): "Social modelling theory: a possible framework for understanding avian vocal learning". *Auk* 102, 854-864.

Pepperberg, I. M. (1985b): "Acquisition of anomalous communicatory systems: implications for studies on inter-species communication". In *Dolphin Behaviour and Cognition: comparative and ecological aspects* (R. Buhr, R. Schusterman, J. Thomas and F. Wood, eds), Erlbaum Assoc., New York.

Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B. and Stebbins, W. C. (1978): "Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*)". *Science* 202, 324-327.

Peterson, G. E. and Barney, H. L. (1952): "Control methods used in the study of the vowels". *J. Acoust. Soc. Am.* 24, 175-184.

Phillips, R. E. and Peek, F. W. (1975): "Brain

organization and neuromuscular control of vocalization in birds". In *Neural and Endocrine Aspects of Behaviour in Birds* (P. Wright, P. G. Caryl and D. M. Vowles, eds), Elsevier, Amsterdam.

Phillips, R. E. and Youngren, O. M. (1981): "Effects of denervation of the tracheo-syringeal muscles on frequency control in vocalization in chicks". *Auk* 98, 299-306.

Phillips, R. E., Youngren, O. M. and Peek, F. W. (1972): "Repetitive vocalizations evoked by local electrical stimulation of avian brains I: awake chickens (*Gallus gallus*)". *Anim. Behav.* 20, 689-705.

Pisoni, D. B. (1977): "Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops". *J. Acoust. Soc. Am.* 61, 1352-1361.

Pisoni, D. B. (1979): "On the perception of speech sounds as biologically significant signals". *Brain Behav. Evol.* 16, 330-350.

Plomp, R. (1967): "Pitch of complex tones". *J. Acoust. Soc. Am.* 41, 1,526-1,533.

Plomp, R. (1976): *Aspects of Tone Sensation*, Academic Press, New York and London.

Plomp, R. (1983): "The role of modulation in hearing". In *Hearing - Physiological Bases and Psychophysics* (R. Klinke and R. Hartmann, eds), Springer-Verlag, Berlin.

Pollack, I. (1968): "Detection of rate of change of auditory frequency". *J. Exp. Psychol.* 77, 535-541.

Portmann, A. (1950): "Les organes respiratoires". In *Traité de Zoologie* (P.-P. Grasse, ed.) Vol. 15: "Oiseaux", Masson, Paris.

Potash, L. M. (1970): "Vocalizations elicited by electrical brain stimulation in *Coturnix coturnix japonica*". *Behaviour* 36, 149-167.

Potter, R. K., Kopp, G. A. and Green, H. C. (1947): *Visible Speech*, D. van Nostrand, New York.

Potter, R. K. and Steinberg, J. C. (1950): "Toward the

specification of speech". *J. Acoust. Soc. Am.* 22, 807-820.

Poulsen, H. (1951): "Inheritance and learning in the song of the chaffinch". *Behaviour* 3, 216-228.

Price, P. H. (1979): "Developmental determinants of structure in zebra finch song". *J. Comp. Physiol. Psychol.* 93, 260-277.

Pumphrey, R. J. (1961): "Hearing", in *Biology and Compartive Physiology of Birds* Vol. 2 (A. J. Marshall, ed.), Academic Press, New York and London.

Pye, J. D. (1980): "Echolocation signals and echoes in air". In *Animal Sonar Systems* (R.-G. Busnel and J. F. Fish, eds), Plenum Press, New York and London.

Rausch, G. and Scheich, H. (1982): "Dendritic spine loss and enlargement during maturation of the speech control system in the mynah bird (*Gracula religiosa*)". *Neurosci. L.* 29, 129-133.

Rayleigh, Lord (1878): *The Theory of Sound* Vol. 2, Macmillan, London.

Remez, R. E. and Rubin, P. E. (1984): "On the perception of intonation from sinusoidal sentences". *Perc. & Psych.* 35, 429-440.

Remez, R. E., Rubin, P. E., Pisoni, D. B. and Carrell, T. D. (1981): "Speech perception without traditional speech cues". *Science* 212, 947-950.

Rice, J. O. and Thompson, W. L. (1968): "Song development in the indigo bunting". *Anim. Behav.* 16, 462-469.

Ritsma, R. J. (1962): "Existence region of the tonal residue". *J. Acoust. Soc. Am.* 34, 1,224-1,229.

Ritsma, R. J. (1967): "Frequencies dominant in the perception of the pitch of complex sounds". *J. Acoust. Soc. Am.* 42, 191-198.

Robinson, F. N. (1975): "Vocal mimicry and the evolution of bird song". *Emu* 75, 23-27.

Rogers, L. J. and Anson, J. M. (1979): "Lateralization of function in the chicken forebrain". *Pharmacol. Biochem. Behav.* 10, 679-686.

Romanes, G. J. (1886): *Animal Intelligence*, Kegan Paul, Trench and Co., London.

Rosen, S. (1986): "Monaural phase sensitivity: frequency selectivity and temporal processes". In *Auditory Frequency Selectivity* (B. C. J. Moore and R. D. Patterson, eds), Plenum Press, New York.

Rosen, S. and Fourcin, A. (1986): "Frequency selectivity and the perception of speech". In *Frequency Selectivity in Hearing* (B. C. J. Moore, ed.), Academic Press, London.

Rosen, S. and Howell, P. (1987): "Auditory, articulatory and learning explanations of categorical perception in speech". In *Categorical Perception* (S. Harnad, ed.), Cambridge University Press.

Rosenzweig, M. R. (1951): "Representations of the two ears at the auditory cortex". *Am. J. Physiol.* 167, 147-158.

Rossing, T. D. (1982a): *The Science of Sound*, Addison-Wesley, Reading, MA.

Rossing, T. D. (1982b): "The physics of kettledrums". *Sci. Am.* 247 (5), 147-152.

Rowell, T. E. and Hinde, R. A. (1962): "Vocal communication by the rhesus monkey (*Macaca mulatta*)". *Proc. Zool. Soc. Lond.* 138, 279-294.

Rubel, E. W. and Parks, T. N. (1975): "Organization and development of the brain stem auditory nuclei of the chicken: tonotopic organization of *nucleus magnocellularis* and *nucleus laminaris*", *J. Comp. Neur.* 164, 411-434.

Rubel, E. W. and Ryals, B. M. (1983): "Development of the place principle: acoustic trauma". *Science* 219, 512-514.

Rüppell, W. (1933): "Physiologie und Akustik der Vögelstimme". *J. Orn. Berl.* 81, 433-542.

Russ, K. (1884): *The Speaking Parrots*, L. Upcott Gill, London.

Ryals, B. M. and Rubel, E. W. (1985): "Ontogenetic

changes in the position of hair cell loss after acoustic overstimulation in avian basilar papilla". *Hearing Res.* 19, 135-142.

Sachs, M., Woolf, N. K. and Sinnott, J. M. (1980): "Response properties of neurons in the avian auditory system: comparisons with mammalian homologues and consideration of the neural encoding of complex stimuli". In *Comprative Studies of Hearing in Vertebrates* (A. N. Popper and R. R. Fay, eds), Springer-Verlag, New York.

Sachs, M. B., Young, E. D. and Lewis, R. H. (1974): "Discharge patterns of single fibres in the pigeon auditory nerve". *Brain Res.* 70, 431-447.

Saito, N. (1980): "Structure and function of the avian ear". In *Comparative Studies of Hearing in Vertebrates* (A. N. Popper and R. R. Fay, eds), Springer-Verlag, New York.

Saito, N. and Maekawa, M. (1981): "Integration of voco-auditory centres in song birds", in *Neuronal Mechanisms in Hearing* (J. Syka and L. Aitkin, eds), Plenum Press, New York.

Saito, S., Fukuda, H., Isogai, Y. and Ono, H. (1981): "X-ray stroboscopy". In *Vocal Fold Physiology* (K. N. Stevens and M. Hirano, eds), Tokyo University Press.

Saunders, J. C., Bock, G. R. and Fahrbach, S. E. (1978): "Frequency selectivity in the parakeet (*Melopsittacus undulatus*) studied with narrow-band noise masking". *Sensory Processes* 2, 80-89.

Saunders, J. C., Denny, R. M. and Bock, G. R. (1978): "Critical bands in the parakeet (*Melopsittacus undulatus*)". *J. Comp. Physiol.* 125, 359-365.

Saunders, J. C. and Dooling, R. J. (1974): "The relations between vocalizations and hearing capability in the parakeet (*Melopsittacus undulatus*)". *J. Acoust. Soc. Am.* 55, 480-481.

Saunders, J. C., Else, D. V. and Bock, G. R. (1978): "Frequency selectivity in the parakeet (*Melopsittacus*

undulatus) studied with psychophysical tuning curves", *J. Comp. Physiol. Psych.* 92, 406-415.

Saunders, J. C., Rintelmann, W. F. and Bock, G. R. (1979): "Frequency selectivity in bird and man: a comparison among critical ratios, critical bands and psychophysical tuning curves", *Hearing Res.* 1, 303-323.

Savart, F. (1826): "Mémoire sur la voix des oiseaux". *Ann. de Chim. et de Physique* 32, 5-24.

Scanlan, J. P. (1983): "Talking birds: an analysis of non-human speech". *Abstracts of the 18th International Ethological Conference*, Brisbane.

Scharf, B. (1970): "Critical bands". In *Foundations of Modern Auditory Theory* (J. V. Tobias, ed.), Academic Press, New York.

Scheich, H. (1977): "Central processing of complex sounds and feature analysis". In *Recognition of Complex Acoustic Signals* (T. H. Bullock, ed.), Dahlem Konferenzen, Berlin.

Scheich, H., Bock, W., Bonke, D., Langner, G. and Maier, V. (1983): "Acoustic Communication in the Guinea fowl (*Numida meleagris*)". In *Advances in Vertebrate Neuroethology* (J.-P. Ewert, R. R. Capranica and D. J. Ingle, eds), Plenum, New York.

Scheich, H., Bonke, D. and Langner, G. (1979): "Tonotopy and analysis of wide-band calls in field L of the Guinea fowl". In *Exp. Brain Res. Supp. 2: Hearing Mechanisms and Speech* (O. Creutzfeldt, H. Scheich and Chr. Schreiner, eds), Springer-Verlag, Berlin and Heidelberg.

Scheich, H., Langner, G. and 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". *J. Comp. Physiol.* 132, 257-276.

Scheich, H., Langner, G. and Koch, R. (1977): "Coding of

narrow-band and wide-band vocalizations in the auditory midbrain nucleus (MLD) of the Guinea fowl (*Numida meleagris*)". *J. Comp. Physiol.* 117, 245-265.

Schouten, J. F., Ritsma, R. J. and Cardozo, B. L. (1962): "Pitch of the residue". *J. Acoust. Soc. Am.* 34, 1418-1424.

Schubert, G. (1971): "Experimentelle Untersuchungen über die artkennzeichnenden Parameter im Gesang des Zilpzalps, *Phylloscopus c. collybita* (Vieillot)". *Behaviour* 38, 289-314.

Schwartzkopff, J. (1955): "On the hearing of birds". *Auk* 72, 340-347.

Schwartzkopff, J. (1973): "Mechanoreception", in *Avian Biology* Vol. 3 (D. S. Farner and J. R. King, eds), Academic Press, New York and London.

Sedlacek, K. (1971): "Hearing and communication in birds". *Acta Otolaryng.* 71, 194-196.

Seller, T. J. (1979): "Unilateral nervous control of the syrinx in Java sparrows (*Padda oryzivora*)". *J. Comp. Physiol.* 129, 281-288.

Serpell, J. A. (1985): "Parrots, lories and cockatoos". In *The Encyclopaedia of Birds* (C. M. Perrins and A. L. A. Middleton, eds), Allen and Unwin, London.

Shankweiler, D., Strange, W. and Verbrugge, R. (1977): "Speech and the problem of perceptual constancy". In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* (R. E. Shaw and J. Bransford, eds), Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Shankweiler, D. and Studdert-Kennedy, M. (1967a): "Identification of consonants and vowels presented to left and right ears". *Q. J. Exp. Psychol.* 19, 59-63.

Shankweiler, D. and Studdert-Kennedy, M. (1967b): "An analysis of perceptual confusions in identification of dichotically presented CVC syllables". *J. Acoust. Soc. Am.* 41, 1581.

Shower, E. G. and Biddulph, R. (1931): "Differential pitch sensitivity of the ear". *J. Acoust. Soc. Am.* 2,

275-287.

Singh, R. P. (1980): *Anatomy of Hearing and Speech*, Oxford U. P., New York.

Sinnott, J. M. (1980): "Species-specific coding in bird song". *J. Acoust. Soc. Am.* 68, 494-497.

Sinnott, J. M., Beecher, M. D., Moody, D. B. and Stebbins, W. C. (1976): "Speech sound discrimination by monkeys and humans". *J. Acoust. Soc. Am.* 60, 687-695.

Sinnott, J. M., Sachs, M. B. and Hienz, R. D. (1976): "Differential sensitivity to frequency and intensity in songbirds". *J. Acoust. Soc. Am.* 60, S87.

Sinnott, J. M., Sachs, M. B. and Hienz, R. D. (1980): "Aspects of frequency discrimination in passerine birds and pigeons". *J. Comp. Physiol. Psychol.* 94, 401-415.

Skinner, B. F. (1957): *Verbal Behaviour*, Appleton-Century-Crofts, New York.

Slater, P. J. B. (1983): "Chaffinch imitates canary song elements and aspects of organization". *Auk* 100, 493-495.

Smith, C. A. (1985): "Inner ear", in *Form and Function in Birds* Vol. 3 (A. S. King and J. McLelland, eds), Academic Press, London and New York.

Smith, D. G. (1977): "The role of the sternotrachealis muscles in bird song production". *Auk* 94, 152-155.

Smith, N. and Wilson, D. (1979): *Modern Linguistics*, Penguin Books.

Smith, S. (1947): "Analysis of vowel sounds by ear". *Arch. Neerl. Phonet. Exp.* 20, 78-96.

Smith, S. (1954): "Remarks on the physiology of the vibrations of the vocal cords". *Folia Phoniat.* 6, 166-178.

Smolders, J. W. Th., Gummer, A. W. and Klinke, R. (1986): "Basilar membrane tuning in the pigeon ear". In *Auditory Frequency Selectivity* (B. C. J. Moore and R. D. Patterson, eds), Plenum Press, New York and London.

Snow, D. W. (1968): "The singing assemblies of little hermits". *Living Bird* 7, 47-55.

Snow, D. W. (1977): "Duetting and other synchronized displays of the blue-backed manakins (*Chiroxiphia* spp.). In *Evolutionary Ecology* (B. Stonehouse and C. Perrins, eds), Macmillan, London.

Snowdon, C. T. (1979): "Response of nonhuman animals to speech and to species-specific sounds". *Brain Behav. Evol.* 16, 409-429.

Snowdon, C. T. (1987): "A naturalistic view of categorical perception". In *Categorical Perception* (S. Harnad, ed.), Cambridge University Press.

Snowdon, C. T. and Pola, Y. V. (1978): "Interspecific and intraspecific responses to synthesized pygmy marmoset vocalizations". *Anim. Behav.* 26, 192-206.

Soli, S. D. (1983): "The role of spectral cues in discrimination of voice onset time differences". *J. Acoust. Soc. Am.* 73, 2150-2165.

Sonesson, B. (1970): "The functional anatomy of the speech organs". In *Manual of Phonetics* (B. Malmberg, ed.), North Holland, Amsterdam.

Sparling, D. W. (1979): "Evidence for vocal learning in prairie grouse". *Wilson Bull.* 91, 618-621.

Springer, S. P. (1971): "Ear asymmetry in a dichotic detection task". *Perc. Psych.* 10, 239-241.

Springer, S. P. (1973): "Hemispheric specialization for speech opposed by contralateral noise". *Perc. Psych.* 13, 391-393.

Stein, L. (1949): *The Infancy of Speech and the Speech of Infancy*, Methuen, London.

Stein, R. C. (1968): "Modulation in bird sounds". *Auk* 85, 229-243.

Stevens, K. N. (1972): "Quantal nature of speech". In *Human Communication: A Unified View* (E. E. David and P. B. Denes, eds), McGraw-Hill, New York.

Stevens, K. N. (1977): "Physics of laryngeal behaviour and larynx modes". *Phonetica* 34, 264-279.

Stevens, K. N. (1980): "Acoustic correlates of some phonetic categories". *J. Acoust. Soc. Am.* 68, 836-842.

Stevens, K. N. (1981): "Constraints imposed by the auditory system on the properties used to classify spech sounds: data from phonology, acoustics, and psychoacoustics". In *The Cognitive Representation of Speech* (T. Myers, J. Laver and J. Anderson, eds), North Holland, Amsterdam.

Stevens, K. N. and House, A. S. (1955): "Development of a quantitative description of vowel articulation". *J. Acoust. Soc. Am.* 27, 484-493.

Stevens, K. N. and Klatt, D. H. (1974): "Role of formant transitions in the voiced-voiceless distinction for stops". *J. Acoust. Soc. Am.* 55, 653-659.

Storer, R. W. (1960): "The classification of birds". In *Biology and Comparative Physiology of Birds* (A. J. Marshall, ed.) Vol. 1, Academic Press, New York and London.

Storer, R. W. (1971): "Classification of birds". In *Avian Biology* (D. S. Farner and J. R. King, eds) Vol. 1, Academic Press, New York and London.

Streeter, L. A. (1976): "Kikuyu labial and apical stop discrimination". *J. Phonet.* 4, 43-49.

Stresemann, E. (1937): "Syrinx". In *Handbuch der vergleichenden Anatomie der Wirbeltiere* (L. Bolk, E. Göppert, E. Kallius and W. Lubosch, eds), Urban und Schwarzenberg, Berlin and Vienna.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S. and Cooper, F. S. (1970): "Motor theory of speech perception: a reply to Lane's critical review". *Psychol. Rev.* 77, 234-249.

Studdert-Kennedy, M. and Shankweiler, D. P. (1970): "Hemispheric specialization for speech perception". *J. Acoust. Soc. Am.* 48, 579-594.

Studdert-Kennedy, M., Shankweiler, D. and Pisoni, D. (1972): "Auditory and phonetic proceses in speech perception: evidence from a dichotic study". *Cognitive*

Psychol. 3, 455-466.

Summerfield, Q. and Haggard, M. (1977): "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants". *J. Acoust. Soc. Am.* 62, 435-448.

Sutherland, C. A. and McChesney, D. S. (1965): "Sound production in two species of geese". *Living Bird*, 99-106.

Suthers, R. A. and Hector, D. H. (1982): "Mechanism for the production of echolocating clicks by the grey swiftlet, *Collocalia spodiopygia*". *J. Comp. Physiol.* 148, 457-470.

Takasaka, T. and Smith, C. A. (1971): "The structure and innervation of the pigeon's basilar papilla". *J. Ultra. Res.* 35, 20-65.

Tanaka, K. and Smith, C. A. (1978): "Structure of the chicken's inner ear: SEM and TEM study", *Am. J. Anat.* 153, 251-271.

Tenaza, R. R. (1976): "Wild mynahs mimic wild primates". *Nature* 259, 561.

Thielcke, G. A. (1966): "Ritualized distinctiveness of song in closely related sympatric species". *Philosoph. Trans. Royal Soc. Lond.* (Series B) 251, 493-497.

Thielcke, G. A. (1969): "Geographic variation in bird vocalizations". In *Bird Vocalizations* (R. A. Hinde, ed.), Cambridge University Press.

Thielcke, G. A. (1970): "Lernen von Gesang als möglicher Schrittmacher der Evolution". *Z. Sys. Evolutionsforsch* 8, 309-320.

Thielcke, G. A. (1976): *Bird Sounds*, Ann Arbor, Michigan.

Thielcke-Polz, H. and Thielcke, G. (1960): "Akustisches Lernen verschieden alter schallisolierter Amseln (*Turdus merula* L.) und die Entwicklung erlernter Motive ohne und mit künstlichem Einfluss von Testosteron". *Z. Tierpsychol.* 17, 211-244.

Thorndike, E. L. (1898): "Animal intelligence". *Psychol. Rev.* Monogr. Supp. 8.

Thorpe, W. H. (1951): "The learning abilities of birds". *Ibis* 93, 252-296.

Thorpe, W. H. (1954): "The process of song-learning in the chaffinch as studied by means of the sound spectrograph". *Nature* 173, 465-469.

Thorpe, W. H. (1955): "Comments on *The Bird Fancyer's Delight*, together with notes on imitation in the sub-song of the chaffinch". *Ibis* 97, 247-251.

Thorpe, W. H. (1958a): "The learning of song patterns by birds, with especial reference to the song of the chaffinch". *Ibis* 100, 535-570.

Thorpe, W. H. (1958b): Further studies on the process of song learning in the chaffinch (*Fringilla coelebs gengleri*)". *Nature* 182, 554-557.

Thorpe, W. H. (1959): "Talking birds and the mode of action of the vocal apparatus of birds". *Proc. Zool. Soc. Lond.* 132, 441-455.

Thorpe, W. H. (1961): *Bird-song*, Cambridge University Press.

Thorpe, W. H. (1963): "Antiphonal singing in birds as evidence for avian auditory reaction time". *Nature* 197, 774-776.

Thorpe, W. H. (1967): "Vocal imitation and antiphonal song and its implications". In *Proc. XIVth Int. Ornith. Congr.* (D. W. Snow, ed.), Blackwell, Oxford.

Thorpe, W. H. (1968): "Perceptual basis for group organization in social vertebrates, especially birds". *Nature* 220, 124-128.

Thorpe, W. H. (1969): "The significance of vocal imitation in animals, with special reference to birds". *Acta Biol. Exp.* 29, 251-269.

Thorpe, W. H. (1972): *Duetting and Antiphonal Song in Birds*, *Behaviour* Sup. 18.

Thorpe, W. H. and North, M. E. W. (1965): "Origin and significance of the power of vocal imitation, with special reference to the antiphonal singing of birds". *Nature* 208, 219-222.

Thorpe, W. H. and North, M. E. W. (1966): "Vocal imitation in the tropical bou-bou shrike (*Laniarius aethiopicus major*) as a means of establishing and maintaining social bonds". *Ibis* 108, 432-435.

Thorpe, W. H. and Pilcher, P. M. (1958): "The nature and characteristics of sub-song". *Br. Birds* 51, 509-514.

Thurlow, W. R. (1963): "Perception of low auditory pitch: a multicue, mediation theory". *Psychol. Rev.* 70, 461-470.

Todt, D. (1975a): "Social learning of vocal patterns and modes of their application in grey parrots". *Z. Tierpsychol.* 39, 178-188.

Todt, D. (1975b): "Spontaneous recombinations of vocal patterns in parrots". *Naturwissen.* 62, 399-400.

Tretzel, E. (1965): "Imitation und Variation von Schäferpfiffen durch Haubenlerchen (*Galerida c. cristata* L.): ein Beispiel für spezielle Spottmotiv-Prädisposition". *Z. Tierpsychol.* 22, 784-809.

Tretzel, E. (1967): "Imitation und Transposition menschlicher Pfiffe durch Amseln (*Turdus m. merula* L.): ein weiterer Nachweis relativen Lernens und akustischer Abstraktion bei Vögeln". *Z. Tierpsychol.* 24, 137-161.

Tsunoda, T. (1969): "Contralateral shift of cerebral dominance for nonverbal sounds during speech perception". *J. Auditory Res.* 9, 221-229.

Turney, T. H. (1982): "The association of visual concepts and imitative vocalizations in the mynah (*Gracula religiosa*)". *Bull. Psychon. Soc.* 19, 59-62.

Van den Berg, J. (1958): "Myoelastic-aerodynamic theory of voice production". *J. Sp. Hear. Res.* 1, 227-244.

Van den Berg, J. (1970): "Mechanism of the larynx and the laryngeal vibrations". In *Manual of Phonetics* (B. Malmberg, ed.), North-Holland, Amsterdam.

Van den Berg, J., Zantema, J. T. and Doornaenbal, P. (1957): "On the air resistance and the Bernoulli effect of the human larynx". *J. Acoust. Soc. Am.* 29,

626-631.

Vanden Berge, J. C. (1979): "Myologia". In *Nomina Anatomica Avium* (J. J. Baumel *et al.*, eds), Academic Press, London and New York.

Van Tyne, J. and Berger, A. J. (1959): *Fundamentals of Ornithology*, John Wiley and Sons, New York.

Van Tyne, J. and Berger, A. J. (1976): *Fundamentals of Ornithology* (2nd edn), John Wiley and Sons, New York.

Waite, E. R. (1903): "Sympathetic song in birds". *Nature* 68, 322.

Wallschläger, D. (1981): "Analyse von konstruktionsprinzipien der Lautäusserungen eines 'sprechenden' Wellensittichs (*Melopsittacus undulatus*)". *Mitt. Zool. Mus. Berlin*, Suppl. Band 57, Ann. Orn. 5, 1-13.

Warner, R. W. (1969): *The Anatomy of the Avian Syrinx* (Ph. D. thesis), University of London.

Warner, R. W. (1971): "The structural basis of the organ of voice in the genera *Anas* and *Aythya* (Aves)". *J. Zool. Lond.* 164, 197-207.

Warner, R. W. (1972a): "The syrinx in family Columbidae". *J. Zool. Lond.* 166, 385-390.

Warner, R. W. (1972b): "The anatomy of the syrinx in passerine birds". *J. Zool. Lond.* 168, 381-393.

Waters, R. A. and Wilson, W. A. (1976): "Speech perception by rhesus monkeys". *Perc. & Psych.* 19, 285-289.

Watkins, W. A. (1967): "The harmonic interval: fact or artifact in spectral analysis of pulse trains". In *Marine Bio-acoustics* (W. N. Tavolga, ed.), Pergamon Press, Oxford.

West, M. J., Straud, A. N. and King, A. P. (1983): "Mimicry of the human voice by European starlings: the role of social interaction". *Wilson Bull.* 95, 635-640.

White, S. J., White, R. E. C. and Thorpe, W. H. (1970): "Acoustic basis for individual recognition by voice in the gannet". *Nature* 225, 1,156-1,158.

White, S. S. (1968a): "Movements of the larynx during crowing in the domestic cock". *J. Anat.* 103, 390-392.

White, S. S. (1968b): "Mechanisms involved in deglutition in *G. domesticus*". *J. Anat.* 104, 177.

White, S. S. (1975): "Larynx" (in A. S. King: "Aves respiratory system"). In Sisson and Grossman's *The Anatomy of the Domestic Animals* (R. Getty, ed.), W. B. Saunders Co., Philadelphia.

White, S. S. and Chubb, J. C. (1968): "The muscles and movements of the larynx of *Gallus domesticus*". *J. Anat.* 102, 575.

Whitefield, I. C. (1970): "Central nervous processing in relation to spatio-temporal discrimination of auditory patterns". In *Frequency Analysis and Periodicity Detection in Hearing* (R. Plomp and G. F. Smoorenburg, eds), Sijthoff, Leiden.

Whitfield, I. C. and Evans, E. F. (1965): "Responses of auditory cortical neurons to stimuli of changing frequency". *J. Neurophysiol.* 28, 655-672.

Wildenthal, J. L. (1965): "Structure in primary song of the mockingbird (*Mimus polyglottos*)". *Auk* 82, 161-189.

Wiley, R. H. (1971): "Song groups in a singing assembly of little hermits". *Condor* 73, 28-35.

Wilkinson, R. and Howse, P. E. (1975): "Time resolution of acoustic signals by birds". *Nature* 258, 320-321.

Wilson, T. A., Beavers, G. S., DeCoster, M. A., Holger, D. K. and Regenfuss, M. D. (1971): "Experiments on the fluid mechanics of whistling". *J. Acoust. Soc. Am.* 50, 366-372.

Wind, J. (1976): "Phylogeny of the human vocal tract". *Ann. N.Y. Acad. Sci.* 280, 612-630.

Wollberg, Z. and Newman, J. D. (1972): "Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations". *Science* 175, 212-214.

Wood, C. C. (1974): "Parallel processing of auditory and phonetic information in speech discrimintion". *Perc.*

Psych. 15, 501-508.

Wood, C. C. (1975): "Auditory and phonetic levels of processing in speech perception: neurophysiological and information-processing analyses". *J. Exp. Psychol.* 104, 3-20.

Wood, C. C., Goff, W. R. and Day, R. S. (1971): "Auditory evoked potentials during speech perception". *Science* 173, 1,248-1,251.

Woolf, N. K. and Sachs, M. B. (1977): "Phase locking to tones in avian auditory-nerve fibres", *J. Acoust. Soc. Am.* 62, S46.

Wurdinger, I. (1970): "Erzengung, Ontogenie und Funktion der Lautausserungen bei vier Gansearten (*Anser indicus, A. caerulescens, A. albifrons und Branta canadensis*)". *Z. Tierpsychol.* 27, 257-302.

Wyndham, E. (1980): "Diurnal cycle, behaviour and social organization of the budgerigar (*Melopsittacus undulatus*)". *Emu* 80, 25-33.

Yoshimoto, C., Hirahara, T. and Ifukube, T. (1978): "Speech production processes in the mynah". *J. Acoust. Soc. Am.* 64, S94.

Youngren, O. M., Peek, F. W. and Phillips, R. E. (1974): "Repetitive vocalizations evoked by local electrical stimulation of avian brains III: evoked activity in the tracheal muscles of the chicken (*Gallus gallus*)". *Brain, Behav. Evol.* 9, 393-421.

Zaretsky, M. D. (1978): "A new auditory area of the songbird forebrain: a connexion between auditory and song control centres". *Exp. Brain Res.* 32, 267-273.

Zaretsky, M. D. and Konishi, M. (1976): "Tonotopic organization in the avian telencephalon". *Brain Res.* 111, 167-171.

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B. and Stebbins, W. (1979): "Species-specific perceptual processing of vocal sounds by monkeys". *Science* 204, 870-873.

Zweers, G. A. (1982a): "Drinking in the pigeon (*Columba*

livia L.)". *Behaviour* 80, 274-317.

Zweers, G. A. (1982b): "Pecking of the pigeon (*Columba livia* L.)". *Behaviour* 81, 173-230.

Zweers, G. A. (1982c): *The Feeding System of the Pigeon (*Columba livia *L)* (Vol. 73 of Advances in Anatomy, Embryology and Cell Biology), Springer-Verlag, Berlin.

Zweers, G. A. (1985): "Generalism and specialism in avian mouth and pharynx". *Fortschr. Zool.* 30, 189-201.

Zweers, G. A., van Pelt, H. C. and Beckers, A. (1981): "Morphology and mechanics of the larynx of the pigeon (*Columba livia* L.): a drill-chuck system". *Zoomorphology* 99, 37-69.

Zwicker, E. (1952): "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones". *Acoustica* 2, 125-133.

Zwicker, E., Flottorp, G. and Stevens, S. S. (1957): "Critical bandwidths in loudness summation". *J. Acoust. Soc. Am.* 29, 548-557.