

## ABSTRACT

Title of Document: THE STRUCTURE AND PERCEPTION OF  
BUDGERIGAR (*MELOPSITTACUS*  
*UNDULATUS*) WARBLE SONGS

Hsiao-Wei Tu, Ph.D., 2009

Directed By: Professor Robert J. Dooling, Department of  
Psychology

The warble song of male budgerigars (*Melopsittacus undulatus*) is an extraordinarily complex, multi-syllabic, learned vocalization that is produced continuously in streams lasting from a few seconds to a few minutes without obvious repetition of particular patterns. As a follow-up of the warble analysis of Farabaugh et al. (1992), an automatic categorization program based on neural networks was developed and used to efficiently and reliably classify more than 25,000 warble elements from 4 budgerigars. The relative proportion of the resultant seven basic acoustic groups and one compound group is similar across individuals. Budgerigars showed higher discriminability of warble elements drawn from different acoustic categories and lower discriminability of warble elements drawn from the same category psychophysically, suggesting that they form seven perceptual categories corresponding to those established acoustically. Budgerigars also perceive individual voice characteristics in addition to the acoustic measures delineating categories. Acoustic analyses of long sequences of natural

warble revealed that the elements were not randomly arranged and that warble has at least a 5<sup>th</sup>-order Markovian structure. Perceptual experiments provided convergent evidence that budgerigars are able to master a novel sequence between 4 and 7 elements in length. Through gradual training with chunking (~5 elements), birds are able to master sequences up to 50 elements. The ability of budgerigars to detect inserted targets taken in a long, running background of natural warble sequences appears to be species-specific and related to the acoustic structure of warble sounds.

THE STRUCTURE AND PERCEPTION OF BUGERIGAR (*MELOPSITTACUS  
UNDULATUS*) WARBLE SONGS

By

Hsiao-Wei Tu

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2009

Advisory Committee:

Professor Robert J. Dooling, Chair  
Distinguished Univ. Prof. Emeritus William Hodos  
Associate Professor David D. Yager  
Associate Professor Rochelle S. Newman  
Professor Juan Uriagereka

© Copyright by  
Hsiao-Wei Tu  
2009



## Dedication

To my family and my friends all over the world.

Thank you all.

## Acknowledgements

First, I would like to thank my advisor, Bob Dooling, for six years of advice, guidance, and mentorship. I would also like to thank my dissertation committee members, Bill Hodos, David Yager, Juan Uriagereka, and Rochelle Newman, for all the help, comments, and of course, the mental therapy sessions you gave me.

I am grateful for all my colleagues and friends, both past and present. I could not have made this far without each of you. I thank Beth Brittan-Powell for always being there to calm me down and helping me communicate with Bob; Peter Marvit for numerous discussions and suggestions; Ed Smith for technical support; Chen Chiu for her encouragement and friendship over the years; and Rex Butler for being a good listener and dance partner. I also thank all my friends in Maryland: Sandra Blumenrath, Arthur Chang, Amanda Lauer, Daniel Lindsey, Bernie Lohr, Vanessa Medley, Mike Osmanski, Yoshimasa Seki, Megan Shaw, Beth Vernaleo, Vivian Wang; and in Taiwan: Dr. Jiun-Hong Chen, Tsai-Yu Wu, Po-Yi Tu, and all my friends in Physics and Zoology.

I would like to thank Yichun Cheng in particular for her kindness, patience, and emotional support that helped me keep my sanity along this long journey. Her encouragement allowed me to brave all the frustrations with a smile. Things would not have been the same without her.

Finally, I would like to thank my parents, my brother, and my grandfather for having faith in me, knowing that I can take care of myself overseas and letting me chase my dream.

# Table of Contents

Dedication .....	ii
Acknowledgements .....	iii
Table of Contents .....	iv
List of Tables .....	vii
List of Figures .....	viii
Chapter 1: Introduction .....	1
Songbirds .....	1
Budgerigars .....	7
Syntactic capacities of nonhuman animals .....	11
The goals of this research .....	15
Chapter 2: General Methods .....	18
Subjects .....	18
Vocalization recording .....	19
Segmentation of vocalizations .....	20
Training .....	22
Apparatus .....	22
Procedures .....	25
Chapter 3: Acoustic Categories of Budgerigar Warble Elements .....	29
Experiment 1: Acoustic categorization by humans .....	31
Background and rationale .....	31
Method .....	32
Results .....	35
Discussion .....	39
Chapter 4: Perceptual Categories of Budgerigar Warble Elements .....	43
Experiment 1: Perceptual categorization by budgerigars .....	44
Background and rationale .....	44
Method .....	46
Results .....	47
Discussion .....	51
Experiment 2: Variations within categories .....	54
Background and rationale .....	54
Method .....	55
Results .....	57
Discussion .....	59
Chapter 5: A Special Case: Contact Calls and Warble Calls .....	62
Experiment 1: Acoustic analysis .....	63
Vocal stimuli .....	63
Method .....	64
Results .....	66
Discussion .....	70
Experiment 2: Perceptual analysis .....	72
Vocal stimuli .....	72
Method .....	72

Results.....	73
Discussion.....	76
Chapter 6: Sequential Analysis of Budgerigar Warble.....	79
Experiment 1: Acoustic analysis of warble sequences.....	80
Background and rationale.....	80
Method.....	81
Results.....	84
Discussion.....	88
Chapter 7: Perception of Sequences of Warble Elements.....	91
Experiment 1: Budgerigars' sequencing ability on familiar sequences.....	91
Background and rationale.....	91
Method.....	92
Results.....	94
Discussion.....	95
Experiment 2: Budgerigars' sequencing ability on novel sequences.....	97
Background and rationale.....	97
Method.....	98
Results.....	98
Discussion.....	99
Chapter 8: Detection of Insertions in Warble Sequences.....	101
Experiment 1: Detection of non-budgerigar vocalizations in natural warble sequences.....	102
Background and rationale.....	102
Method.....	102
Results.....	104
Discussion.....	105
Experiment 2: Detection of budgerigar calls in natural warble sequences.....	106
Background and rationale.....	106
Method.....	107
Results.....	108
Discussion.....	113
Experiment 3: Detection of warble calls in warble sequences.....	115
Background and rationale.....	115
Method.....	115
Results.....	117
Discussion.....	121
Chapter 9: General Discussion.....	126
Acoustic perceptual categories for communication signals.....	127
Contact calls vs. warble calls.....	129
The structure of budgerigar warble and its biological importance.....	130
The salience of sequence cues in warble.....	131
Relevance of present studies for animal models of human speech communication.....	133
Conclusions.....	134
Appendix I: Measures for the Classifier.....	135
Introduction.....	135

Features .....	135
Footnote .....	141
Appendix II: Multiple Neural Networks using Majority Vote Classifier .....	142
Objective .....	142
The classifier is based on neural networks (NNs) .....	142
Large and redundant feature set .....	143
Approach .....	143
Classifier construction .....	143
Classifier operation .....	144
One competing NN classifier approach .....	145
Bibliography .....	146

## List of Tables

Table 1: Auto-shaping training phases. ....	26
Table 2: Twenty acoustic measures (see Appendix I) taken by the neural networks and their relative merit in categorizing warble segments. ....	34
Table 3: Twenty measurements used in the comparison between contact calls and contact call-like warble elements.....	65
Table 4: Comparisons of 20 acoustic measures.....	69
Table 5: Summary of PCA result.....	70
Table 6: Transition matrix of the longest sequence in each budgerigar. ....	85
Table 7: Summary of $d'$ values and standard errors of budgerigars' performance on the original sequences and the new sequences. ....	98
Table 8: Summary of $d'$ values and standard errors of the subjects' sensitivity for detecting pure tones and zebra finch song syllables. ....	104
Table 9: Summary of $d'$ values and standard errors of the subjects' sensitivity for detecting contact calls and warble calls.....	109
Table 10: Summary of $d'$ values and standard errors of the subjects' sensitivity for detecting contact calls (A) and warble calls (B) from the same individual as the background and from different individuals than the background.....	110
Table 11: Summary of $d'$ values and standard errors of the subjects' sensitivity for detecting contact calls (A) and warble calls (B) from a collection of unrelated warble calls and against a natural warble sequence.....	112
Table 12: Summary of $d'$ values and standard errors of the budgerigars' performance on detecting warble call insertions in different backgrounds. ....	118
Table 13: Summary of $d'$ values and standard errors of the subjects' sensitivity for detecting normal warble calls and reversed warble calls.....	123

## List of Figures

Figure 1: Demonstration of the custom-written MATLAB segmentation program. ..	21
Figure 2: Photograph of a budgerigar in the operant testing apparatus. ....	24
Figure 3: Example of a 20-second piece of warble recorded from a male budgerigar; frequency on the y-axis and time in ms on the x-axis. ....	31
Figure 4: Examples of the seven elemental groups (A – G) and the compound group (H) where a contact call-like element immediately followed by a broadband sound. ....	36
Figure 5: The range of acoustic measures in each group. Six measures were shown here as an example. They have the highest relative merit (shown in parentheses) in simulating human categorization. ....	38
Figure 6: Distribution of the elements in different categories among individuals. ....	39
Figure 7: D prime of between-group and within-group discriminations. ....	48
Figure 8: D primes for within-group comparisons for budgerigars within-group comparisons. ....	49
Figure 9: Three-dimensional spatial representation of the result from INDSCAL. ...	50
Figure 10: Stimuli used in Experiment 2. Eight elements form a similarity gradient within the same category. Three categories, contact call-like group, alarm call- like group, and short harmonic group, were tested. (Cross-correlation values with the prototype were indicated on top of each column.).....	56
Figure 11: Three different species perceived variations in each of the three acoustic categories differently. ....	58
Figure 12: Overall performance of all four species on discriminating within-category variations in alarm call-like, contact call-like, and short harmonic warble elements groups. ....	59
Figure 13: Three-dimensional MDS plots for each budgerigar showing clustering patterns for contact calls and warble calls. ....	67
Figure 14: Results of the psychophysical test on calls of each individual budgerigar. .....	74
Figure 15: Canaries and zebra finches were less sensitive to the difference between contact calls and warble calls.....	75

Figure 16: Salience of individual vocal characteristics. ....	76
Figure 17: Distribution of different categories in the longest sequence of each bird. ....	84
Figure 18: Conditional uncertainty plotted as a function of the order of Markov chain for each budgerigar. ....	87
Figure 19: The performance of all subjects detecting insertions in a sequence.....	95
Figure 20: (A) Comparison of the sensitivity between males ( $n = 2$ ) and females ( $n = 2$ ). (B) Comparison of the sensitivity between different stimuli. ....	97
Figure 21: Comparison of budgerigars' sensitivity to detect a change in the original familiar sequences and the new sequences composed of the same elements. ....	99
Figure 22: The performance of all three species detecting insertions of pure tones and zebra finch syllables.....	105
Figure 23: The performance of all three species detecting insertions of contact calls and warble calls. ....	108
Figure 24: Comparison between targets from the same individual in the background and targets from a different individual than the one in the background. ....	111
Figure 25: Cross-species comparison of the sensitivity of detecting a contact call (A) or a warble call (B) among a group of individual calls (Chapter 5) and in a continuous warble sequence (present experiment). ....	113
Figure 26: Comparison of budgerigars' ability to detect a warble call inserted in different background warble sequences.....	119
Figure 27: Comparison of budgerigars' ability to detect insertions of warble calls in different background warble sequences. (A) Analysis of inserted warble calls from the same recording session as the background warble sequence. (B) Analysis of inserted warble calls from a different recording session of the same individual being inserted in different background warble sequences.....	120
Figure 28: Cross-species comparison between detection of normal warble calls and temporally reversed warble calls. ....	124



# Chapter 1: Introduction

## Songbirds

Communication systems of nonhuman animals have attracted researchers for many years. Intrinsically intriguing and complex, the topic provides a chance to examine comparatively the unique and common properties of human language. Bird vocalization, especially songs of oscines, may be the most widely-used and well-understood system among all the studies.

Compared to bird calls that are simple and short vocalizations, songs are relatively longer, more complex, and mainly produced by male songbirds for reproductive purposes such as attracting mates and defending territories in most songbirds (Catchpole & Slater, 2008). Birdsong can be described at many levels where elements (defined as a continuous marking on a sound spectrogram) are often combined to make syllables – a group of elements separated by brief silent intervals but always uttered together. Syllables can further be combined to make phrases or motifs, which in turn are combined to create songs (Catchpole & Slater, 2008). Similarly, this hierarchical structure also appears in human speech where phonemes are combined to make words, which in turn are combined to make phrases and sentences. This shared feature is one of a numbers of parallels between these two complex acoustic systems.

There is a substantial body of research covering various aspects of songbird songs since the 1950s (see review in Marler, 2004). Developmentally, songbirds (Passeriformes) are one of the four avian orders that have shown evidence of vocal learning like humans do (e.g., Kroodsma, 1982; Nottebohm, 1972). Significant studies

have examined the two-step learning process of species-specific songs in songbirds (Konishi, 1965; Konishi & Nottebohm, 1969; Marler, 1970b). First, in the sensory learning phase, hatchling birds listen and memorize what they hear to form a template for later vocal production. Although they are unable to translate the sounds into the correct motor pattern, and therefore no song is produced in this phase, it is believed that young birds are predisposed to be more sensitive to the characteristics of conspecific vocalizations that direct their future song learning than those of heterospecific vocalizations (Baptista, 1996; Dooling & Searcy, 1980; Marler & Peters, 1989; Nelson & Marler, 1993).

Shortly after the sensory learning phase, young birds enter the sensory-motor phase and start to vocalize (Doupe & Kuhl, 1999; Hultsch & Todt, 2004; Konishi, 1965; Konishi & Nottebohm, 1969; Marler, 1970b). At this stage, the vocalizations that babies remember from their sensory phase become the “templates” that serve as guides for vocal development (Catchpole & Slater, 2008; Kuhl & Meltzoff, 1996). At first, similar to the babbles of human infants (Doupe & Kuhl, 1999), birds produce subsongs that are low-pitched and unstructured (Hultsch & Todt, 2004). After repeatedly comparing and matching their own vocalization to the templates, the vocalization becomes more structured and more similar to the templates (plastic song in birds (Hultsch & Todt, 2004); canonical babble in human infants (Doupe & Kuhl, 1999)). However, the variations are overgenerated more than the amount of a usual adult repertoire. Eventually, overproduced sounds are abandoned, and crystallized adult-like species-specific vocalizations are produced (Doupe & Kuhl, 1999; Hultsch & Todt, 2004).

During vocal learning, auditory feedback is important for songbirds. It allows them to compare what they hear from other individuals, their tutors, to what they produce in order to refine, rehearse, and develop normal vocalization. For example, white-crowned sparrows and song sparrows deafened early in life fail to develop normal species-specific songs but produce abnormal series of sounds (Marler & Sherman, 1983). The abnormalities range from changes in the forms and patterns of elements to completely unrecognizable song structures (Konishi & Nottebohm, 1969).

Another crucial factor during vocal development and learning is social interaction. It plays an essential role both quantitatively and qualitatively. For instance, white-crowned sparrows learn from a heterospecific song sparrow as long as they are exposed to a live tutor (Baptista & Petrinovich, 1986). Male cowbirds modify their song syllables and song rate according to the reaction of female birds even though the females do not sing (A. P. King, West, & Goldstein, 2005). In fact, it is not necessarily for some species to interact with a “live” tutor. If a key peck triggered playback of a conspecific song, zebra finches actively pay attention to the task and learn the song, but they do not learn from passive exposure to taped conspecific songs (Adret, 1993).

Anatomically, bird vocalizations are generated in the syrinx, the vocal organ located at the base of the trachea (A. S. King, 1989) which is lower than where the human larynx is. More specifically, the syrinx in songbirds is situated at the junction of the two bronchi with the trachea, creating two independent sound sources (Catchpole & Slater, 2008; Mindlin & Laje, 2005). Above the syrinx, the air flow goes through the trachea and the larynx to the beak. Although it is still debated whether the vocal tract (resonator) is coupled to the syrinx (source) in songbirds (Nowicki, 1987; Nowicki & Marler, 1988;

Westneat, Long, Hoese, & Nowicki, 1993), evidence has shown that the characteristics of the vocal tract, especially the size of beak opening, play a central role in the quality of bird vocalizations (see review in Podos & Nowicki, 2004).

Physiologically, songbirds have evolved specialized, hierarchical neural pathways for vocal control, like humans. At the lowest level, motor neurons of the nucleus ambiguus control the vocal organ through cranial nerves, while medullary motor neurons control the respiratory muscles as well as the vocal tract structures (Butler & Hodos, 2005; Doupe & Kuhl, 1999; Wild, 1997). At the next higher level, midbrain areas are crucial for vocalization. The mesencephalic nucleus, nucleus dorsomedialis, is involved in eliciting innate species-specific vocalizations (Balaban, Teillet, & Douarin, 1988; Doupe & Kuhl, 1999).

At higher levels, instead of a multi-layered cortex like humans, birds' forebrain is organized as discrete nuclei. There are seven telencephalic vocal nuclei interconnected into two main pathways in songbirds (Brenowitz & Kroodsma, 1996; Jarvis, 2004). Along these two pathways, several important nuclei, their projections, and their corresponding roles in song production and learning have been identified and compared to specific brain areas in humans. For instance, in songbirds, premotor neurons in the high vocal center (HVC) seem to encode song patterns and syllable identity during singing, which leads to their parallel to human's language control areas. The nucleus robustus arcopallii (RA), which is thought to be parallel to human's motor cortex (Doupe & Kuhl, 1999; Vu, Mazurek, & Kuo, 1994; Yu & Margoliash, 1996), receives signals from HVC and is associated with the components of syllables. Moreover, lesions to the cortical interfacial nucleus (NIf) in the nidopallium lead to highly variable syntax in their

songs (Hosino & Okanoya, 2000) or deficits in vocal imitation (Plummer & Striedter, 2002).

By and large, our understanding of the communication system in songbirds has been greatly advanced over the past five decades. Numerous superficial parallels between birdsongs and human speech have been found along the way, making oscine songs the leading animal model for examining similarities to human language (Brainard & Doupe, 2002; Doupe & Kuhl, 1999; Goldstein, King, & West, 2003; Marler, 1970a; Marler & Peters, 1981; Todt, 2004; Wilbrecht & Nottebohm, 2003).

In addition to the similar hierarchical structure of birdsong and human speech mentioned above, developmentally human infants also acquire their language through a two-stage process – perceptual learning phase where no speech-like is produced but babies are perceptually able to discriminate phonetic contrasts of all languages (Kuhl, 1994; Kuhl, Kiritani, Deguchi, & Hayashi, 1997; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 2002), and sensory-motor learning where they babble and finally start to form words and sentences (Doupe & Kuhl, 1999; Stone & Stoel-Gammon, 1994). It has been shown that specific language experience shapes babies' perception (Kuhl, 1994; Kuhl, et al., 1997), and by the time they are ready for language-specific speech production, they are perceptually oriented toward the language that they have been exposed to the most, usually their mother tongue. In other words, like what was found in songbirds, human are born with the ability to learn and be perceptually prepared before specific vocal production (Doupe & Kuhl, 1999).

Furthermore, auditory feedback and social interaction during learning of normal vocalization in human are as important as they are in songbirds. Deaf infants begin to

babble much later than normal babies and do not reach beyond a certain stage of babbling due to lack of auditory feedback (Oller, Eilers, & Bull, 1985). Rare cases like Genie, who was reared with little language input and social interaction for 12 years beginning around 14 months of age, show that the progress in learning language is retarded and less successful than that in normal children (Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974). Studies of mothers' responsiveness to their babies' vocalizations showed that infants in an experimental group where mothers were told to respond immediately to their child's vocalizations (smile at, touch, etc.) not only produced more utterances (quantitatively), but also had more adult-like vocalizations (qualitatively) than those in a control group (Goldstein, et al., 2003).

However, as more is known about these two systems, and deeper comparisons reveal the limitations of the similarities, songbirds no longer seem to be an adequate model of human communication systems. First, most studies of songbirds have focused on age-dependent learners, who have a strictly defined timing of vocal learning beyond which they are no longer able to learn new songs (e.g., white-crowned sparrows (Marler, 1970b), zebra finches (Zann, 1996), and song sparrows (Marler & Peters, 1987)), or seasonal learners, whose ability to learn production of new song reopens seasonally in adulthood (e.g., canaries (Nottebohm, Nottebohm, & Crane, 1986)). Research on open-ended learners, like humans who have the ability to learn new vocalizations throughout their life (e.g., European starlings (Eens, Pinxten, & Verheyen, 1992) and nightingales (Hultsch & Todt, 2004)) is relatively sparse.

Secondly, unlike humans who talk for multiple purposes under various scenarios, songbirds sing primarily in order to defend their territory and attract mates. They seldom

use their songs in other situations. More specifically, although birdsongs and human speech both have a hierarchical structure, individual elements in oscine songs are usually not produced alone without a complete sequence, whereas human speech can be separated into single words that can be produced alone.

Lastly, songbird songs are generally short (a few seconds) and stereotyped. They usually sing one or more “song types” (particular patterns of elements) repeatedly in one song bout. The sequential order of song elements is so typical and predictable that it is unlikely that information is encoded in different combinations of song elements. Indeed, syntax, the rules of combining words into sentences, seems to be a unique feature that makes an infinite range of expressions in human language.

### *Budgerigars*

Recently, more and more work has been done on a social, non-oscine parrot, the budgerigar (*Melopsittacus undulatus*), increasing our understanding of their species-specific vocal system, and also suggesting that they may serve as a better model when compared with human language capabilities (Brittan-Powell, Dooling, & Farabaugh, 1997; Dooling, Best, & Brown, 1995; Dooling & Brown, 1990; Dooling, Okanoya, & Brown, 1989; Farabaugh & Dooling, 1996).

The budgerigar is a small parrot native to Central Australia. They are non-territorial, group-living parakeets that usually form large flocks (from hundreds of individuals to over 25 thousands), foraging and pairing together (Brockway, 1964b; Wyndham, 1980). They are opportunistic breeders that can breed whenever conditions are favorable (Trillmich, 1976c). In order to coordinate social behaviors in the flock, reinforce pair bond between mates, and breed successfully, vocal communication is very

important for them (Brockway, 1964a, 1964b; Farabaugh & Dooling, 1996; Trillmich, 1976c).

Contact calls and warble songs are the two main classes of vocalization in budgerigars' repertoire. Contact calls (approximately 100-300 ms) are narrowband (2-4 kHz) and strongly frequency-modulated, produced as single instances without a fully constructed sequence by both genders. It is believed that budgerigars use contact calls to coordinate, localize and synchronize the flock (Farabaugh, Linzenbold, & Dooling, 1994). Contact calls have been extensively investigated in many aspects, including basic auditory perception (e.g., Brown, Dooling, & O'Grady, 1988; Dooling, et al., 1995; Dooling & Brown, 1990; Dooling, Brown, Park, Okanoya, & Soli, 1987; Dooling, et al., 1989; Dooling, Park, Brown, Okanoya, & Soli, 1987; Park & Dooling, 1986), vocal development (e.g., Brittan-Powell, Dooling, & Farabaugh, 1997), vocal plasticity (e.g., Farabaugh, et al., 1994; Hile, Plummer, & Striedter, 2000; Hile & Striedter, 2000), and vocal control (e.g., Manabe, Dooling, & Brittan-Powell, 2008; Manabe, Sadr, & Dooling, 1998; Osmanski & Dooling, 2009), mainly because they are easy to elicit in experimental setups and straightforward to categorize and analyze.

The other vocalization in budgerigars' repertoire is the warble song, characterized as a melodic, continuous multi-syllabic vocalization that lasts as long as several minutes (Farabaugh, Brown, & Dooling, 1992). It is primarily produced by males when courting females, accompanied by various courtship behaviors (Brockway, 1964b), though occasionally sung by females as well (Wyndham, 1980). Warble, especially certain low-pitched components like "tuk-tuk" and "whedelee," plays an important part in budgerigar courtship (Brockway, 1961, 1962, 1969). Male warble initiates pair formation and



reinforces the pair bond (Brockway, 1964b). Hearing warble promotes egg laying and ovarian development in females (Brockway, 1961, 1964b, 1965, 1967b) and stimulates more extensive precopulatory behaviors and sperm production in males (Brockway, 1964c, 1968). Moreover, warble of others increases a male's tendency to perform his own warble (Brockway, 1964b, 1964c, 1969), which in turn raises the level of circulating gonadal steroids and further lowers the thresholds for warbling (Brockway, 1964c, 1969). This vocal-endocrinological feedback interaction has been shown to be critical in the reproduction of budgerigars (Brockway, 1969).

As a proposed non-songbird model to compare with human language, budgerigars have a syrinx that is entirely tracheal, which results in a unitary sound source similar to the larynx in human (Brittan-Powell, Dooling, Larsen, & Heaton, 1997; Heaton, Farabaugh, & Brauth, 1995; A. S. King, 1989). Their vocal learning ability is also more closely analogous to humans than to songbirds. Budgerigars' sensory learning phase starts about 11 days after hatching when they develop their hearing sensitivity, and their sensory-motor learning phase probably begins from the time the bird first produces a contact call, usually around 35-60 days after hatching (Brittan-Powell, 2002; Brittan-Powell, Dooling, & Farabaugh, 1997). They are subsequently open-ended learners that are capable of learning both biological and non-biological sounds and incorporating them into their repertoire during their lifetime (Farabaugh & Dooling, 1996; Gramza, 1970). Flock mates usually learn from each other to form a shared call type (Brown, et al., 1988; Farabaugh, et al., 1994; Hile, et al., 2000; Hile & Striedter, 2000; Striedter, Freibott, Hile, & Burley, 2003), independent of sex, age and season (Brittan-Powell, Dooling, & Farabaugh, 1997; Farabaugh & Dooling, 1996; Farabaugh, et al., 1994; Hile, et al., 2000;

Hile & Striedter, 2000; Striedter, et al., 2003). They also share a higher proportion of similar warble elements than those living apart (Farabaugh, et al., 1992).

In addition, auditory feedback during vocal learning in budgerigars is important and necessary as in songbirds and human. Deafened nestlings are able to produce food-begging calls, but they never successfully transform those calls into typical contact calls (Heaton & Brauth, 1999). Individuals raised in acoustic isolation develop highly aberrant warble song (Eda-Fujiwara & Okumura, 1992). Even deafening in adulthood makes the bird produce fewer and severely abnormal vocalizations (Heaton, Dooling, & Farabaugh, 1999). Recent experiments further show that delayed auditory feedback disrupt budgerigars' vocal production (Osmanski & Dooling, 2009), just like a small manipulation (e.g., incorrect or delayed) of auditory feedback could affect the fluency of human speech (Howell & Archer, 1984).

Likewise, social interaction is crucial to vocal learning in budgerigars. Individuals that can see, hear, and interact with one another develop similar contact calls through imitation within a very short time. Vocal imitation is greatly reduced in situations where birds can only hear each other but cannot see and interact with each other (Farabaugh & Dooling, 1996; Farabaugh, et al., 1994).

Perhaps more interestingly and beyond the limit of songbird models, the acoustic complexity, non-repeating structure, and unusual length of budgerigar warble open the door to extend our knowledge of temporal processing to the perception of the serial order of elements. Warble allows us to further probe higher order questions such as how information is encoded and whether the sequential order of warble elements is produced and perceived according to some type of rule. In other words, it provides a natural system

to study the possibility that there may be learned syntactical structure in an animal vocalization – another potential parallel to human language.

### *Syntactic capacities of nonhuman animals*

In the simplest form, syntax might be generally defined as the rules of combining discrete components into fully-structured utterances. But the exact nature of syntax in complex communication systems like human language is still a matter of debate by cognitive psychologists and linguists (Bates, 2003; Chomsky, 1965, 1995; Pinker & Jackendoff, 2005). It is apparently unique to human language because of the production of an infinite range of expressions by the use of recursion and compositionality (Kirby, 2002). Recursion is the ability to combine finite lexica into infinite expressions (Hauser, Chomsky, & Fitch, 2002; Kirby, 2002), whereas compositionality is defined as the ability to recombine component sequences into different strings, where the meaning of each string is a product of the assembled meanings of its components and the way they are put together (Kirby, 2002; "lexical syntax" in Marler, 2000).

When studying animals' capacity for "syntax," some scientists focus on examining the animals' concept of human language and where their limits are on this capability by training and teaching animals with different tasks. For example, Fitch & Hauser (2004) presented cotton-top tamarins with two different AB grammars where A and B were two classes of consonant-vowel syllables. The animals could master the lowest level of grammar – finite state structures  $(AB)_n$  (for example, ABABAB when  $n = 3$ ) – but they were unable to master a grammar at a higher level – recursive  $A_nB_n$  (for example, AAABBB when  $n = 3$ ) structure as humans can. However, another experiment suggested that by using two different types of the species' own vocalizations, European

starlings can accurately recognize acoustic patterns defined by a recursive grammar ( $A_n B_n$ ) and reliably exclude agrammatical patterns by operant conditioning (Gentner, Fenn, Margoliash, & Nusbaum, 2006). Nevertheless, criticisms have been made about the thousands and thousands of training trials required before those birds could successfully complete the task, and that the achievement of the birds could be explained simply by a counting strategy instead of actual “understanding” of recursion (Corballis, 2007).

Taking the investigation of animals’ syntactical capacity one step further, some scientists have explicitly taught animals linguistic analogies of human language using signs or arbitrary symbol systems. Through intensive social interaction with human trainers, the famous African grey parrot (*Psittacus erithacus*), Alex, was able to produce and seemed to understand over a hundred English words (Pepperberg, 1992, 1997). Not only did he spontaneously recombine those vocabularies to make requests like “wanna go X” (where X is a location) and “wanna X” (where X is an object or food) (Pepperberg, 1988, 1990), but he also knew that the words are comprised of individual phonemes that can be recombined to create new referential vocalizations (Pepperberg, 2007). Another example comes from two artificial language-trained bottlenosed dolphins (*Tursiops truncatus*), Phoenix (trained on an acoustic language generated by computer and presented through an underwater speaker) and Akeakamai (trained on a visually-based language given by the gestures of a trainer's arms and hands). The dolphins showed the ability to comprehend 35-40 words, including objects, object modifiers, and actions, plus a set of syntactic rules that recombine the vocabulary elements to make novel, meaningful new sentences from two to five words in length (Herman, Kuczaj, & Holder, 1993; Herman, Richards, & Wolz, 1984).

Perhaps the most striking breakthrough in animal syntactic capacities came from the bonobo (*Pan paniscus*) Kanzi. Kanzi learned an artificial lexigram system “Yerkish” through both observation and ordinary, unreinforced “conversations” with his trainers (Savage-Rumbaugh, McDonald, Sevcik, Hopkins, & Rubert, 1986). He learned to use 90 symbols on a lexigram keyboard and has the capacity to comprehend spoken language in grammatical constructions of a complexity comparable to a human 2.5-year-old child. Clearly, Kanzi understands something about syntax. He responds differently to instructions such as “Put the juice in the egg” and “Put the egg in the juice,” which suggests that the order of words in spoken English sentences appears to be meaningful to him, even when the combination is novel and never presented to him before (see Savage-Rumbaugh, et al., 1993 for more details).

Research has thus shown that some animals are able to spontaneously recombine human vocabularies to make meaningful utterances. However, arguments could be made that it was achieved through intensive training, and the expressions were not what the animals naturally would do. Therefore, another group of scientists has chosen to approach the question of animal syntactic capacities by studying a variety of species to find linguistic parallels in the natural, spontaneous communication of other species. In this case, syntax is defined in a broad sense as any system with a set of rules that generates predictable sequences of behavior (Snowdon, 1990), even if the recombined sequence means the same as individual components (“phonological syntax” in Marler, 2000).

Indeed, some “rules” do exist in animals that govern the pattern of their acoustic communication. For example, the particular arrangement of syllables in a song is related to individual identities in gibbons (*Hylobates agilis*) (Mitani & Marler, 1989) and to

motivational states in capuchins (*Callicebus moloch*) (Robinson, 1984). A more specific analysis on the vocalization of the forest monkey showed that these primates not only recombine their calls, but also possibly change the meaning in them. Wild Diana monkeys (*Cercopithecus diana*) respond to male Campbell's monkeys' (*C. campbelli*) alarm calls with their own alarm calls. However, if Campbell's males emit a pair of low, resounding "boom" calls before their alarm calls, which indicates a less dangerous situation, or if recordings are experimentally played back with a boom-introduced Campbell's alarm call, Diana monkeys no longer respond. Additionally, when the booms precede the alarm calls of Diana monkeys, they are not effective either (Zuberbühler, 2002). Besides nonhuman primates, Kanwal et al. (1994) also showed that there may be several syntactical rules formulated in the repertoire of mustached bats (*Pteronotus parnellii*). For instance, a fixed sinusoidal frequency modulation syllable is usually followed by a short quasi-constant frequency or a short quasi-constant frequency-like sound, rather than any random combination.

Nevertheless, phonological syntax is most widely studied in songbirds. For instance, songs of willow warblers (*Phylloscopus trochilus*) have a hierarchically branching pattern, i.e., at some point along the song, certain elements always appear together with high predictability, while other positions along the song are "points of decision" where more choices can be made (Gil & Slater, 2000).

Chickadees are extensively studied for their syntactic songs. The two call systems of black-capped chickadee (*Poecile atricapilla*), "gargles" and "chick-a-dee" calls, are both combinatorial and composed according to certain syntactical rules (Ficken & Popp, 1992; Hailman, Ficken, & Ficken, 1985, 1987), and so are the chick-a-dee calls of

Carolina chickadees (*P. carolinensis*) (Bloomfield, Phillmore, Weisman, & Sturdy, 2005). Although those rules are relatively simple compared to human speech, they do convey a certain amount of information. By playback experiments, it has been shown that a manipulated order of the song will not produce normal responses from the receiver (see review in Lucas & Freeberg, 2007).

Sparrows are another species of focus. Soha & Marler (2001) showed that young, hand-fed white-crowned sparrows (*Zonotrichia leucophrys*), tutored with separate phrases of conspecific songs, could spontaneously reorganize the pieces they learned into species-typical sequences. This not only shows that there is a species-specific syntax in the songs of white-crowned sparrow, but also shows that syntax is to some extent pre-encoded in this species. Similarly, song sparrows (*Melospiza melodia*) would only learn swamp sparrow's (*Melospiza georgiana*) vocalizations when they are arranged in the same order as their own song (Marler & Peters, 1989). Another experiment showed that swamp sparrows from different populations have different song element ordering, and this difference is behaviorally salient to both males and females (Balaban, 1988).

### The goals of this research

This dissertation is focused on the structure and perception of budgerigar warble. Since relatively little research has been done on warble for its complexity and variability compared to contact calls, any finding related to warble would contribute to our understanding of budgerigar vocal system.

The first challenge is to provide a more complete portrait of male budgerigar warble songs. A set of acoustic categories of warble elements that occurs across birds will be computationally established so that long streams of warble can be analyzed more

precisely and efficiently. Budgerigars will then be tested by a psychoacoustic paradigm to show whether their perceptual categories match these acoustic categories. Acoustic and perceptual variations between and within categories will be examined, and the animals' categorization ability will be compared across species, including budgerigars, canaries, zebra finches, and human, to look for species-specificity. In particular, how similar, acoustically and perceptually, a bird's contact calls are to the contact call-like warble elements will be discussed.

The second half of this thesis is centered at the structure and perception of the sequential order of warble elements. Acoustic analysis will be used to investigate whether warble elements are arranged randomly or there are certain rules underlying their production, while perceptual experiments will test the budgerigar's ability to learn a novel sequence and investigate possible strategies they use to perform the task. Finally, budgerigars will be trained to detect different kinds of insertions in long streams of warble elements, as well as insertions in various types of warble streams. This will show the extent to which element order is important in warble perception. Other species will also be tested for a cross-species comparison and uncover possible species-specific advantages. Findings from these experiments will go a long way to advance our understanding of budgerigar warble, especially its potential "human language-like" characteristics.

To summarize, this thesis addresses the following questions:

- Are there well-defined acoustic categories of budgerigar warble elements?
- Do budgerigars have perceptual categories that match these human defined acoustic categories?



- How discriminable is acoustic variation with each acoustic category?
- Are contact calls (produced as single utterances) related to warble calls (call-like vocalizations occurring only in warble song) acoustically or perceptually?
- Is information coded in the sequences of elements in natural warble?
- Do budgerigars show species-specific perceptual sensitivities to element ordering in warble sequences?

## Chapter 2: General Methods

### Subjects

Three bird species, budgerigars (*Melopsittacus undulates*), canaries (*Serinus canarius*), and zebra finches (*Poephila guttata castanotis*), were used in this study. The actual number of individuals tested in the operant conditioning experiment will be specified later in each chapter. These birds were either bred from a laboratory flock at the University of Maryland or purchased from a local breeder. They were housed individually in small cages and kept on a constant 12-12 light-dark cycle. Since food was used as reinforcement, they were maintained at approximately 85-90% of their free-feeding weight with *ad libitum* access to water at all times. Budgerigars whose vocalizations were recorded for further analyses were all housed together in another big cage and kept on a constant light-dark cycle. They had *ad libitum* access to food and water all the time. The Animal Care and Use Committee of the University of Maryland, College Park approved all experimental procedures.

All human subjects were above twenty years old with normal hearing. (Specific numbers and genders are detailed in each chapter.) Before the experiment, the subjects were given detailed instructions by the experimenter and required to sign a consent form. All questions from the subjects regarding the experimental process were answered by the experimenter to ensure that the subjects understood the task. The human subject protocol was approved by the Institutional Review Board at UMCP.

### Vocalization recording

Warble vocalizations were obtained from several birds. In selecting subjects for recording, the birds' behaviors in the flock were observed for at least one hour a day for several days for evidence of pair bonding. If a male was seen warbling to a particular female at least once per day throughout the observation period, those two individuals were designated a pair and selected for recording. Approximately two to four weeks prior to the start of recording, all of the observed pairs were moved to a large flight cage in another room and housed together. Establishing a new 'flock' group helps to promote pair bonding and increase male warbling. Animals had *ad libitum* access to both food and water at all times.

Observations in our lab also suggest that budgerigars temporarily kept in isolation are more likely to vocalize. Therefore, prior to recording, a pair of budgerigars was separated and placed in a small animal acoustic isolation chamber (Industrial Acoustic Company model AC-1). After an isolation period of at least one hour, the doors of the chambers were opened and a recording session was begun.

During a recording session, the male was stimulated to vocalize by playing a low-amplitude recording of birds from the budgerigar flock room. Furthermore, experience suggests that being around other birds facilitates warble production. So, the male bird's mate was placed in close proximity so that the two could interact visually and acoustically (since both are important components for courtship during which warble song is produced). A single directional microphone (Audio-Technica Pro 35ax clip-on instrument microphone) was aimed at the male's cage so that the male's warble recordings were not contaminated by any vocalizations from the female. All vocalizations

were stored as a single channel of a PCM WAV file at a sampling rate of 48 kHz on a Marantz PMD670 digital recorder. Each recording session was terminated after the male stopped singing of his own accord. An aggregation of more than one hour of warble was collected over approximately four hours of recording. Animals were returned to the flight cage following each recording session.

Warble recordings were also obtained from a previous study by Farabaugh et al. (1992). The mated pairs were placed in an experimental chamber and separated by a Plexiglas divider, so that they could see and hear each other. Vocalizations from the two birds were recorded simultaneously on separate audio tracks of the same videotape using a Realistic Electret dynamic microphone and a Panasonic Omnivision VHS hifi-stereo video cassette recorder, model P-4960. Only the warble songs of one bird (Yuri) were used in the present study. They were also digitized at a 48 kHz sampling rate and stored on a computer together with the new warble recordings.

#### Segmentation of vocalizations

Each recording session was transferred from the Marantz digital recorder to a computer. Spectral analysis with an FFT was used to estimate the widest possible frequency range of vocalizations. Then, energy above and below this range (300 Hz – 12k Hz) was filtered out using Adobe Audition 2.0.

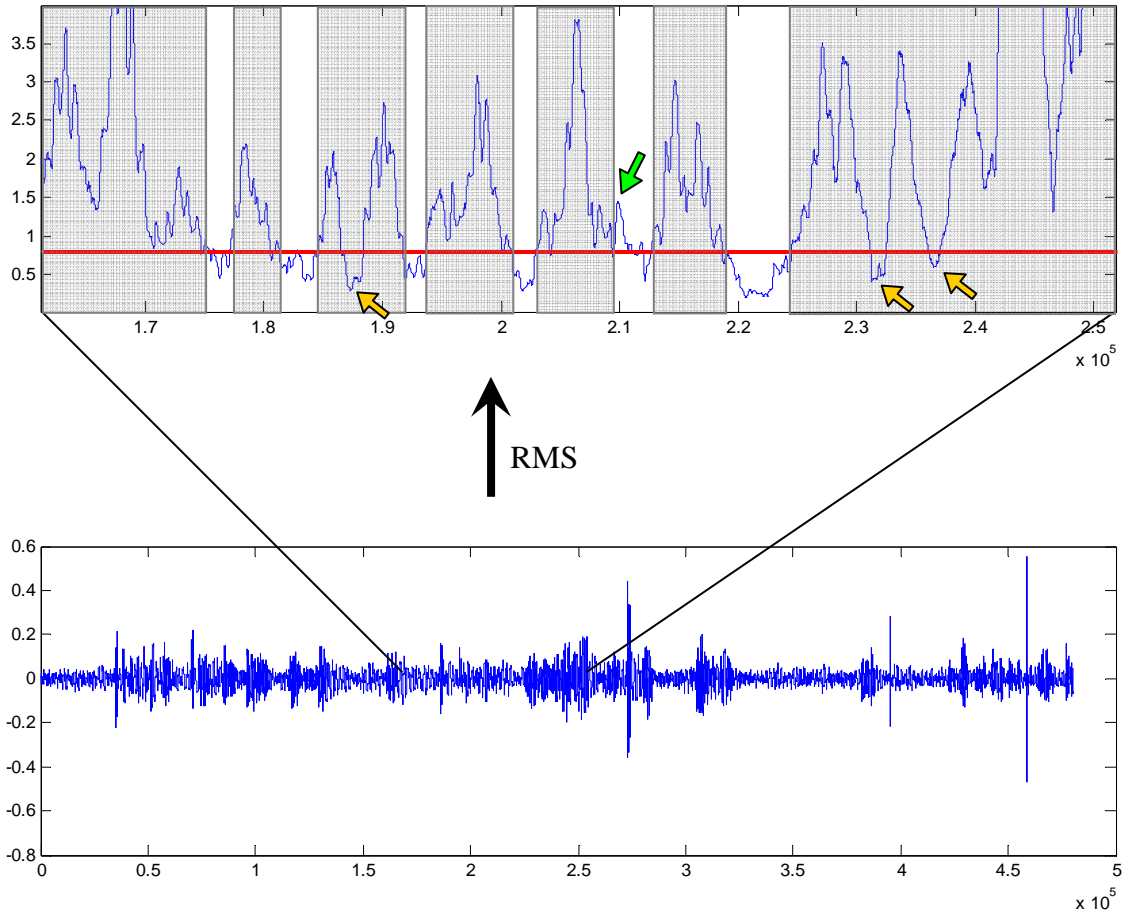


Figure 1: Demonstration of the custom-written MATLAB segmentation program. The lower panel shows the waveform of a section of warble where the rms amplitude envelope of part of it is enlarged in the upper panel. (The x axis in both panels represents sample points.) The red line indicates the amplitude threshold; the green arrow points to an instance where a segment was ignored because it was shorter than the duration threshold (1 ms); the yellow arrows points to instances where the interval between two potential segments was shorter than the interval threshold (25 ms) so they were combined as one segment. The shaded areas suggest the final segments that output from the program.

After filtering, a custom-written MATLAB segmentation program was used to segment the warble songs into acoustic elements or syllables. This program advanced through each WAV file, computing root-mean-square amplitude values using a 0.83 ms

window. From the resulting amplitude envelope of the whole warble song, the segmentation algorithm was used to break the warble into individual syllables (Figure 1). Three parameters were used to segment each warble song. First, an rms amplitude threshold was set according to the condition of each recording session, so that any amplitude envelope that continuously exceeded this threshold was considered as a segment (red line in Figure 1). Next, a duration threshold was selected as a cutoff point. Segments shorter than this cutoff duration were considered non-vocalizations and ignored. This parameter was set at 1 ms for every warble song in order to maximize the number of discrete segments (i.e., to minimize the number of compound segments) (green arrow in Figure 1). Finally, an interval threshold was selected. Intervals shorter than this cutoff point resulted in the two syllables being combined and counted as one. This parameter was set at a constant 25 ms. A number of pre-tests showed that most warble elements are separated by intervals greater than 25 ms (yellow arrows in Figure 1). In addition to producing individual WAV files of each warble element, the segmentation program also generates a file log indicating the start point, end point, duration, and sequential order of each segment in the original warble recording. With this file log, the complete natural warble sequence could be reconstructed.

### Training

### **Apparatus**

Birds were trained and tested in a small wire cage (23 x 25 x 16 cm<sup>3</sup>) mounted in a sound-attenuated chamber (Industrial Acoustics Company, Bronx, NY, IAC-3) lined with acoustic foam and illuminated with a 60-watt light bulb in a fixture at the top. In addition to the test cage in the chamber, a video camera system was mounted overhead to

monitor the animals at all times. A speaker was mounted from the roof at a 45 degree angle aimed toward the front of the test cage, approximately 25 cm from the bird's head.

Inside the test cage, a perch was mounted on the floor in front of a small light bulb (the hopper light) and an opening on the floor through which food was accessible when a hopper was raised by activation of a solenoid. A control panel with two microswitch response keys was mounted vertically in front of the perch and the food opening was within a reachable distance for the bird on the perch. The keys were approximately 5 cm apart and each key had an 8 mm light emitting diode (LED) attached. The left LED key is red and designated as the observation key, and the right LED key is green and designated as the report key. See Figure 2 for an image of this operant apparatus.

The experiments were controlled by a PC microcomputer controlling Tucker-Davis Technologies (TDT, Gainesville, FL) System III modules. Stimuli were stored digitally and output via a 2-channel signal processor (TDT, Model RX6) at a sampling rate of 24.4 kHz. Each signal was then output at a mean level of about 70 dB SPL with a 3 dB rove from a separate channel of the D/A converter to a separate digital attenuator (TDT, Model PA5), combined in an analog summer (TDT, SM5) and then amplified (Crown, Model D-75) to a loudspeaker (KEF Model 80V, England) in a sound-attenuated chamber.

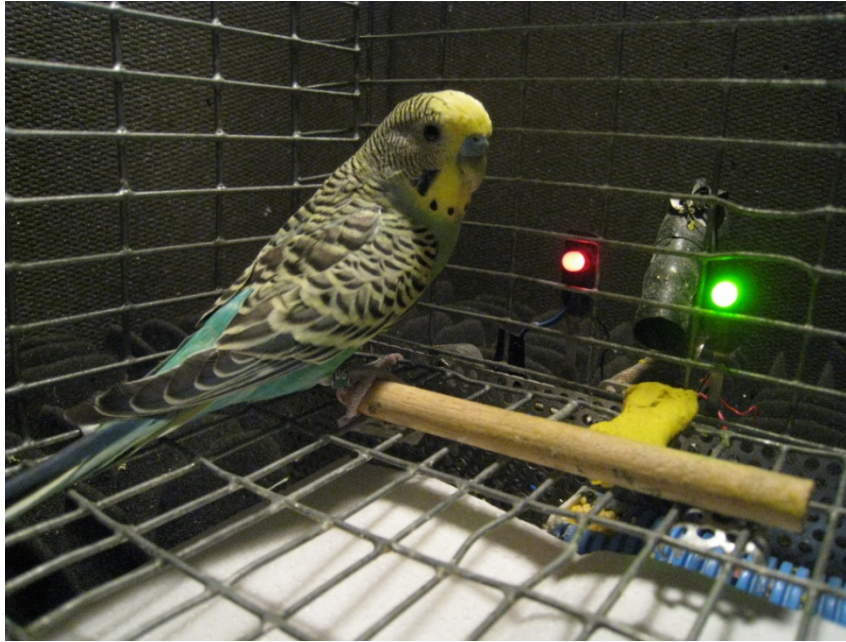


Figure 2: Photograph of a budgerigar in the operant testing apparatus.

Stimulus calibration was performed with a Larson-Davis sound level meter (Model 825, Provo, UT) with a 20-foot extension cable attached to a ½ inch microphone. The microphone was positioned in the place normally occupied by the birds' head during testing. Typically, stimuli were calibrated individually, but for some experiments where a large number of sound files (usually more than 1000) were needed, a different method was used to calibrate more efficiently. First, the amplitude of each stimulus was normalized to a constant rms (root-mean-square) value. Then a 2500 Hz, 150 ms pure tone with the same rms amplitude was created to represent the sound pressure level of the entire stimulus set. The sound pressure level of that tone is therefore defined to be the sound pressure level of the complex stimulus set with the same rms amplitude.



All test sessions were conducted using custom-designed MATLAB software driving the external hardware as described earlier. Data was stored digitally and analyzed using both MATLAB and commercially available statistics software.

## **Procedures**

Ultimately, birds were trained on a task requiring them to discriminate a change in a recurring pattern of sounds. In one case, this background consisted of a particular class of warble sounds. In other cases, the background consisted of natural or artificial warble sequences which could be up to 6 minutes in length. To start, birds first completed a five-phase operant auto-shaping process (Table 1) in order to get used to receiving food reward from the hopper by pecking keys in a specific order.

Once the bird moved through all five phases of the auto-shaping program, the random observation interval was gradually increased from 2 to 6 sec. In other words, once the bird begins pecking the observation (left, red) key, a random interval of 2 to 6 seconds occurred before a target or syllable insertion was presented. Also, the maximum response interval was gradually decreased from 3 to 2 seconds. During this 2-second response interval, the target alternates with the background sound in some experiments, while in other experiments there was only one single target presentation at the start of the response interval. Regardless, if the bird pecked the report (right, green) key within this 2-second response interval, the food hopper was activated for 1.5 seconds and the bird received access to food. This was recorded as a “hit.” If the bird failed to peck the report key within the response interval, it was recorded as a “miss.” If the bird did not peck the report key during sham trials where no target/insertion was presented, it was recorded as a “correct rejection.” Pecks to the report key during sham trials were recorded as “false

alarms” and punished with a blackout period (2 to 10 seconds) during which all of the room lights were turned off and no sound was played back. Any other incorrect report key pecks were also recorded and punished with blackouts. The same trial (or next trial if it is a false alarm) resumed after the blackout period.

Table 1: Auto-shaping training phases.

<b>Training Phase</b>	<b>Behavior Required to Move to Next Phase</b>
Hopper Training	Food hopper is up all the time, allowing free access to food. Here the bird learns to eat seeds out of the hopper.
Phase 1	Observation key LED blinks, a tone plays, and the hopper raises every 40 sec. Bird eventually begins to peck the observation key. After 10 pecks to the observation key, the program advances to the next phase.
Phase 2	Here, the observation key LED is always on, and the bird must peck it to initiate a trial. The bird must peck the observation key 10 times to move to the next phase.
Phase 3	Observation key LED is always on. A peck on the observation key starts a tone and blinking of the report key LED. Bird must first peck the observation key and then peck the report key, when it is blinking, for 10 times to advance to the next stage.
Phase 4	LEDs for both keys are always on. A tone is presented after each observation key-peck. Bird must peck the observation key followed by the report key for 10 times.
Phase 5	LEDs for both keys are always on. Sham trials are introduced and presented randomly. Bird runs until it withholds pecking the report key during sham trials.

During training sessions, background sounds were introduced at lower amplitude (~ 40 dB SPL) at first and then gradually increased to the same level as the target/insertion (70 dB SPL). The number of trials needed for training varies according to different tasks and species, but the stimuli used to train birds were never used in final data collection.

Usually each running session consists of approximately 90 to 120 trials, among which 20 to 30 % are sham trials. Birds were tested twice a day, 5 days a week. To minimize response biases and practice effects, birds ran on different experimental conditions in a random order.

Birds' behavior (hit/miss/correct rejection/false alarm) in each trial was recorded and later pooled together to calculate hit rate and false alarm rate. These two numbers were then used to derive  $d'$ ,

$$d' = z(\text{Hit rate}) - z(\text{False alarm rate})$$

To avoid infinite values, 100% correct and 0% false alarm rates were converted to  $1/(2N)$  and  $1 - 1/(2N)$ , respectively, where  $N$  is the number of trials which the percentage was based on (Macmillan & Creelman, 2005).

$D$  prime has been widely adopted as a measure of a subject's sensitivity in discrimination experiments (Jesteadt, 2005; Macmillan & Creelman, 2005). By taking the difference between the standard  $z$  score of hit rate and that of false alarm rate,  $d'$  equalizes the performance of conservative subjects, whose hit rate and false alarm rate are both low, and liberal subjects, whose hit rate and false alarm rate are both high.

Therefore, the response bias that might exist between individuals and/or species is eliminated, resulting in a direct comparison of perceptual sensitivity. As a general observation during training and most experiments, individuals of the same species respond with a similar strategy – canaries tended to be conservative and zebra finches tended to be liberal, and budgerigars tended to be in the middle..

To evaluate differences in  $d'$  between two conditions, the standard error (square root of the variance) of  $d'$  was calculated and used to construct a 95% confidence interval around the  $d'$  value of each condition. If the two 95% confidence intervals overlapped, there was no significant difference in the sensitivity of these two conditions. If the two intervals did not overlap, the sensitivity in these two conditions differed significantly (Macmillan & Creelman, 2005).

Additionally, response latencies were recorded. A number of studies have shown that response latencies in such a psychoacoustic discrimination task can be reliably used as a measure of stimulus similarity, especially when the differences are subtle and the hit rates are around ceiling. The longer the latency is, the more similar the two stimuli are, and vice versa (Dooling, Brown, et al., 1987; Dooling & Okanoya, 1995b; Dooling, Park, et al., 1987; Okanoya & Dooling, 1988). However, different species may intrinsically react with different speed, which may create bias when using absolute latency to compare across species; therefore a normalization of latency values is used for cross-species comparisons.

### Chapter 3: Acoustic Categories of Budgerigar Warble Elements

Defining acoustically distinctive categories in a vocal communication system is an essential first step in analyzing the repertoire of animals (Deecke & Janik, 2006). It is often achieved by human categorization of sounds or sonograms, with or without the aid of computer (e.g., Armstrong, 1992; Bloomfield, Charrier, & Sturdy, 2004).

Human categorization is a straightforward method that involves several experimenters (raters) sorting piles of spectrograms based on hearing and/or visual inspection of each of them. Because different raters may make different decisions on how acoustic features should be weighed (Jones, Ten Cate, & Bijleveld, 2001), the experimenter usually requires a high inter-rater reliability and *post hoc* multivariate analyses (e.g., discriminant function analysis (DFA), principal components analysis (PCA), and/or cluster analysis) to ensure objectivity and infer the relative weighting of each acoustic feature (e.g., duration, frequency) based on the acoustic measurements of each vocal signal (e.g., Charrier, Bloomfield, & Sturdy, 2004; Kanwal, et al., 1994). However, it is always possible, of course, that the acoustic features or combination of features that animals pay attention are not those the experimenters identified or the animals do not linearly perceive the acoustic features, which violates the assumption of most statistical methods (Deecke & Janik, 2006).

In order to more precisely and efficiently simulate the behavior of human experimenters or the animals, researchers have recently turned to a new computational technique, the artificial neural network, for the classification of animal vocalizations (e.g., Dawson, Charrier, & Sturdy, 2006; Ranjard & Ross, 2008). Briefly, it is a multi-layer, “brain-like” program consisting of many interconnected processing units that

consider the nonlinearities of sound perception for generating reasonable categories. A set of “input units” extract the pattern of activity of the input stimuli and send it to a set of “internal processing units,” where the information is processed through the interconnection between them. Those connections are either excitatory (amplifies the information being sent) or inhibitory (attenuates the information being sent) and allows the neural network to make an appropriate response to the input stimulus. The final result is again presented as a pattern of activity in a set of “output units.”

Typically, a neural network must be trained with a learning procedure before it can process the input information in a non-linearly fashion to categorize sounds appropriately. The weights of the interconnections in the neural network are randomly assigned at the beginning, and a training set of stimuli, of which the “correct response” is already known, is used as input. When the network generates a response, an error is calculated and sent backwards to modify the property of interconnections. After repeating this procedure multiple times, the errors can be reduced. Once this occurs, the neural network is said to have “learned” a specific pattern of connectivity that can then be used to categorize novel signals analogous to the way used to categorize the training set.

Since budgerigar warble has been studied less than the contact calls, the purpose of this chapter was to develop an automatic procedure for reliably categorizing warble elements on the basis of acoustic features to set the stage for perceptual investigations.

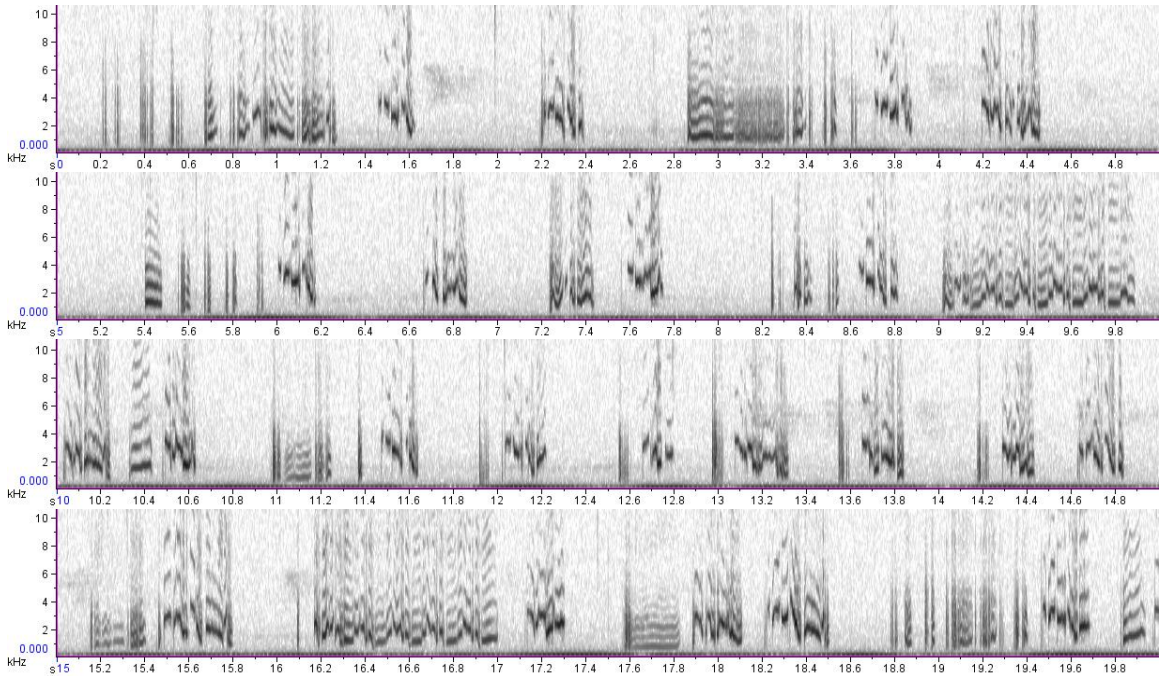


Figure 3: Example of a 20-second piece of warble recorded from a male budgerigar; frequency on the y-axis and time in ms on the x-axis.

### Experiment 1: Acoustic categorization by humans

#### **Background and rationale**

Budgerigar warble is intriguing because of its length and complexity. It contains learned components that are combined in a highly varying pattern (Brockway, 1964b; Farabaugh, et al., 1992). Males are often seen producing warble at different tempos, loudnesses, directions (whether directed to a mate or not), and lengths (ranging from 0.02 to more than 4 minutes) depending on social contexts, as well as the level of arousal and reproductive state of the bird (Brockway, 1964b, 1969).

As with running human speech, the acoustic complexity of warble makes it extremely difficult to analyze and manipulate. More than a decade ago, Farabaugh and

her colleagues (1992) carried out a categorization analysis of budgerigar warble. In that study, human subjects were asked to visually inspect and classify spectrograms of warble syllables, using a human-determined hierarchy where certain acoustic features were more important than others. The result showed that a total of about 2800 syllables could be sorted into 42 groups, among which 15 were basic “elemental” units that were never subdivided while the other 27 groups were compound units where two or more elemental units were combined together. For the purpose of simplicity and clarity in data presentation but not in the analyses, the 15 basic groups were further lumped into narrowband (including contact call-like elements), nonharmonic broadband (including alarm call-like elements), and harmonic broadband sounds.

Unfortunately, analyzing warble ‘by eye’ like Farabuagh, et al. (1992) did limits the amount of warble that can be analyzed in a timely manner. In order to evaluate the importance (i.e., information bearing capability) of higher order acoustic or perceptual aspects of budgerigar warble, very long sequences of warble must be analyzed. The use of neural network methods coupled with human categorization allows us to establish a set of categories that reliably classify more than twenty-five thousand warble elements for later perceptual tests with budgerigars.

## **Method**

As mentioned in the previous chapter, a custom-written MATLAB segmentation program was used to segment the warble recordings from four male budgerigars (Buzz, Puffy, Ricky, and Yuri) into acoustic elements. Vocalizations of three budgerigars were used to develop the automatic classification procedure described below, and warble



elements from the fourth bird were used later to verify that this procedure can be reliably applied to other budgerigars.

Three human raters, experienced with classification of budgerigar vocalizations, were asked to categorize a random subset of 860 warble segments from three budgerigars (283 segments from Puffy; 291 segments from Ricky; 286 segments from Yuri) with the help of another custom-written MATLAB program GROUPER (see below). Raters were allowed to use both auditory (playback) and visual (spectrogram) cues, as well as their past experience with budgerigar vocalizations, to complete the task. Each segment could only be assigned to one group to avoid ambiguity. Additionally, raters were required to give a proper description of each group after classification. Raters were encouraged to ‘clump’ as much as possible, i.e., to reduce the number of groups that contain only one segment.

The MATLAB program “GROUPER” was developed to aide in the classification of warble segments online, instead of printing out thousands of spectrograms. It allows users to load in sounds, play back as needed, show spectrograms on the screen, and assign them to an open-ended number of groups. As a group was established, one segment in that group was chosen randomly and shown on the screen as an exemplar allowing comparisons with other unsorted segments. If the user opened the group folder, all the group members could be seen and heard. Users were able to go back and change their categorization at any time, which means 1) multiple groups could be combined; 2) any group could be divided into multiple groups; 3) users could change the group membership of any single segment at any point. Moreover, the “Advisor” function in GROUPER provided a suggestion as to which group a particular segment may belong to

based on the highest correlation value. Users decided freely whether to accept the suggestion or not.

Table 2: Twenty acoustic measures (see Appendix I) taken by the neural networks and their relative merit in categorizing warble segments.

Measures	Parameters	Relative merit (%)
Spectral roughness	Quality	83
Tonality	Quality	61
Harmonic strength	Frequency	48
1 <sup>st</sup> frequency quartile	Frequency	47
Duration	Temporal	28
Skewness of power	Amplitude	24
Zero-crossing frequency	Frequency	22
3 <sup>rd</sup> frequency quartile	Frequency	18
2 <sup>nd</sup> frequency quartile	Frequency	13
Average peak spacing	Frequency	11
Amplitude modulation	Amplitude	9
No. of harmonic lines	Frequency	7
Frequency of max amplitude	Frequency	6
80% bandwidth	Frequency	5
Entropy	Quality	5
Time to peak amplitude	Temporal	3
Kurtosis of power	Amplitude	3
Frequency modulation	Frequency	2
Standard deviation of power	Amplitude	0
Average power per sample	Amplitude	0

After human categorization using GROUPER, these 860 segments and their corresponding groups were used to train a 3-layered feed-forward neural network (see Appendix II) where 20 acoustic measures (Table 2; see Appendix I) were taken from each segment to nonlinearly simulate human classification. Another 500 segments from each of the three budgerigars were chosen randomly and categorized both by the neural network-based classification program and by human experimenters using GROUPER. The extent to which humans agreed with the program was used as validation of this automatic classification program. The relative merit of each measure was evaluated by the extent of change in grouping before and after any one of the measures was excluded in the neural network. If one measure is crucial to “correctly” group the elements according to human decisions, eliminating it in the program would result in large differences in grouping of the same elements, i.e., more elements would be put into the “wrong” group than before.

Once these procedures were working satisfactorily, all segments from these three budgerigars (7357 segments in Puffy warble, 5633 segments in Ricky warble, and 7204 segments in Yuri warble) and one new budgerigar (6027 segments in Buzz warble) that were not included in the developing and training procedure were categorized, and the categorization of warble was compared among individuals.

## **Results**

Using GROUPER, three raters categorized 860 warble segments into 7 elemental groups and two “special” groups – one contained segments that have a contact call-like element immediately followed by a broadband sound (Group H) and the other included cage noise the bird produced during recording. Raters also provided clear descriptions of

the 7 elemental groups: A) alarm call-like elements, defined as loud, broadband non-harmonic sounds, approximately 100 ms; B) contact call-like elements, defined as narrowband, frequency-modulated tonal sounds, approximately 100-300 ms; C) long harmonic calls, defined as any harmonic sound that is longer than 100 ms; D) short harmonic calls, defined as any harmonic sound that is longer than 100 ms; E) short harmonic calls, defined as any harmonic sound that is shorter than 100 ms; F) “noisy” calls, defined as any broadband sound that sounds noisy (not harmonic) and is approximately shorter than 70 ms; G) clicks, defined as extremely short broadband calls that sounds like clicks; H) pure tone-like elements, defined as calls that show no frequency modulation (Figure 4). A subset of 469 segments was randomly picked out from those 860 segments and showed an inter-rater reliability of 89.3%.

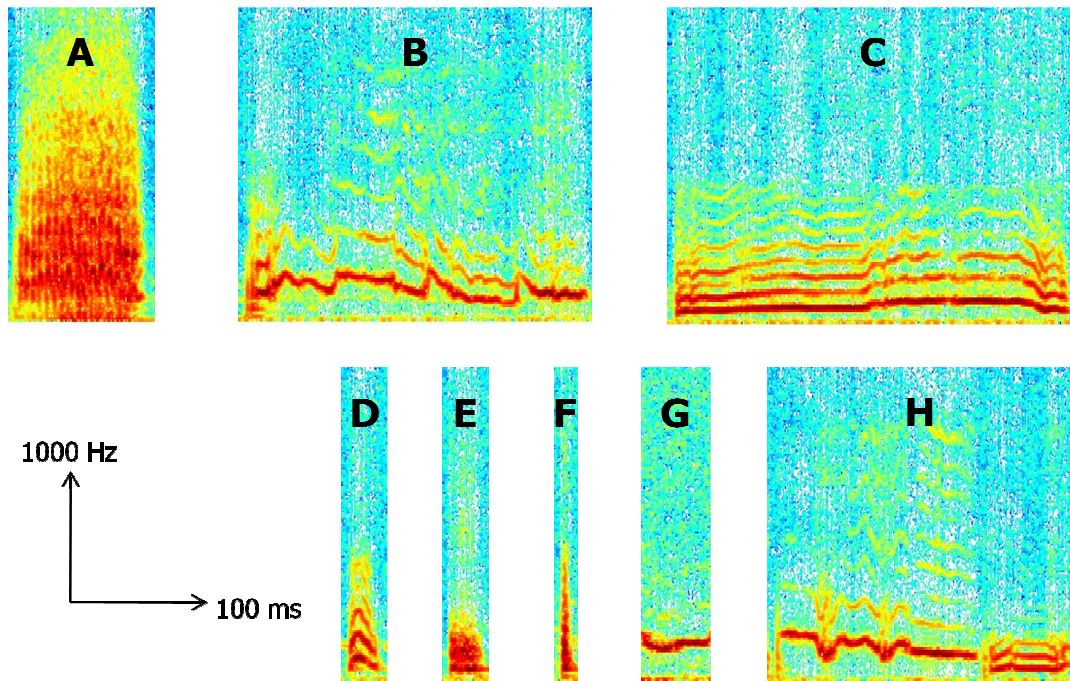


Figure 4: Examples of the seven elemental groups (A – G) and the compound group (H) where a contact call-like element immediately followed by a broadband sound.

Subsequently, a neural network was developed and trained based on these 860 segments of 7 groups and applied to another 500 random segments from each of these three budgerigars and one new individual (Buzz). The same 2000 segments were also categorized by human experimenters, and the average human-program reliability was 83.2% (82.2% for Puffy, 88.2% for Ricky, 81.6% for Yuri, and 80.8% for Buzz).

Table 2 showed the relative merit of each measure in this automatic categorization procedure. Quality parameters such as spectral roughness and tonality were relatively more important than other acoustic features, while amplitude parameters like standard deviation of power and average power per sample were less influential in categorization (Figure 5).

Eventually, all warble segments recorded from the four birds were categorized by the classification program. Eliminating those sounds identified as cage noise, there were 7357 segments in Puffy warble, 5633 segments in Ricky warble, 7204 segments in Yuri warble, and 6027 segments in Buzz warble. Overall, contact call-like elements were the most common segments, comprising 33.22% of warble; pure tone-like elements were the least common segments, only 3.89% of warble.

Further analysis showed that the distribution of all eight categories varied a lot across individuals ( $\chi^2 = 1379.89$ ,  $p < 0.001$ ) (Figure 6). Most of the variations between individuals existed in the noisy group. A closer look showed that Yuri was somewhat unlike the other three birds especially in the amount of clicks and pure tone-like segments produced. Approximately 15% of Buzz, Ricky, and Puffy's warble were clicks, but only about 10% of Yuri warble were clicks. Yuri had less than 1% of pure tone-like segments, but Buzz, Ricky, and Puffy had 6.70%, 4.08%, and 7.01% respectively.

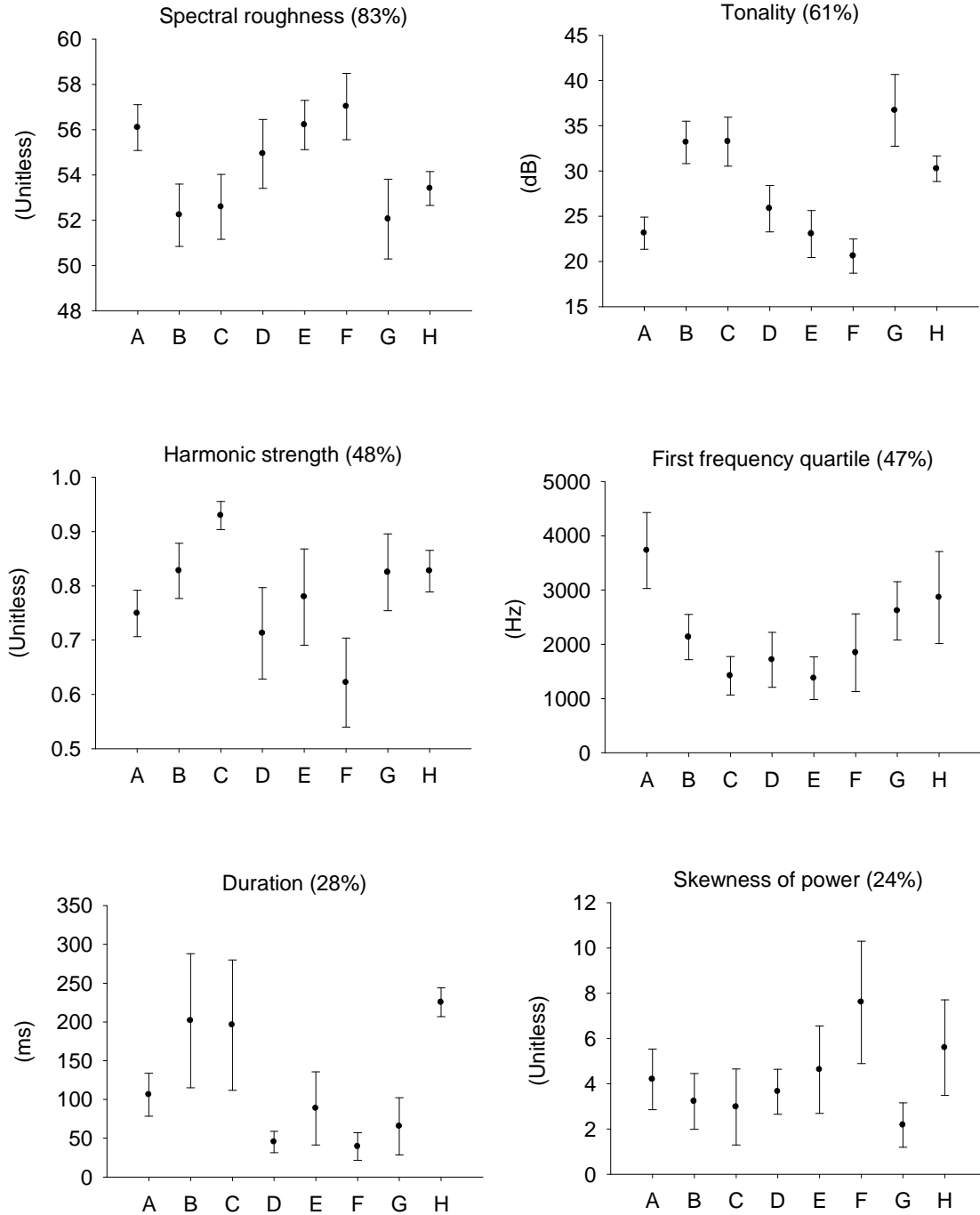


Figure 5: The range of acoustic measures in each group. Six measures were shown here as an example. They have the highest relative merit (shown in parentheses) in simulating human categorization. The x axis indicates different groups. A: Alarm call-like; B: Contact call-like; C: Long harmonic; D: Short harmonic; E: Noisy sounds; F: Clicks; G: Pure tone-like; H: Compound group.

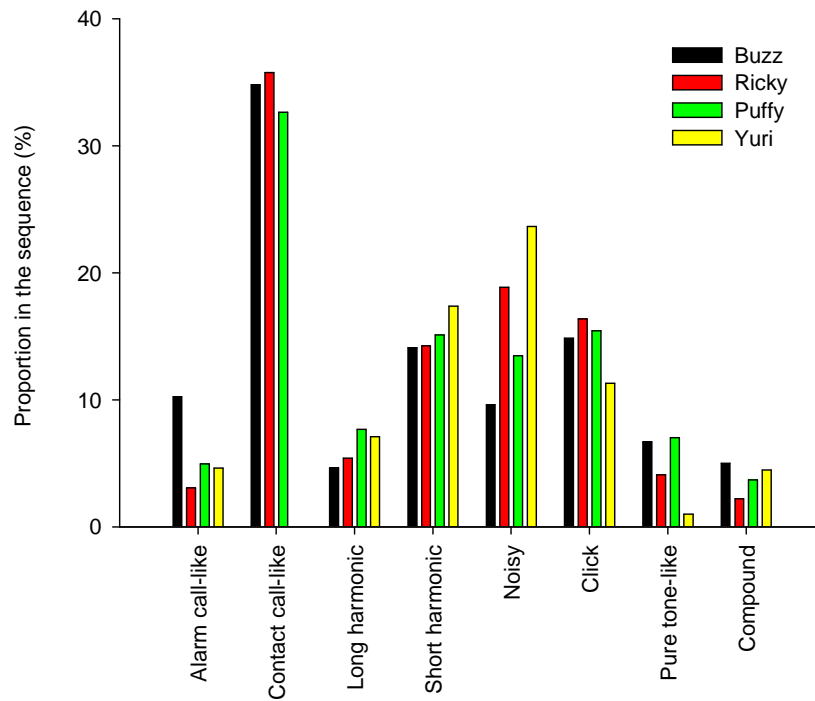


Figure 6: Distribution of the elements in different categories among individuals.

## Discussion

The goal of this experiment was to find acoustic categories that human raters agreed on with high reliability using conventional techniques and to develop a computational process (neural network) that would reliably arrive at the same categories so that large numbers of warble syllables could be automatically classified.

Human raters agreed on eight acoustic categories (seven elemental categories and one compound category), far fewer than the 42 groups, including 15 elemental groups, found in the Farabaugh et al. study. Comparing the two studies, except the pure tone-like elements, all of the other categories here could be found in one or more of Farabaugh's

classes. For example, four out of six nonharmonic broadband classes were lumped together as the short noisy sounds in the current study while the other two stood out as alarm call-like group and clicks. Also, five harmonic broadband classes were re-categorized as long and short harmonic groups now, but all four narrowband classes before were put together as contact call-like group here. However, among the 15 elemental groups in Farabaugh's study, six groups contained less than 2% of the whole warble samples. Since the experimenters in this study were encouraged to combine small groups together to avoid groups that only a few elements, it is reasonable that the current study found fewer groups and more variation within a group. Moreover, the parameters during segmentation were set to minimize the number of compound segments and to have as many elements as possible, so only one compound group (a contact call-like element immediately followed by a broadband sound) was seen in the present analysis. A similar group was also found before (Farabaugh, et al., 1992).

Besides these eight acoustic categories, the existence of the "cage noise" group showed that the segmentation process was not perfect. In other words, the interval threshold might be too high to further separate some compound segments, but the amplitude threshold may be too low to eliminate some noise. On the other hand, setting the interval threshold too low may accidentally combine two individual segments together, and setting the amplitude threshold too high affects the integrity of a segment (low amplitude at the beginning and the end of an element may be cut off). A compromise was reached based on experience and a *post hoc* analysis that showed the settings could be reached where less than 10% in the entire warble set of three



budgerigars were misclassifications. Given the convenience and efficiency of segmentation, this error rate was deemed tolerable.

Overall, the average human-program reliability was 83.2%. This degree of correspondence between human raters and the automatic classification program was considered acceptable. Moreover, an 80.8% human-network correspondence for warble elements from a new bird (Buzz) showed that this neural network based, automatic categorization procedure is also applicable to warble elements of individuals that were not included in the training set.

Acoustic analysis of the elements showed that quality parameters, especially spectral roughness and tonality, weighed more than other acoustic measures when human experimenters (and later the neural network) made a grouping decision (Table 2; Figure 6). This was similar to what Farabaugh *et al.* (1992) used to manually sort warble spectrograms. The first major criterion raters distinguished was the overall bandwidth of energy (broadband or narrowband), which directly related to whether the segment sounds tonal or not. Next, the presence and pattern of harmonic structure and the pattern and range of frequency modulation were used to separate different elements. Furthermore, intensity, which correlates with the amplitude parameters in this study, was least relevant to the classification of warble elements. However, unlike the stepwise procedure using relative importance of each criterion in Farabaugh *et al.* (1992), the neural network technique applied in this study non-linearly connects all the measures and simulates the raters' decision on categorizing warble elements.

The relative proportion of each category is quite consistent across the four birds analyzed (Figure 5). Most of the variations exist in the noisy group, which makes sense

because this group is actually a combination of four nonharmonic broadband groups in the previous study (Farabaugh, et al., 1992). There may be subgroups that have different proportions in the warble of each individual.

Farabaugh *et al.* (1992) showed that cage mates usually share a significant portion of their warble classes, and this may be partially due to vocal imitation of conspecifics. Nevertheless, the similarities between the warble songs of two acoustically isolated groups suggest that some aspects of warble song structure may be common to all budgerigars. Buzz, Ricky, and Puffy were cagemates for more than a month; whereas, Yuri, whose warble was recorded for the Farabaugh *et al.* study, died many years before these birds entered the flock. Although the difference of element distribution between Yuri's warble and others' warble is not significant, it gives a reason for the similar compositions of Buzz's, Ricky's, and Puffy's warble, and the subtle difference in distribution between them and Yuri (Figure 5). Additionally, the relatively small amount of pure tone-like elements in Yuri's warble compared to the other three birds indicates that it may be a newly shared sound within the flock and may explain why this acoustic group did not appear in the previous analysis (Farabaugh, et al., 1992).

One of the goals of the current analysis was to provide a set of universal categories suitable to describe warble from all budgerigars. Thus, in the present study, experimenters were instructed to 'lump' rather than 'split' in categorizing warble elements in order to reduce the number of categories with only a few elements in them. One negative consequence of this strategy is that categories unique to a certain individual may be grouped into other similar categories due to their sparse existence.

## Chapter 4: Perceptual Categories of Budgerigar Warble Elements

In the previous chapter, I established that there are acoustic categories for warble elements that can be reliably identified by both human raters and an automatic classification program based on neural networks. Whether budgerigars have perceptual categories and whether their perceptual categories match those acoustic categories is another question. It is likely that human auditory perception of animal sounds and/or human visual perception of sonograms are fundamentally different from how animals perceive their own species-specific sounds.

There are a number of operant conditioning or playback studies that have asked animals to categorize the vocal signals within their own repertoire. For example, using a habituation/recovery design, researchers found that animals' observed behaviors should be significantly different between categories than within categories (e.g., Fischer, 1998; Searcy, Podos, Peters, & Nowicki, 1995). This method directly answers the question of animals' own perceptual organization of conspecific vocalizations, but the sample repertoires were manageable. As the number of categories increases, the time required to complete the task increases exponentially. This is because every possible stimulus pair needs be tested in order to have a complete matrix of similarity (or dissimilarity).

By strategically using a psychoacoustic discrimination paradigm, the two experiments in this chapter demonstrate how the acoustic categories established in Chapter 3 are perceptually relevant to budgerigars and how discriminable (i.e., salient) acoustic variation within one category versus another is to the bird.

### Experiment 1: Perceptual categorization by budgerigars

#### **Background and rationale**

Eight acoustic categories that describe warble elements across individual budgerigars were determined using a neural network-based automatic classification program trained by human judgments of sonograms. However, while it is experimentally useful that budgerigar warble elements can be reasonably categorized by humans according to a hierarchy of different acoustic features, there is no evidence that the birds categorize their warble elements using the same measures weighed in the same way.

Earlier research shows that budgerigars perceive tone sequences (Dooling, Brown, et al., 1987) and conspecific contact calls (Brown, et al., 1988) differently than humans do. Not only this, but the relative salience of spectral characteristics used for perceptual organization in budgerigars was also dependent on experience with the vocalization (Brown, et al., 1988). Other research suggests that budgerigars are especially sensitive in the 2- to 4-kHz spectral region where their major acoustic communication occurs (Dooling, Brown, et al., 1987). All these studies suggest that budgerigars have a specialized auditory perceptual system for processing species-specific vocal signals that may be different from humans and other bird species.

Thus, establishing how budgerigars perceive warble elements is an essential step. In order to compare warble sequences of different individuals or under different contexts and to furthermore set the foundation for more advanced communicative functions such as “syntax” in human language, individuals must reliably perceive elements as belonging to different categories in spite of considerable acoustic variation within each category. This general definition of the perception of warble sound categories is different from the

more narrow definition of categorical perception familiar from speech work that requires a peak in discrimination performance at a location along a speech sound continuum that falls at the boundary between two speech sound categories.

This experiment tested whether budgerigars categorically perceive the seven acoustic groups (the only compound group was omitted) that have been previously developed by human raters and the automatic classification program (see Chapter 3). If perceptual categories exist in budgerigar warble, we can also gain some insight as to whether these categories are unique to budgerigars by testing humans and other birds on the same stimuli.

Generally, categorization can be inferred from discrimination as long as the subjects have more difficulty discriminating variations among stimuli within the same category than those among stimuli that span two categories (Goldstone, 1994; Horn & Falls, 1996). Most discrimination experiments are focused on making fine distinctions between one target stimulus and one background stimulus that are continuously presented close together in time. However, here we are interested in the relative discriminability within categories versus between categories so a modified discrimination task was adopted as an indication of categorization in the current experiment. Specifically, multiple, but not identical, elements (tokens) were selected from the same acoustic category and played as the background. Likewise, multiple, but not identical, targets from all categories (including the background category) were presented during trials. In order to successfully detect targets that were from different categories than the background sound, the birds had to ignore irrelevant acoustic variation within the background

category and only respond to relevant acoustic variation that reliably differentiated the background category from all other categories.

## **Method**

Four budgerigars (two males and two females), two zebra finches, and two canaries were trained to perform a discrimination task. All stimuli were randomly chosen from the warble segments of Puffy, Ricky, and Yuri that were analyzed in the previous study. None of the subjects had been previously housed with these three birds. In addition, two humans with normal hearing were also tested with the same psychoacoustic method.

The psychophysical methods were described in Chapter 2. In each test session, there were 90 trials (70 test trials where targets were presented and 20 sham trials where no target was presented) among one continuous “background set.” The 70 test trials included 10 targets randomly selected from each of the 7 element categories, making it a “target set.” Three target sets (a total of 210 sounds) were prepared so that each bird (Puffy, Ricky, and Yuri) contributed 10 targets from each category. The background set consisted of 150 elements randomly selected from one element category, evenly drawn from the 3 individuals (i.e., 50 elements from each bird). Since there were 7 categories, 7 background sets were constructed. As a result, each subject bird was tested for 21 sessions (3 target sets \* 7 background sets).

In all, there were a total of 28 possible pairs of group comparisons (7 within-category pairs and 21 between-category pairs). Since all possible combinations of background and target categories were tested and the birds received equal numbers of

background and target categories, it follows that they were not trained to respond to particular category differences over others.

For each possible pair, the subjects' response (hit/miss/false alarm/correct rejection) was recorded as well as their response latency (maximum response interval = 3000 ms if the animal did not respond and the trial was scored as a miss). In order to ensure that any species differences in response bias were not influencing the results, each animal's overall correct percentage and overall false alarm rate were used to calculate a  $d'$  as a measure of discriminability between any two sounds. Standard error (se) was calculated for each  $d'$  and construct a 95% confidence interval around  $d'$  for evaluation of the difference between conditions (see Chapter 2 for more detail).

In addition, response latency was used to construct a seven-by-seven similarity matrix from the final data for each individual. Each cell contained the average latency across all trials ( $n = 30$ ) between two given stimuli. To make every cell value between 0 and 1 as required by the software (0 = completely different; 1 = identical), all latencies were divided by 3000 (maximum response interval). Resulting matrices were analyzed by individual differences scaling (INDSCAL) (SYSTAT 11), which finds a common solution of the perceived relationship of the sounds for each species. The variance in response latencies accounted for by a spatial representation provides a measure of the goodness of fit and can be seen as a "perceptual map" of the stimuli. All data were plotted in a three-dimensional space by SigmaPlot 10.0.

## **Results**

Budgerigars, canaries, zebra finches, and humans were all significantly more sensitive when discriminating two elements from different groups (budgerigar:  $d' = 3.62$ ,

se = 0.07; canary:  $d' = 2.14$ , se = 0.07; zebra finch:  $d' = 2.47$ , se = 0.07; human:  $d' = 3.32$ , se = 0.09) than when discriminating two elements from the same group (budgerigar:  $d' = 0.86$ , se = 0.08; canary:  $d' = 0.31$ , se = 0.11; zebra finch:  $d' = 0.26$ , se = 0.09; human:  $d' = 0.20$ , se = 0.12) (Figure 7). However, there are species differences evident in these perceptual data. Between-group sensitivities are significantly different among all four species. Budgerigars ( $d' = 3.62$ , se = 0.07) were obviously much more sensitive than humans ( $d' = 3.32$ , se = 0.09), zebra finches ( $d' = 2.47$ , se = 0.07), and canaries ( $d' = 2.14$ , se = 0.07) (Figure 7).

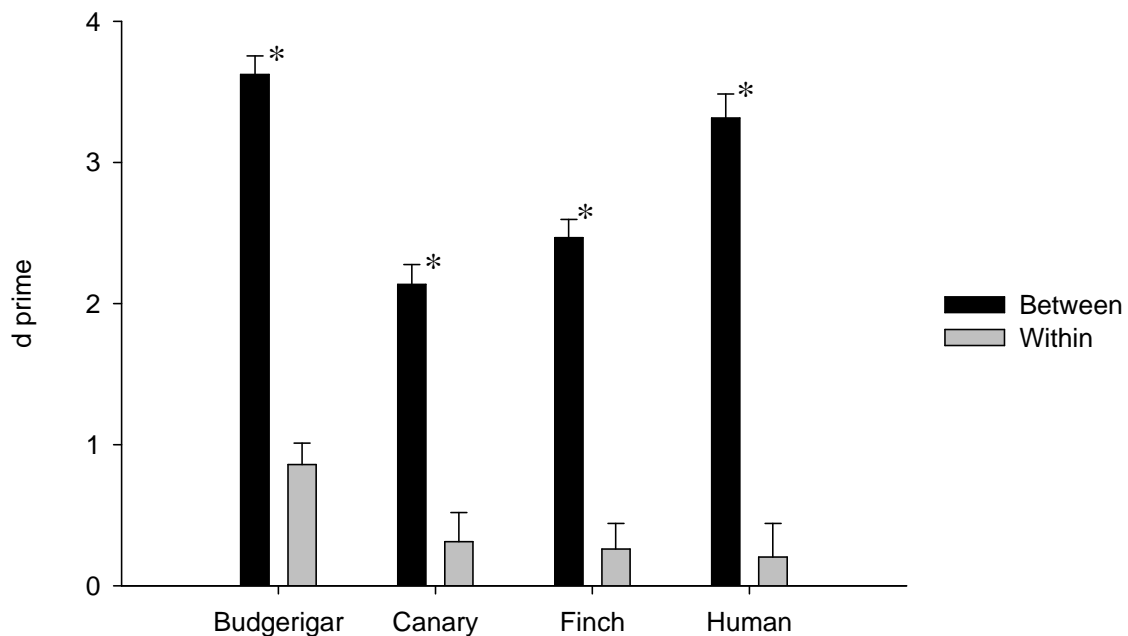


Figure 7: D prime of between-group and within-group discriminations. Error bars indicate 95% confidence interval.



When only within-group comparisons are considered, discriminability is not zero as would be expected if elements in the same group are treated as the same. Moreover, budgerigars are significantly better at detecting subtle variations within the same category than other two species and human (budgerigar:  $d' = 0.86$ ,  $se = 0.08$ ; canary:  $d' = 0.31$ ,  $se = 0.11$ ; zebra finch:  $d' = 0.26$ ,  $se = 0.09$ ; human:  $d' = 0.20$ ,  $se = 0.12$ ) (Figure 7). A close look showed that they are especially good at discriminating variations in contact call-like warble elements (Figure 8). As for other species, within-group discriminability is so low ( $d' < 0.35$ ) that it is difficult to tell whether they are even capable of detecting variations within a certain category.

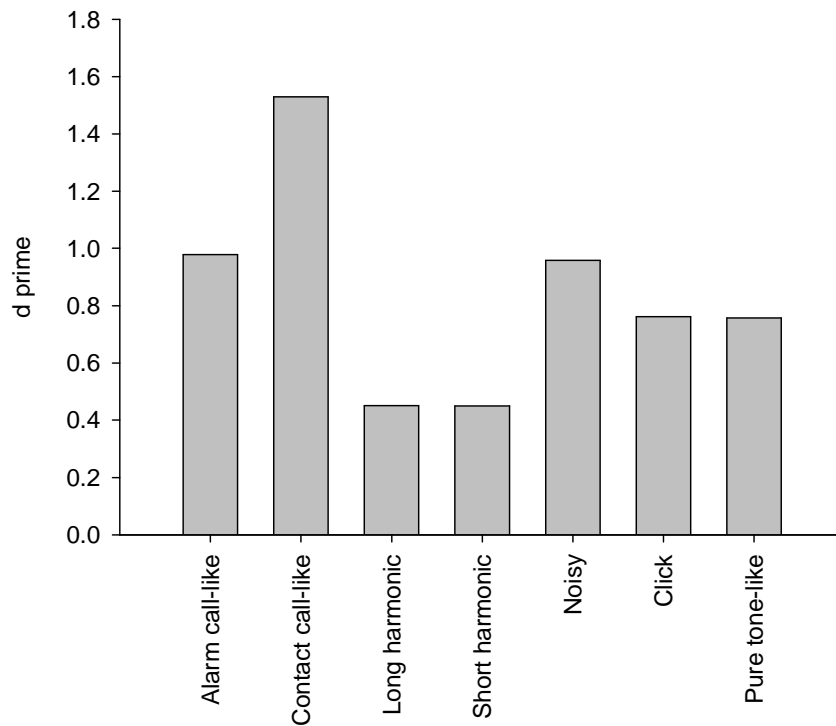


Figure 8: D primes for within-group comparisons for budgerigars within-group comparisons. Data were pooled from all four budgerigars.

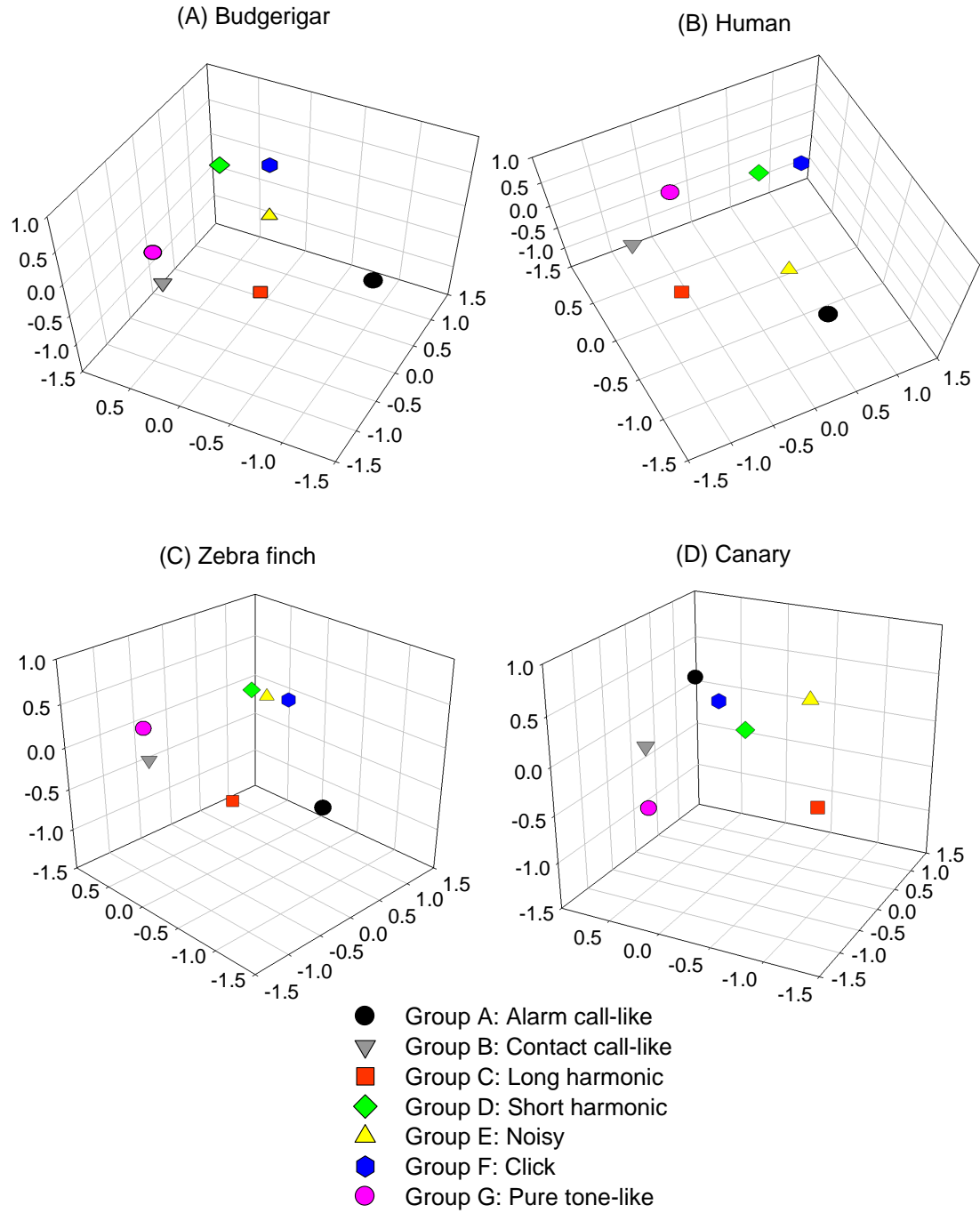


Figure 9: Three-dimensional spatial representation of the result from INDSCAL.

Lastly, each species' perceptual arrangement of these groups can be presented on a three-dimensional space by INDSCAL, where each symbol represents a category of warble elements, and the perceived similarity is represented by the spatial proximity among them (Figure 9). In other words, elements that are perceived by the subject as similar will result in a longer response latency and on the MDS will be represented as closer in perceptual space. Generally speaking, budgerigars perceive their warble elements as seven separate groups as human experimenters have built up, but they are not evenly distributed on the "perceptual map" (Figure 9A). For example, although  $d'$  is still larger than 1, the contact call-like group and the pure tone-like group are relatively close together, so are short harmonic group and click group. On the other hand, the other three groups (alarm call-like, long harmonic, and noisy) are relatively separated in perceptual space. In other species, warble elements are perceived as seven discrete categories, but the perceptual organization of warble elements is quite different. Humans perceive the seven groups relatively evenly (Figure 9B). Zebra finches treat short harmonic calls, noisy calls, and clicks similarly (Figure 9C); canaries put alarm call-like sounds and clicks closer (Figure 9D).

## **Discussion**

This experiment robustly demonstrates that acoustic categories derived by humans in Chapter 3 can also be perceived categorically by budgerigars. The birds' discrimination performance indicates that they can put warble elements into separate perceptual categories similar to the acoustic categories defined by human experimenters. Since it is impossible for budgerigars (or humans) to produce identical utterances of the same element every time, these results clearly show that birds tolerate some variations

within each category yet still treat them as the same. In other words, they are able to focus on acoustically essential differences that are related to grouping and filter out irrelevant features that occurs across warble element or across individuals within a category. This facility probably enables to categorize individual elements of the warble on-line at a fast rate.

Just like in human speech, there is considerable variability in different utterances of the same vowel due to different speakers, different contexts, different rates of speaking, etc., but the identification of those tokens is highly accurate (Hillenbrand, Getty, Clark, & Wheeler, 1995; Pickett, 1999). It has been shown that animals can perceive human speech sounds categorically across talker and genders (e.g., Burdick & Miller, 1975; Dooling & Brown, 1990; Kluender, Diehl, & Killeen, 1987). Here I show that budgerigars form acoustic perceptual categories of their own vocalization.

Among these seven acoustic/perceptual groups, the contact call-like group is a special case. These warble elements are similar to the contact calls produced out of warble as single utterances, which are hypothesized to encode more information than other vocalizations. For instance, budgerigars are able to recognize the identity and gender of the signaler based on a number of spectrotemporal characteristics of the call (Ali, Farabaugh, & Dooling, 1993; Brown, et al., 1988; Dooling, Park, et al., 1987; Park & Dooling, 1985). While there is no parallel study on contact call-like warble elements, the finding that budgerigars, but not zebra finches and canaries, are able to discriminate small variations in the category of contact call-like warble elements suggests that these warble elements may contain species-specific information that other species are not privy to.

In Figure 9, the seven groups are not evenly distributed on the “perceptual map.” For budgerigars, it makes sense that the contact call-like group and the pure tone-like group are relatively close. As discussed above, budgerigars are more sensitive to the differences among contact call-like warble elements, possibly due to the amount of information encoded in them. Pure tone-like elements may simply be “contact calls with less frequency modulation” and be treated as a subgroup in the bigger contact call-like group. On the other hand, the long harmonic group seems to stand alone and away from other elements, perhaps because they are relatively rare and each of them may carry specific information that is perceived differently by budgerigars. In addition, short harmonic group and click group are relatively close. Given that both types of sounds are broadband and very short (~70 ms or shorter), the addition of harmonic structure may be difficult to discriminate and thus no longer salient to the birds for these short sounds. Alternatively, all those short elements may actually be perceived by birds as a loosely-clustered, large group with a lot of variation that encodes complex information. Further research on budgerigars’ temporal resolution is needed to determine whether they can perceive the spectral structure when the sound is short in duration.

Figure 9 also showed that the perceptual categories are species-specific. Although zebra finches and canaries also perceive warble elements categorically, they do not have the same “perceptual map” as budgerigars. It is consistent with early findings on the auditory perception of conspecific and heterospecific contact calls in budgerigars, canaries, and zebra finches. Each species forms perceptual categories corresponding to vocalizations of difference species, and are more efficient at discriminating among calls

of their own species over the calls of the others (Dooling & Brown, 1992; Okanoya & Dooling, 1991).

Besides the differences in basic auditory ability and sensitivity to various acoustic features, it is possible that those species, especially budgerigars, categorize warble elements according to their functions. For example, even though zebra finches do not warble, Group D (short harmonic) and Group E (noisy) sound similar to some of their own vocalizations, which are also noisy and harmonic. This may be the reason why they perceive these two groups of sounds as very similar. Interestingly, the perceptual map of warble elements in humans is relatively equally divided into seven categories, perhaps because the original acoustic categories were built by human experimenters (see Chapter 3). Although the human subjects here were not the same as the raters in Chapter 3, the acoustic features used to categorize warble elements may be similar.

### Experiment 2: Variations within categories

#### **Background and rationale**

In the experiment above, budgerigars responded faster and more correctly when discriminating element pairs between groups than within the same group, indicating that they have perceptual categories similar to the acoustic categories found in Chapter 3. Additionally, budgerigars showed particularly better discriminability of within-group elements than other three species. As with human speech, it is impossible to produce identical utterances of the same phoneme. And also, as with humans, the within category variation that budgerigars can clearly discriminate and perceive in their vocalizations may be communicatively significant.

To further explore how discriminable elements drawn from the same acoustic category are, budgerigars, canaries, finches, and humans were tested psychophysically to discriminate subtle variations within three categories: the alarm call-like group, the contact call-like group, and the short harmonic group which correspond roughly to the nonharmonic broadband sounds, narrowband sounds, and harmonic broadband sounds in Farabaugh *et al.* (1992). The purpose of this experiment was to confirm that budgerigars' response varied with fine details of acoustic stimulus similarity – the higher the similarity, the lower the discriminability – and to determine whether the discriminability of fine acoustic variations is species-specific or not.

## **Method**

Three individuals of each species (budgerigar, zebra finch, and canary) were used as subjects. Some of them were new and not the same as those used in Experiment 1. All stimuli were extracted from Yuri's warble ensuring that no subject had experience with the vocalization before testing.

One segment was randomly selected as a prototype sound from the contact call-like group, the alarm call-like group, and the short harmonic group respectively. Then, seven other elements from the same group that fell within 20% of the duration of the prototype were selected based on their spectrographic cross-correlation value with the prototype (highest, 0.75, 0.70, 0.65, 0.60, 0.55, and 0.50). Here the cross-correlation value was defined as the maximum value obtained by comparing the spectrograms (256-point Hanning window, 50% window overlap) of two vocal signals along all possible temporal offsets, using a MATLAB 2-dimensional cross correlation algorithm (MATLAB function XCORR2). The end result of this procedure was to create three sets

of eight segments, one for each warble element category, which varied systematically along a similarity gradient (Figure 10).

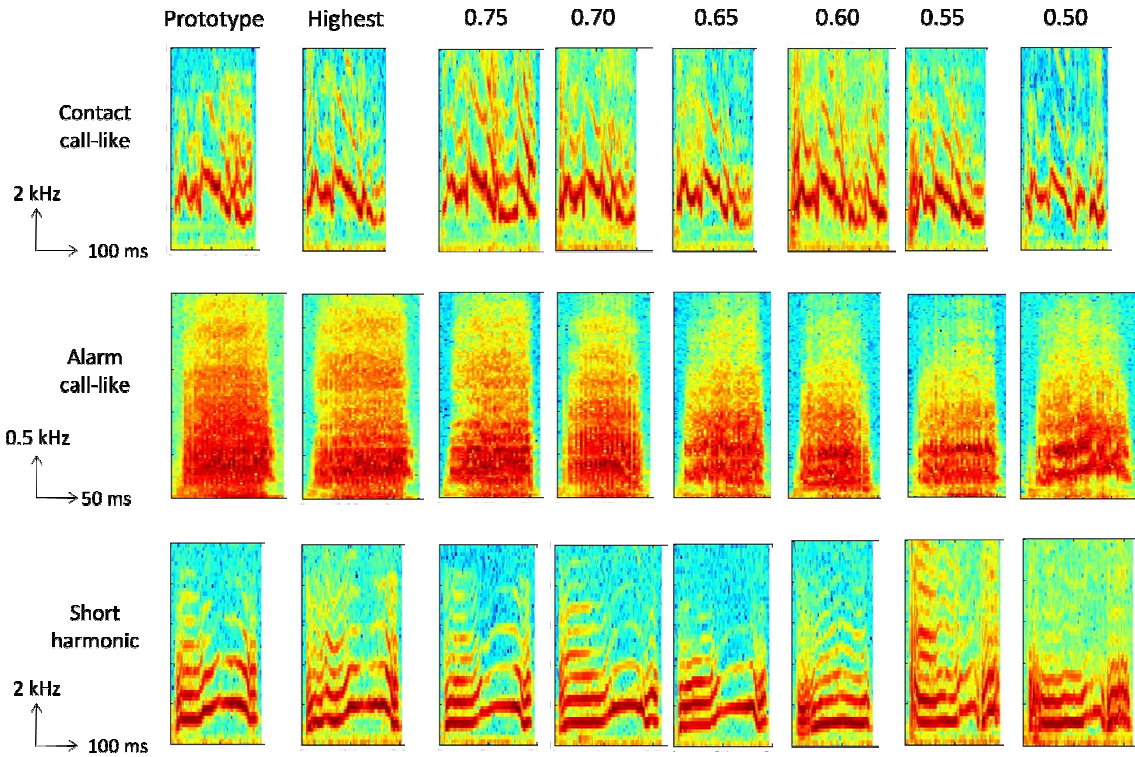


Figure 10: Stimuli used in Experiment 2. Eight elements form a similarity gradient within the same category. Three categories, contact call-like group, alarm call-like group, and short harmonic group, were tested. (Cross-correlation values with the prototype were indicated on top of each column.)

Each element along the acoustic similarity gradient was paired with the prototype, resulting in 7 possible pairs in each element category. Each subject was tested for 40 trials on each pair and the response latencies were recorded. In each element category of each species, response latency of comparisons of each sound along the similarity gradient and the prototype were pooled together and analyzed by ANOVA using the software



SYSTAT 11. A Tukey's *post hoc* test was performed if there was significant difference, but only comparisons of adjacent pairs were reported here since we are more interested in how continuous the birds' perception is relative to continuous acoustic change.

## Results

In Figure 11A, budgerigars showed a gradual increase (no significant difference between adjacent pairs,  $p > 0.05$ ) in response latencies as the cross-correlation increased in the contact call-like group. This means their discriminability declines as the elements became more and more similar. Canaries and zebra finches both showed similar trends as budgerigars, but zebra finches found it significantly more difficult to detect the target when the cross-correlation with the background was higher than 0.75 ( $p < 0.05$ ).

Similar results were found in the short harmonic call group (Figure 11B). All three species' sensitivity gradually increases as element similarity decreases as expected ( $p > 0.05$ ), but zebra finches were significantly less sensitive ( $p < 0.05$ ) to the most similar pair (maximal target-background cross-correlation).

Budgerigars' species-specific discriminability of their warble elements may be the most obvious in the alarm call-like group (Figure 11C). Their responses were considerably different than the other two species. Both zebra finches and canaries showed a sudden, significant increase in the response latency of 0.75 cross-correlation ( $p < 0.05$ ) as if this target was especially difficult to detect. This was not seen in budgerigars, where almost all elements were successfully discriminated with short latencies except the most similar one ( $p < 0.001$ ).

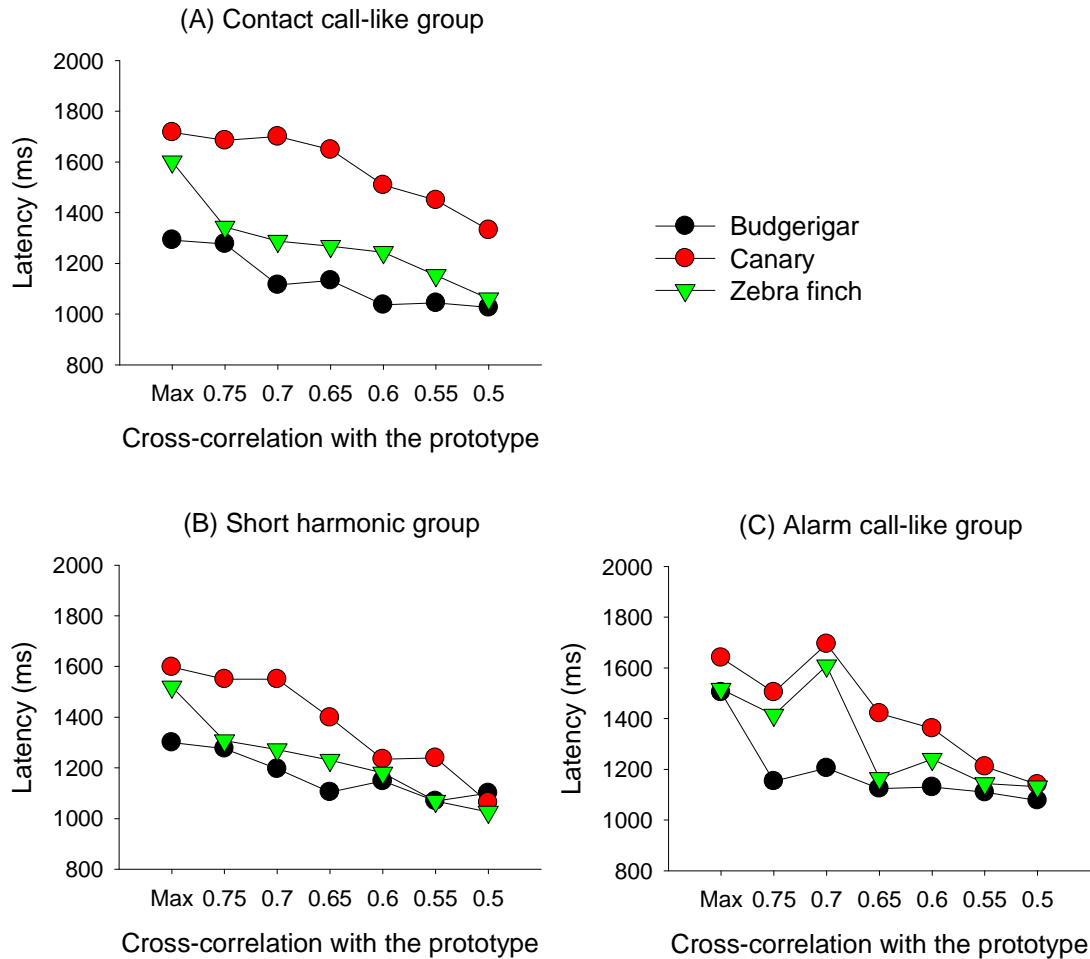


Figure 11: Three different species perceived variations in each of the three acoustic categories differently.

Overall, budgerigars performed better than zebra finches and canaries when discriminating within-category variations (Figure 12). Canaries are generally not very sensitive to subtle variations in budgerigar warble elements, especially not in contact call-like elements. Finches are capable of detecting some variations in every group, but still not as sensitive as budgerigars when the target became extremely similar to the background.

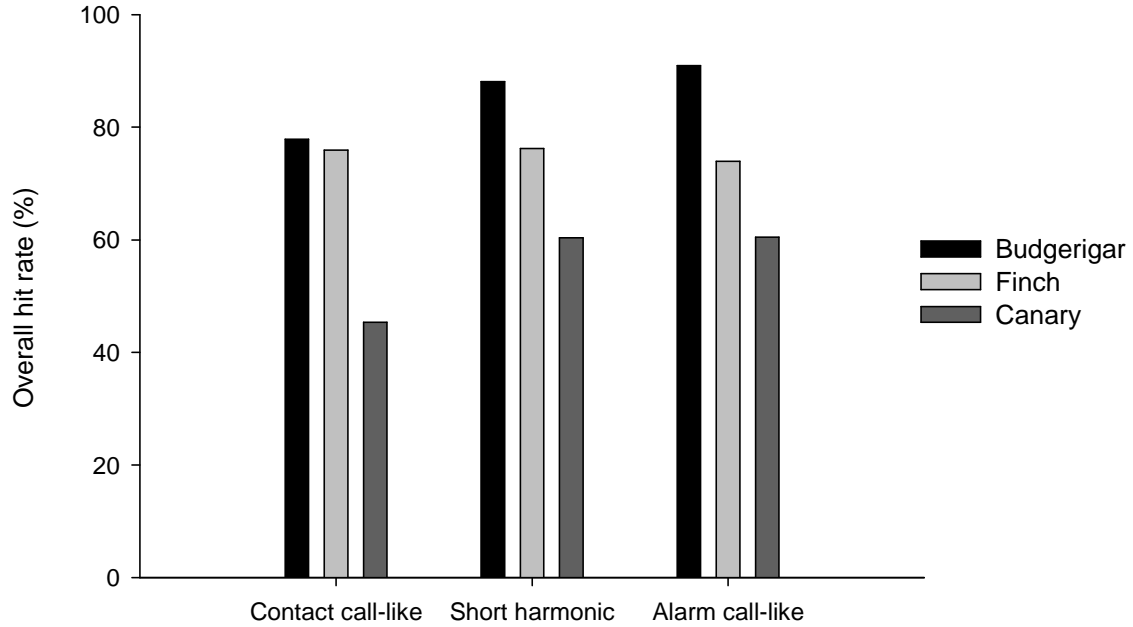


Figure 12: Overall performance of all four species on discriminating within-category variations in alarm call-like, contact call-like, and short harmonic warble elements groups.

## Discussion

The aim of this experiment was to determine budgerigars' ability to discriminate minor acoustic variations within three categories of warble elements. Generally, the more similar the two elements are (the higher the cross-correlation value), the more difficult it is to detect the variation (the longer the response latency), and the higher sensitivity it requires to perform the task. In other words, the relationship between acoustic variations and discriminability is usually linear and continuous.

Earlier experiments comparing budgerigars' perception of contact calls and alarms calls showed that alarm calls form a tighter perceptual category than contact calls

in the MDS plot (Dooling, Park, et al., 1987). Surprisingly, the result here showed that alarm call-like warble elements are clearly perceived as different at a cross-correlation threshold of higher than 0.75 (Figure 11C), while contact call-like warble elements and short harmonic warble elements are perceived linearly and continuously with the level of acoustic difference (Figure 11AB). Budgerigars can discriminate variations in acoustic features, and they can also categorize their own vocalizations by function (Dooling, Park, et al., 1987). It is difficult to know the “meaning” of each single element in warble, but the general functions of contact calls and alarm calls are different. It is possible that contact call-like warble elements and alarm call-like warble elements also have different functions in warble, and that budgerigars use both acoustic cues and functional cues to accomplish the task.

Comparison between budgerigars and other bird species suggest that some of these results are unique to budgerigars (Figure 11). For example, no other species could discriminate alarm call-like elements the same way as and as well as budgerigars do. Since the third element in this stimulus set (cross-correlation = 0.70) was chosen by the same procedure as other elements used in the experiment and did not sound especially different to human experimenters, it is unclear why canaries and zebra finches were especially insensitive to this particular element.

Moreover, while budgerigars and canaries seem to use a “continuous linear discriminator” to detect small, continuous acoustic changes within the contact call-like category and the short harmonic category, zebra finches showed a two-step fashion of perception in these two acoustic categories with a threshold at cross-correlation of 0.75, i.e., two warble elements become difficult to discriminate once their cross-correlation is

higher than 0.75. Even though zebra finches' fine structure discrimination is excellent for their own signals, and even better than budgerigars in detecting harmonic structures (Lohr & Dooling, 1998), they may not be as sensitive to the fine structures in other species' vocalizations.

Overall, budgerigars not only perceptually categorize their warble elements into seven basic acoustic perceptual groups (see Experiment 1 in this chapter), but also have species-specific ability to discriminate acoustic details within each group by the salience of different acoustic features and possibly different functions of sounds (Dooling & Brown, 1992; Okanoya & Dooling, 1991).

## Chapter 5: A Special Case: Contact Calls and Warble Calls

Warble song of budgerigars is composed of a large number of elements uttered in streams. Previous experiments (see Chapter 3 and 4) showed that these warble elements can be acoustically and perceptually categorized into seven basic groups. One particular group, contact call-like warble elements, is particularly interesting. This is the largest group of warble elements, comprising over 30% of the entire warble bout. These calls are about 150 ms in duration, highly frequency modulated, with most energy occurring between 2-4 kHz. They look and sound very similar to contact calls that birds produce as single utterances.

In budgerigars, single contact calls have been extensively studied in many aspects for decades (see review in Farabaugh & Dooling, 1996), not only because they are the most common vocalization found in a flock (Brockway, 1964a), but also because they are very easy to elicit and record under experimental conditions. Each adult individual has at least one or two major call types that it produces the most. As open-ended learners, both male and female budgerigars are able to learn and share contact calls through social interactions (Farabaugh, et al., 1994; Hile, et al., 2000; Hile & Striedter, 2000; Striedter, et al., 2003), and they can perceptually discriminate cagemates and non-cagemates by means of the temporal and spectral cues in their contact calls (Brown, et al., 1988).

Warble song, on the other hand, is considerably less well studied. Evidence has shown that budgerigars deafened as young or raised in acoustic isolation develop highly aberrant warble song (Eda-Fujiwara & Okumura, 1992), showing the need for auditory feedback and perhaps an adult model in order to have normal conspecific songs. Moreover, individuals living together share a higher proportion of similar warble

elements than those living apart (Chapter 3 in this dissertation and Farabaugh, et al., 1992), indicating that vocal learning of warble occurs in adulthood as well. But, most relevant to the present study, is the fact that budgerigars can incorporate a variety of novel environmental sounds into their warble songs (Brockway, 1969; Gramza, 1970).

As mentioned above, a major proportion of budgerigar warble is composed of elements that sound like calls produced as single utterances (contact calls), but are these two vocal signals from different social and vocal contexts different in their acoustic or perceptual aspects? As far as we know, no investigation has ever compared warble calls and contact calls in terms of their acoustic features or perceptual consequences. These are important first steps in trying to tease apart the origin and/or function of these similar vocalizations. Here, using signal processing and psychophysical techniques, a fine grain analysis and comparison of contact calls produced as single utterances (referred as “contact calls”) and contact call-like warble elements (referred as “warble calls”) from the same individuals are reported to further clarify the structure of budgerigar vocal repertoire.

### *Experiment 1: Acoustic analysis*

#### **Vocal stimuli**

Both contact calls and warble songs from four male adult budgerigars (Buzz, Ricky, Puffy, and Cosmo) in our laboratory flock were recorded on 2 to 3 separate days. Warble was recorded, segmented, and categorized by custom-made MATLAB programs as described in Chapter 2 and 3. Approximately 100 warble calls were randomly selected from the contact call-like group of each individual.

To elicit contact calls from the birds, we isolated birds individually in small sound

isolation chambers (Industrial Acoustic Company, model AC1) each fitted with a directional Audio-Technica Carotoid microphone (PRO35A) attached to a Marantz solid state digital recorder (PMD670). After at least one hour, the doors to the chambers were cracked open slightly so the birds could hear the faint calls of their companions in the other boxes. Each bird's vocal behavior was stored on a separate channel of a PCM WAV file at a sampling rate of 48 kHz. The birds were allowed to interact acoustically, but not visually. The recording period was terminated after approximately 100-200 calls were recorded from each bird. Only the male's calls were used in subsequent analyses.

## **Method**

For each of the four individuals recorded (Buzz, Ricky, Puffy, and Cosmo), contact calls and warble calls were compared in two ways. First, multidimensional scaling (MDS) techniques, as described in Chapter 4, were used to see whether the calls in and out of warble grouped together based on spectrographic features. A matrix of correlation values was constructed from all vocalizations produced by each bird and was analyzed using a MATLAB function MDSCALE. The MDS output grouped signals into a 3-dimensional space, where spectrographically similar vocalizations cluster together and dissimilar vocalizations separate.



Table 3: Twenty measurements used in the comparison between contact calls and contact call-like warble elements. All sounds were processed in 5 ms windows, advanced 3 ms at a time (50% overlap in successive windows).

Measure	Description
Peak Frequency (Hz)	Average of peak frequency contour
SD Frequency (Hz)	Standard deviation of frequency contour
Maximum Frequency (Hz)	Maximum value of frequency contour
Minimum Frequency (Hz)	Minimum value of frequency contour
Frequency Range (Hz)	Maximum-minimum of frequency contour
Frequency Change (Hz)	Frequency difference across successive windows
Frequency Modulation (Hz)	Modulation envelope of frequency contour
3dB Bandwidth (Hz)	Frequency bandwidth 3dB down from peak amplitude
Peak Amplitude (dB-Hz)	Average of peak amplitude contour
SD Amplitude (dB-Hz)	Standard deviation of amplitude contour
Maximum Amplitude (dB-Hz)	Maximum value of amplitude contour
Amplitude Range (dB-Hz)	Maximum-minimum of amplitude contour
Amplitude Concentration 1 (%)	% of overall spectrum falling within 2-4 kHz
Amplitude Concentration 2 (%)	% of overall spectrum falling within 2.61-3.11 kHz
dB-RMS	Overall amplitude derived from the RMS of the signal
Amplitude Modulation (Hz)	Modulation envelope of amplitude contour
Duration (ms)	Length of the signal
Wiener Entropy	Unitless measure of disorder (Pure tones = -inf; white noise = 0)
Tonal Quality (%)	% of signal with 3dB bandwidth within 0.3 kHz
Phase Linearity	Unitless measure of deviation from phase linearity (Perfectly in phase = -inf)

Next, a MATLAB-based signal analysis program was used to generate power spectra iteratively across each call in 5 ms windows (with 50% window overlap) and derived 20 different acoustic measurements, including: 1) eight frequency variables, including average peak frequency and 3 dB bandwidth of the spectral peak; 2) eight amplitude variables, including peak amplitude and amplitude modulation; 3) four whole-call measurements, including Wiener entropy (a unitless measure of disorder, see Tchernichovski, Mitra, Lints, & Nottebohm, 2001) (Table 3). A *t* test was performed on each of these measures descriptively using SPSS 16.0 software to reveal any acoustic differences between call categories of the same individual.

Finally, a principal components analysis (PCA) with varimax rotation was performed on 12 of the original 20 measures to sort out the relative salience of the acoustic features budgerigars may use to discriminate call categories. The eight amplitude measurements were removed because the large amplitude differences between warble song (which is produced at a much lower level) and contact calls (which are produced at higher levels) were very obvious.

## **Results**

All birds vocalized readily in both the contact call and warble recording sessions. Each bird produced an average of 130 contact calls (individual birds produced 147, 146, 144, and 82, respectively) and an average of 108 call-like warble elements (116, 109, 101, and 106, respectively). Thus, a total of 519 contact calls and 432 warble elements were used in the following analyses.

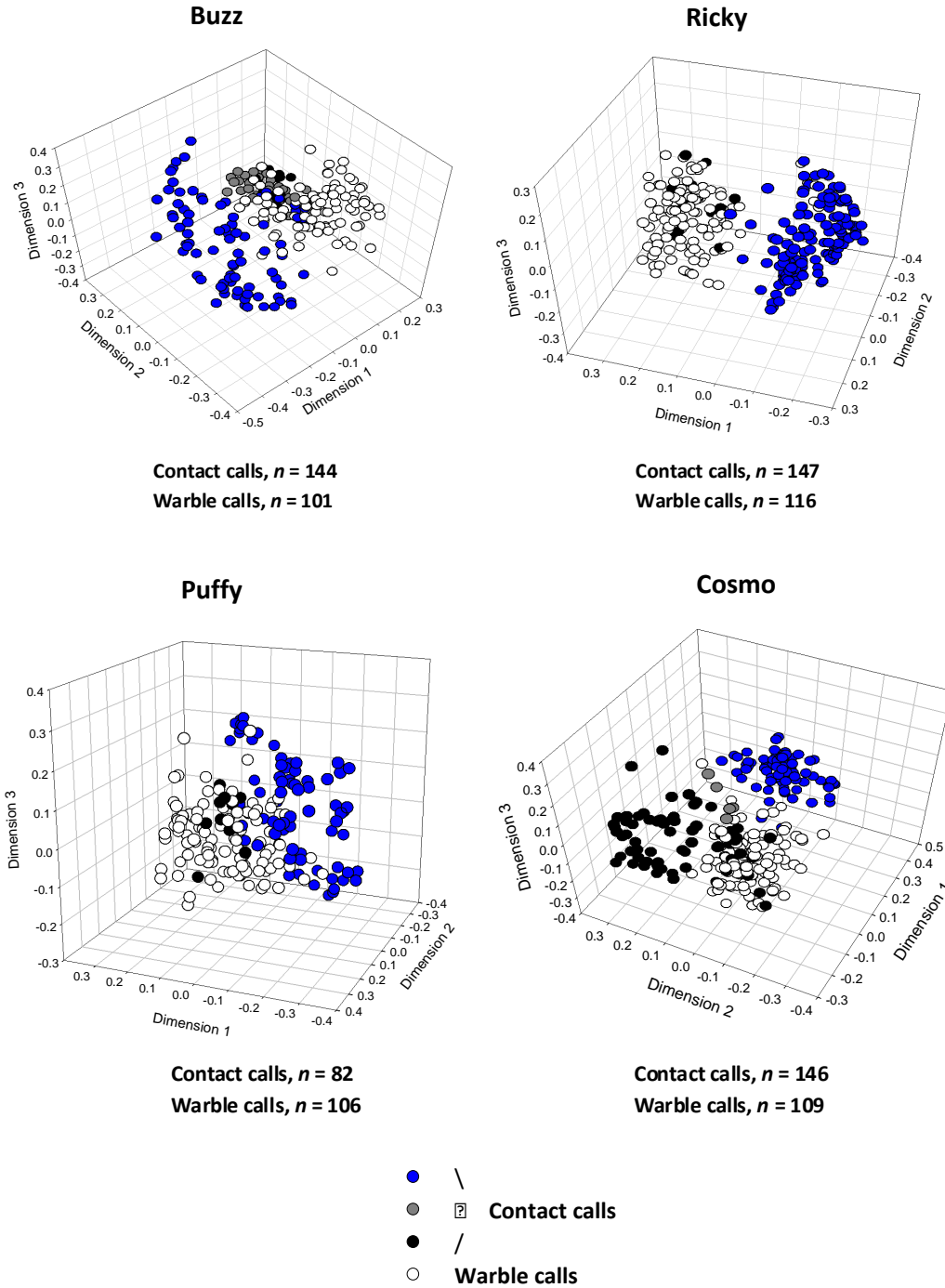


Figure 13: Three-dimensional MDS plots for each budgerigar showing clustering patterns for contact calls and warble calls. Contact calls are shown in color. Each individual may have more than one primary contact call type, indicating by different colors. Warble calls are shown in white. The two call classes are clearly separated in all four birds, showing that the two call groups have distinct spectrotemporal acoustic features.

The three-dimensional MDS plots for each bird are shown in Figure 13. The two call categories (contact calls = colored, each bird may have more than one typical call type; warble calls = white) are clearly separated for each of the four animals in three-dimensional space, showing that the two call groups are spectrotemporally distinct.

These results show that contact calls are significantly different from warble calls on a number of acoustic dimensions (Table 4). Contact calls are generally longer, louder, and have a smaller frequency range than warble calls, but they are also higher in average peak frequency and less frequency-modulated compared to warble calls.

The results of the principal components analysis for the 12 acoustic measures (recall that the 8 amplitude measures were removed) are summarized in Table 5. Four principal components that altogether accounted for approximately 75% of the variation were extracted from the data. The first principal component (PC1) accounted for 34.12% of the variance in the dataset, and the second (PC2), third (PC3), and fourth (PC4) components accounted for 17.22%, 13.17%, and 10.69%, respectively. Measures of frequency (e.g., frequency range, SD frequency) loaded highly on PC1 while PC2 showed high correlations with tonal quality. PC3 was highly correlated with minimal frequency, and PC4 was related to phase linearity. In other words, these four axes mainly presented frequency and “quality” features of the calls.

Table 4: Comparisons of 20 acoustic measures. C = contact calls; W = warble calls. “>” = significantly higher than; “<” = significantly lower than; *n.s.* = no significant difference in *t* tests.

<b>Measure</b>	<b>Buzz (<i>df</i> = 243)</b>	<b>Ricky (<i>df</i> = 261)</b>	<b>Puffy (<i>df</i> = 186)</b>	<b>Cosmo (<i>df</i> = 253)</b>
Peak Frequency	C > W	C > W	C > W	C > W
SD Frequency	C < W	C < W	C < W	C < W
Maximum Frequency	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	C < W
Minimum Frequency	C > W	C > W	C > W	C > W
Frequency Range	C < W	C < W	C < W	C < W
Frequency Change	C < W	C < W	C < W	<i>n.s.</i>
Frequency Modulation	C > W	<i>n.s.</i>	<i>n.s.</i>	C < W
3dB Bandwidth	C > W	C < W	C < W	C > W
Peak Amplitude	C > W	C > W	C > W	C > W
SD Amplitude	C > W	C < W	<i>n.s.</i>	C > W
Maximum Amplitude	C > W	C > W	C > W	C > W
Amplitude Range	C > W	C < W	<i>n.s.</i>	C > W
Amplitude Concentration 1	<i>n.s.</i>	<i>n.s.</i>	C > W	C > W
Amplitude Concentration 2	C > W	C > W	C < W	C > W
dB-RMS	C > W	C > W	C > W	C > W
Amplitude Modulation	<i>n.s.</i>	C > W	<i>n.s.</i>	C > W
Duration	C > W	C > W	<i>n.s.</i>	C > W
Wiener Entropy	C < W	C > W	<i>n.s.</i>	<i>n.s.</i>
Tonal Quality	<i>n.s.</i>	C > W	<i>n.s.</i>	<i>n.s.</i>
Phase Linearity	<i>n.s.</i>	C < W	C < W	C < W

Table 5: Summary of PCA result.

<b>Measure</b>	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>
Peak Frequency	0.055	-0.018	0.815	0.224
SD Frequency	0.855	-0.192	-0.274	0.118
Max Frequency	0.859	-0.008	0.393	0.110
Min Frequency	-0.375	0.226	0.825	-0.160
Frequency Range	0.896	-0.156	-0.249	0.189
Frequency Change	0.829	-0.298	-0.033	-0.127
Frequency Modulation	0.480	0.121	0.466	-0.472
3dB Bandwidth	0.178	-0.902	-0.117	0.110
Duration	-0.083	-0.029	0.218	0.648
Wiener Entropy	-0.133	0.400	-0.097	-0.563
Tonal Quality	-0.162	0.905	0.058	0.007
Phase Linearity	0.226	0.087	-0.206	0.665
Eigenvalue	4.094	2.067	1.580	1.283
Cumulative Variance (%)	34.116	51.337	64.503	75.192

## Discussion

The purpose of the present study was to examine the acoustic characteristics of contact calls and warble calls and determine whether they were acoustically similar to each other. Results show there are considerable acoustic differences between these two groups of vocalizations suggesting that they may represent different phonological systems.

Intuitively, contact calls are louder than warble calls because of the contexts and their social functions. Contact calls are typically produced at very high amplitudes when an individual is isolated from a social group or mate (Farabaugh, et al., 1994; Wyndham, 1980) while warble songs are produced primarily by males and directed toward females at close distances with low amplitudes during courtship behaviors (Brockway, 1964b; Farabaugh, et al., 1992).

In addition, contact calls are longer in duration than warble calls. This may be due to the fast delivery rate of warble song (over 150 elements per minute) that restricts the production space for different warble elements. Likewise, as with human speech, there is a general principle that the greater the number of “subunits” in a unit of speech, the shorter each subunit becomes. This rule can be applied at any level from vowels and consonants as subunits of syllables to words as subunits of sentences (Lindblom, 1963; Pickett, 1999).

There are other differences as well. Warble calls are lower in average peak frequency but have a larger frequency range and higher standard deviation of frequency compared to contact calls. While birds have no comparable articulators like humans that result in “formant patterns,” it still makes sense, from a functional standpoint, that overall warble calls are more variable than contact calls of the same individual. Stereotyped contact calls serve the purpose of individual identification in a large flock (Brown, et al., 1988); whereas, warble is believed to reflect auditory memory of previously heard vocalizations and other environmental sounds (Gramza, 1970). Perhaps, the contact calls are occasionally incorporated into warble in restricted circumstances to increase acoustic complexity and bolster the role of warble in courtship (Brockway, 1964b, 1965, 1969).

There is considerable evidence in songbirds that females prefer males with more complex vocalizations (see review in Searcy & Yasukawa, 1996). Conversely, it could be that contact calls produced as single utterances emerge from warble calls and undergo subsequent modification for distance communication and individual recognition. As in the previous experiment, however, it is one thing to demonstrate acoustic differences in these two categories of vocal signals and another to demonstrate perceptual categories for these signals.

### Experiment 2: Perceptual analysis

#### **Vocal stimuli**

The vocal signals used here were the same as those in acoustic analysis. Four male adult budgerigars were recorded, resulting in 8 “groups” of sounds: Buzz contact calls, Buzz warble calls, Ricky contact calls, Ricky warble calls, Puffy contact calls, Puffy warble calls, Cosmo contact calls, and Cosmo warble calls.

#### **Method**

Four budgerigars, two canaries, and two zebra finches were included as subjects. Once again, multiple elements from the same group were used, and the birds were trained to discriminate them at the level of “groups” instead of comparing them as individual sounds.

There were 100 trials in each running session, where 20 of them were sham trials with no target presented. 10 calls were randomly picked out from each group of the 8 vocal signals, each of which served as target only once, making up the other 80 trials in one session. The background used in each session consisted of 70 random calls from one



group of sounds. Given that there were 8 different groups, each subject had to run 8 sessions to complete the experiment.

The sound level at which the stimuli were played back was normalized and calibrated. Response latencies were recorded (a miss was recorded as 2000 ms, the maximum response latency) and compared by *t* test.

## Results

Figure 14 showed that in general, budgerigars react slower when the background and the target are from the same group. The result was especially obvious when the backgrounds were contact calls and the targets were warble calls. For example, when Buzz's contact calls were in the background, responses to Buzz's warble calls were significantly faster than those to Buzz's contact calls ( $t = 5.53, p < 0.001$ ). Same results were seen when the background was Ricky's contact calls ( $t = 28.06, p < 0.001$ ), Puffy's contact calls ( $t = 41.63, p < 0.001$ ), and Cosmo's contact calls ( $t = 4.87, p < 0.001$ ), respectively. On the other hand, the results were not conclusive when the backgrounds were warble calls and the targets were contact calls. In Buzz's and Puffy's vocalizations, the response latencies were not significantly different between detecting contact calls and detecting warble calls against warble call backgrounds (Buzz:  $t = 0.83, p > 0.05$ ; Puffy:  $t = 0.546, p > 0.05$ ), but in Ricky's and Cosmo's vocalizations, detecting contact calls from a background of warble calls was significantly faster than detecting warble calls from a background of warble calls (Ricky:  $t = 6.78, p < 0.001$ ; Cosmo:  $t = 4.32, p < 0.001$ ).

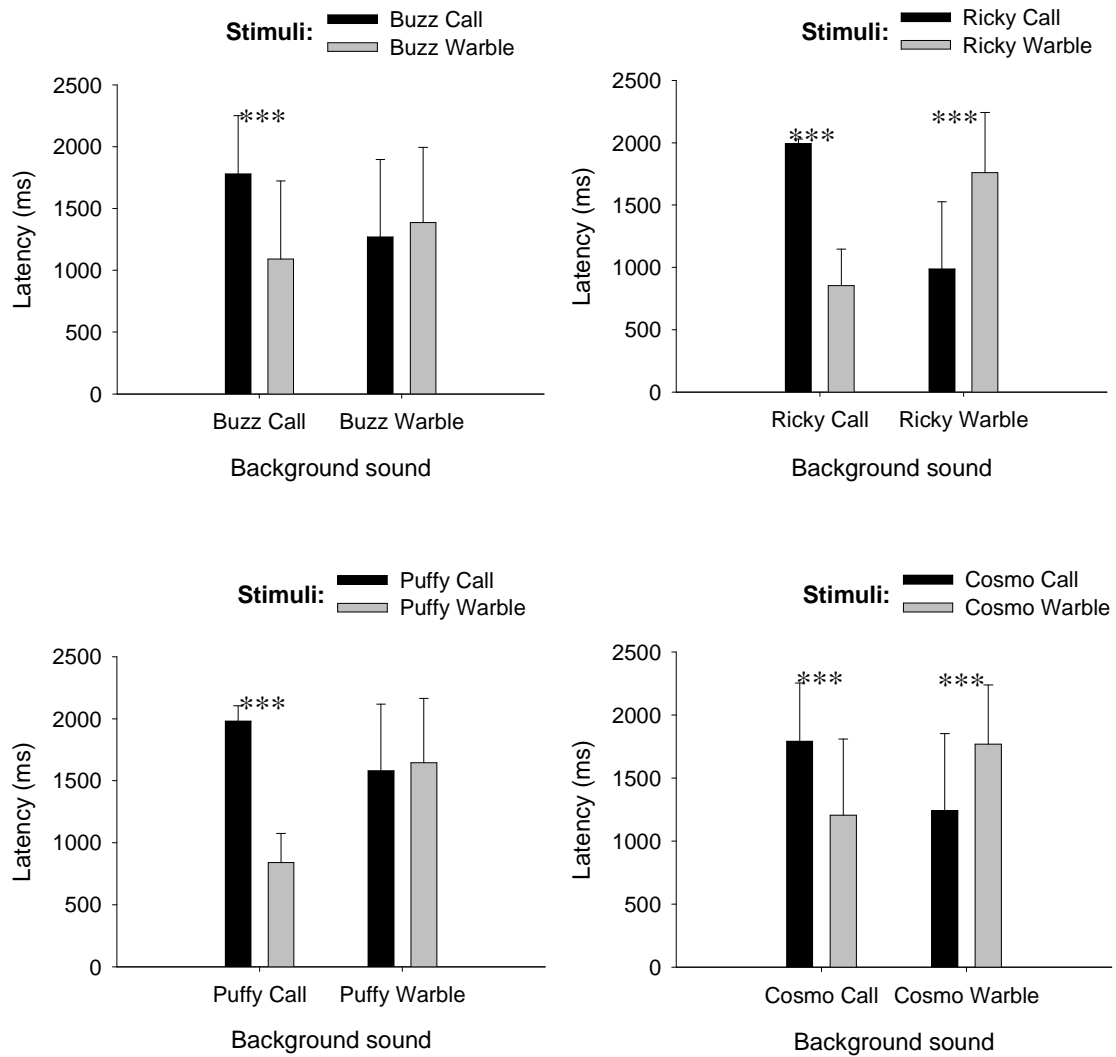


Figure 14: Results of the psychophysical test on calls of each individual budgerigar.

On the other hand, cross-species comparisons showed that canaries and zebra finches were not as sensitive as budgerigars to the difference between single contact calls and warble calls. When contact calls were in the background, both canaries ( $t = 8.95$ ,  $p < 0.001$ ) and zebra finches ( $t = 11.39$ ,  $p < 0.001$ ) showed significantly faster responses to

warble calls than contact calls, but when warble calls were in the background, there was no significant difference between detecting a contact call and detecting a warble call for either canaries ( $t = 0.913, p > 0.05$ ) or zebra finches ( $t = 1.796, p > 0.05$ ) (Figure 15).

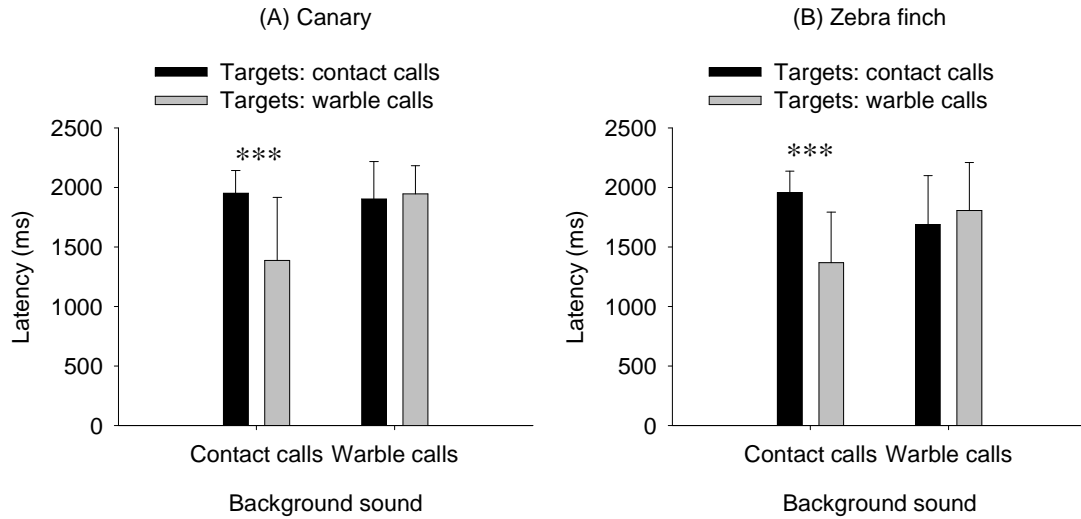


Figure 15: Canaries and zebra finches were less sensitive to the difference between contact calls and warble calls.

A comparison of response latencies collapsed across birds suggests that birds perceive the similarity between contact calls and warble calls produced by the same bird (Figure 16). The latency to response between warble calls and contact calls was significantly longer when both were drawn from the same bird (Buzz:  $F = 8.11, p < 0.001$ ; Ricky:  $F = 0.002, p < 0.01$ ; Puffy:  $F = 0.01, p < 0.05$ ; Cosmo:  $F = 17.03, p < 0.001$ ).

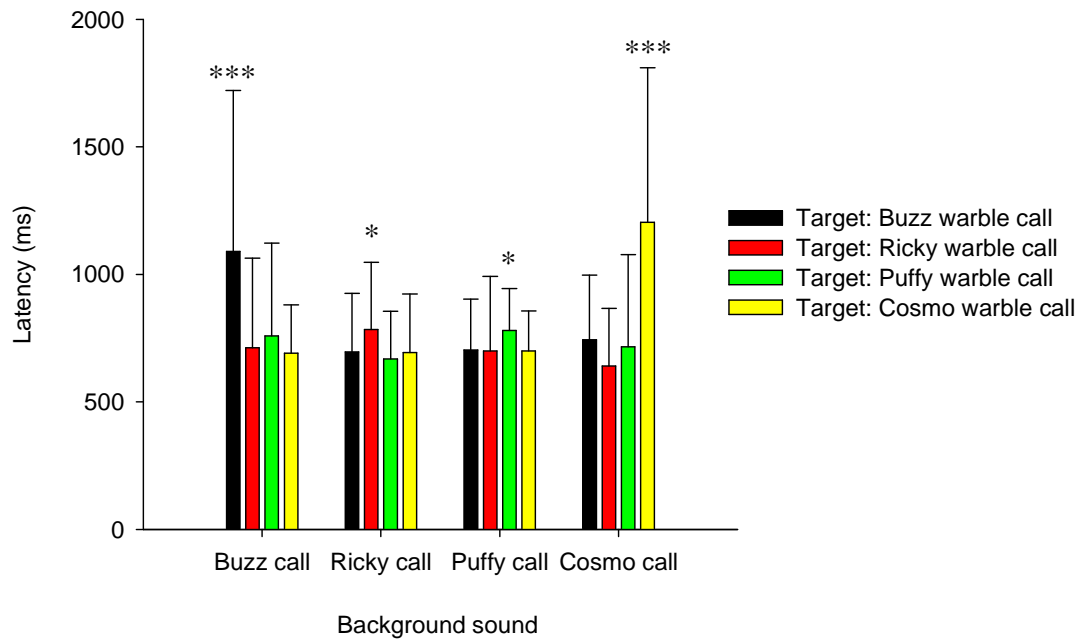


Figure 16: Saliency of individual vocal characteristics.

## Discussion

In aggregate these results suggest that contact calls and warble calls may constitute different phonological systems. The relation between contact calls and warble calls is reminiscent of human speech where single words can be produced alone and can also be strung together in a stream of sentence. Comparisons between words spoken in and out of ongoing speech reveal differences in spectral and temporal features (Lindblom, 1963; Pickett, 1999; Stevens & House, 1963), but these changes do not affect the perception and accuracy of word recognition (Hillenbrand, et al., 1995; Pickett, 1999).

Here I showed that budgerigars form distinctive perceptual groups of contact calls and warble calls, consistent with the result of acoustic analysis in Experiment 1.

However, there are areas of overlap between these two classes of vocalizations. A close look at Figure 13 showed that in Buzz, although the major contact call type (blue) is clearly separated from the contact call-like warble elements (white), two other types of Buzz's contact calls (gray and black) overlap with the contact call-like warble elements on the MDS plot. These two less frequent contact calls may have contributed to the non-significant sensitivity in perceiving Buzz's contact calls and warble calls. Another case was when Puffy warble was the background. Figure 13 showed that Puffy calls are not as clearly separated as other birds, but it could be simply due to a smaller sample size.

Cross-species comparison showed that canaries and zebra finches were not able to discriminate the difference between contact calls and warble calls when the backgrounds were warble calls, but they were able to do so when the backgrounds were contact calls. This is probably because contact calls are less variable in spectrotemporal features (see Experiment 1) than warble calls. Intuitively, detecting variable “oddballs” among a relatively uniform background is easier than detecting relatively constant targets among a very variable background. Nevertheless, even though these two types of calls are acoustically different in many aspects, only budgerigars find these acoustic differences perceptually significant.

All of the above showed that contact calls and warble calls represent different phonological systems both acoustically and perceptually in budgerigars. Moreover, despite the differences between contact calls and warble calls, the evidence suggests budgerigars can also detect the constant features that occur when the same individual produces both calls. It is likely that these “vocal characteristics” of each individual are used in addition to measurable acoustic features to form perceptual categories.

While the developmental process of contact calls and warble calls is still unclear, a full explanation of the differences between these two vocalizations requires longitudinal study with carefully controlled experiments. It could help us determine whether males incorporate old calls (either their own or those of other birds) into their warble, whether new calls emerge from vocal practice in warble, and whether changes in the acoustic structure of warble calls parallel similar changes seen in the normal contact calls during vocal learning. Nevertheless, all these differences indicate that the vocal system of budgerigars is much more complicated than we thought and that the parallels with human speech might be even deeper.

## Chapter 6: Sequential Analysis of Budgerigar Warble

In the previous chapter, I have shown that warble can be decomposed and classified into seven basic acoustic categories by human observers and neural network-based programs. Psychophysical tests further verified that these seven groups coincide with budgerigars' perceptual categories, but they can still hear differences within these acoustic categories and are more sensitive to these differences than other species of birds.

As stated previously, warble is a long, rambling vocalization that is extremely important in coordinating reproductive behavior in budgerigars (Brockway, 1964b, 1965, 1969). The fact that budgerigars perceive warble elements in discrete categories raises the possibility that additional information may be encoded in this vocalization. In other words, besides the characteristics of each individual element, do budgerigars extract useful information from the proportion or distribution properties of each perceptual category in the entire warble bout? More intriguingly, can budgerigars make use of the order of elements in a warble sequence to code important information? That is, if budgerigars are sensitive to the serial order of component elements, is altering this sequence perceptually relevant, and does it lead to different behavior?

The main purpose of this chapter is to examine whether there is any underlying structure in warble song – whether there is any “rule” that governs the ordering of warble elements. If it can be shown that warble is not simply a random delivery of elements, we can ask questions about the perceptual significance of changes in sequential order, or the distribution of different elements produced, or other aspects of the warble sequence that might code information.

### Experiment 1: Acoustic analysis of warble sequences

#### **Background and rationale**

For any sequential data, the ordering of its components can be either independent (i.e., randomly arranged) or relative to each other so that transitions from one element to the next may be predictable. One can imagine a sort of continuum where at one end elements are organized arbitrarily so there is no way for the receiver to predict which element will occur next. At the other end of the continuum, there is a so-called “rigid syntax” – the sequence is perfectly predictable without uncertainty. In the latter case stereotyped song patterns can be found, which implies that there is no information encoded in the sequence (Chatfield & Lemon, 1970; Gottman & Roy, 1990). The vocal sequences of most oscine passerines and human speech fall at somewhere in between these two extremes. In other words, they have “flexible syntax” (Leger, 2005). That is, the animal has some choice and control over which element to produce next. From one perspective, this creates uncertainty in the ordering of song elements. Compared to a purely random sequence, this creates some aspect of determination and predictability in the order of elements.

In budgerigar warble, elements from the eight acoustic/perceptual categories are strung together to form this complex song. However, the sequential organization of warble elements in natural song has never been analyzed. At first glance, it is unlikely that budgerigar warble can be described as having a rigid syntax. It appears highly variable and there is no obviously observable regular pattern of element combination. But it is also unlikely that warble elements are totally unpredictable. The demonstrated importance of warble in budgerigar reproduction strongly suggests that critical



information is conveyed between mates. Information could be transferred in several forms with the sequential order of warble elements a likely possibility given that warble elements are perceived categorically (see Chapter 4) and strung together as long sequences. If it is true, there may be certain degree of predictability in the arrangement of warble elements.

Therefore, the purpose of this experiment is to determine whether there is statistical evidence that the elements in budgerigar warble song are produced in a non-random sequence. Next, Markov chain model and information theory are used to determine “how predictable” the sequence is and what the implications for this kind of organization are.

## **Method**

Four budgerigars whose warble was used in Chapter 3 and 4 were used here. Recall that the entire set of warble recordings for each bird were segmented and identified by element type and location in the sequence.

Budgerigars can sing continuously for more than 10 minutes, and their warble does not have a definitive beginning or end, making it difficult to determine one single “song” like those of songbirds. Thus, for the propose of this analysis, one second is arbitrarily assigned as the cutoff interval between two “songs.” In other words, a song is defined as a sequence of warble elements separated by 1 second or longer silence interval. The resulting sequences are all of different lengths in terms of both the physical time elapsed and the number of elements in the sequence. As a general rule, when one applies a Markov model, it is advisable not to combine sequences of different lengths in the same data set for analysis (Chatfield & Lemon, 1970; Gottman & Roy, 1990).

Therefore, for each of the four birds in this study, the longest sequence generated under natural conditions was used plus four other sequences of more than 100 elements. These sequences were all analyzed separately.

The Markov chain is a relatively easy and straightforward model frequently used in analyzing bird songs (Dobson & Lemon, 1979; Gentner & Hulse, 1998; Lemon & Chatfield, 1971; Martindale, 1980). It is a mathematical model in which the probability of occurrence of an element in a sequence depends on the  $r$  immediately preceding element(s), which we call it an  $r^{\text{th}}$  order Markov chain (Chatfield & Lemon, 1970; Gentner & Hulse, 1998; Gottman & Roy, 1990). This assumption is used to generate some expected values ( $E_i$ ) to compare with the observed values ( $O_i$ ) derived from our data using a likelihood ratio chi-square ( $LR\chi^2$ ) test (Gottman & Roy, 1990):

$$LR\chi^2 = 2 \sum O_i \log_e(O_i/E_i)$$

If it is not statistically significant, the original model is not rejected. If it is statistically significant, our data do not fit the model and some other alternatives should be considered.

In practice, the first step is to assume that adjacent antecedent and consequent elements are independent such that

$$\hat{e}_{ij} = n_{i+} n_{+j} / N_2$$

where  $n_{i+} = \sum_j n_{ij}$  = number of pairs which start with element  $i$

$n_{+j} = \sum_i n_{ij}$  = number of pairs which end with element  $j$

$$N_2 = \sum_{i,j} n_{ij} = \text{total number of pairs}$$

and  $LR\chi^2$  approximates a chi square distribution with  $(c - 1)^2$  degrees of freedom, where  $c$  is the number of categories ( $c = 8$  in this study).

As the order increases, the degrees of freedom become so large that  $LR\chi^2$  no longer approximates a chi square distribution. A better way to analyze data in higher order is to use information theory. As the sequence gets longer, more information is captured and uncertainty decreases. By calculating the amount of information gained (or the amount of uncertainty lost) every time one “additional” element is included into the past, the order of Markov chain can be inferred (Chatfield & Lemon, 1970; Gentner & Hulse, 1998; Gottman & Roy, 1990). Using the Shannon measure of information (Chatfield & Lemon, 1970), the conditional uncertainty of an element given the preceding element is

$$H_2 = - \sum_{i,j} P(i,j) \log_2 P(j|i)$$

The subscript 2 means only 2 elements (pairs) are considered in this equation.

Using the same idea, we calculated  $H_0$  ( $H_0 = \log_2 c$ ),  $H_1$ ,  $H_2 \dots$  until  $H_7$ . It is a decreasing series of number that measures the conditional uncertainty for each order of dependence. In this case, a graphical presentation of those numbers (plot  $H_i$  as a function of  $i$ ) and visual inspection is more convenient and reliable than a series of statistical tests (Chatfield & Lemon, 1970; Gottman & Roy, 1990). The point where  $H_i$  starts to decrease relatively slowly is thought to be the order of dependency.

## Results

Figure 17 showed the distribution of eight categories in the longest (most elements) sequence of Buzz ( $n = 409$ ), Ricky ( $n = 389$ ), Puffy ( $n = 346$ ), and Yuri ( $n = 353$ ). Clearly, the element categories are not evenly distributed in warble, and among individuals. Most variation among birds occurs in the distribution of the noisy elements. This result is consistent with what was found in the previous chapter.

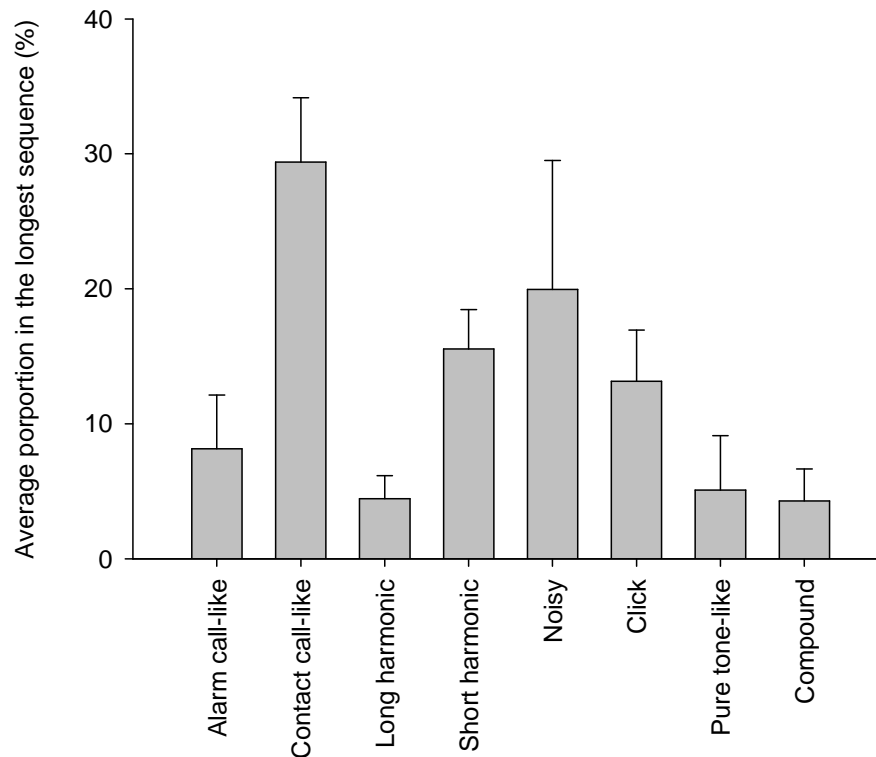


Figure 17: Distribution of different categories in the longest sequence of each bird.

To test the data for independence, the transition matrix for the longest sequence produced by each individual is presented in Table 6 as an example since other sequences all followed the same pattern. Chi square tests indicate the sequencing of elements in warble is not independent. This is true of all four birds.

Table 6: Transition matrix of the longest sequence in each budgerigar.  $LR\chi^2$  is compared with critical  $\chi^2(49) = 66.34$ .

Buzz

		<i>Following category</i>							
<i>Preceding category</i>		<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
	<b>A</b>	8	30	0	4	4	3	3	1
	<b>B</b>	35	38	4	13	7	27	17	3
	<b>C</b>	1	2	0	1	2	0	1	2
	<b>D</b>	2	17	1	30	4	4	7	2
	<b>E</b>	1	11	1	6	5	2	2	1
	<b>F</b>	1	27	1	6	2	8	6	0
	<b>G</b>	5	14	1	7	4	6	7	1
	<b>H</b>	0	6	1	0	1	0	2	0

$LR\chi^2 = 124.74$

Ricky

		<i>Following category</i>							
<i>Preceding category</i>		<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
	<b>A</b>	0	11	1	0	2	0	0	0
	<b>B</b>	12	22	3	11	23	20	4	2
	<b>C</b>	1	5	2	0	2	4	2	0
	<b>D</b>	0	8	2	34	15	9	1	0
	<b>E</b>	0	24	5	15	49	9	5	5
	<b>F</b>	0	18	3	8	15	12	2	0
	<b>G</b>	1	3	0	0	6	4	0	0
	<b>H</b>	0	6	0	1	0	0	0	1

$LR\chi^2 = 158.51$

Puffy

		<i>Following category</i>							
<i>Preceding category</i>		<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
	<b>A</b>	5	8	1	0	3	3	0	3
	<b>B</b>	7	34	7	6	11	12	6	8
	<b>C</b>	2	5	4	1	2	0	1	3
	<b>D</b>	1	6	1	27	15	6	1	0
	<b>E</b>	4	7	4	14	17	16	1	1
	<b>F</b>	2	13	1	5	16	21	1	0
	<b>G</b>	2	6	0	3	0	0	1	1
	<b>H</b>	0	11	0	1	1	1	2	4

$LR\chi^2 = 174.02$

Yuri

		<i>Following category</i>							
<i>Preceding category</i>		<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
	<b>A</b>	5	8	4	4	9	1	1	1
	<b>B</b>	16	30	11	6	30	6	3	7
	<b>C</b>	3	14	1	1	1	1	0	1
	<b>D</b>	1	6	1	11	12	5	2	2
	<b>E</b>	3	36	3	11	17	10	0	8
	<b>F</b>	1	4	1	5	13	3	1	1
	<b>G</b>	1	1	0	2	2	1	0	0
	<b>H</b>	3	10	1	0	4	2	0	4

$LR\chi^2 = 89.37$

For example, in all four birds, group B sound (contact call-like elements) usually repeats itself several times before switching to other categories. This is reasonable since group B has the highest proportion in warble. Group D showed similar repeating feature in Buzz, Ricky, and Puffy, but not so obviously in Yuri. Some interesting pairs can also be detected in the transition matrices. For instance, Yuri and Ricky seem to have a “B → E → B → E” pattern, while Buzz has a “A → B → A → B” cycle. Although Puffy’s transition matrix does not show any particular combination, it is clearly not random either (e.g., group D was almost exclusively followed by another group D sound or a group E sound).

The conclusion that warble elements are not independently ordered drove us to keep looking for a higher-order model. However, when considering third-order dependency, eight categories make  $8^3 = 512$  possible triplets, and many of them never occur even when the observed sequence is quite long. The degrees of freedom rise to  $8 * (8 - 1)^2 = 392$ , and the  $\chi^2$  approximation is invalid. Therefore, H values based on information theory were used in the following analysis.

Figure 18 shows the decreasing pattern of conditional uncertainty. Maximum uncertainty is 3 bits ( $\log_2 8$ ) for all four subjects because every sequence used here contains all 8 categories of elements. The general pattern of these four declining lines is consistent. There is a large drop starting from the second order to the 5<sup>th</sup> order, and the line begins to level off afterwards. It suggests that of the total amount of information conveyed in a warble sequence, most is covered in the 5<sup>th</sup> order transitions between element categories. This means that budgerigar warble follows at least a 5<sup>th</sup> order Markov

chain model. That is, there are significant local structures over approximately 5 contiguous elements throughout the total length of warble.

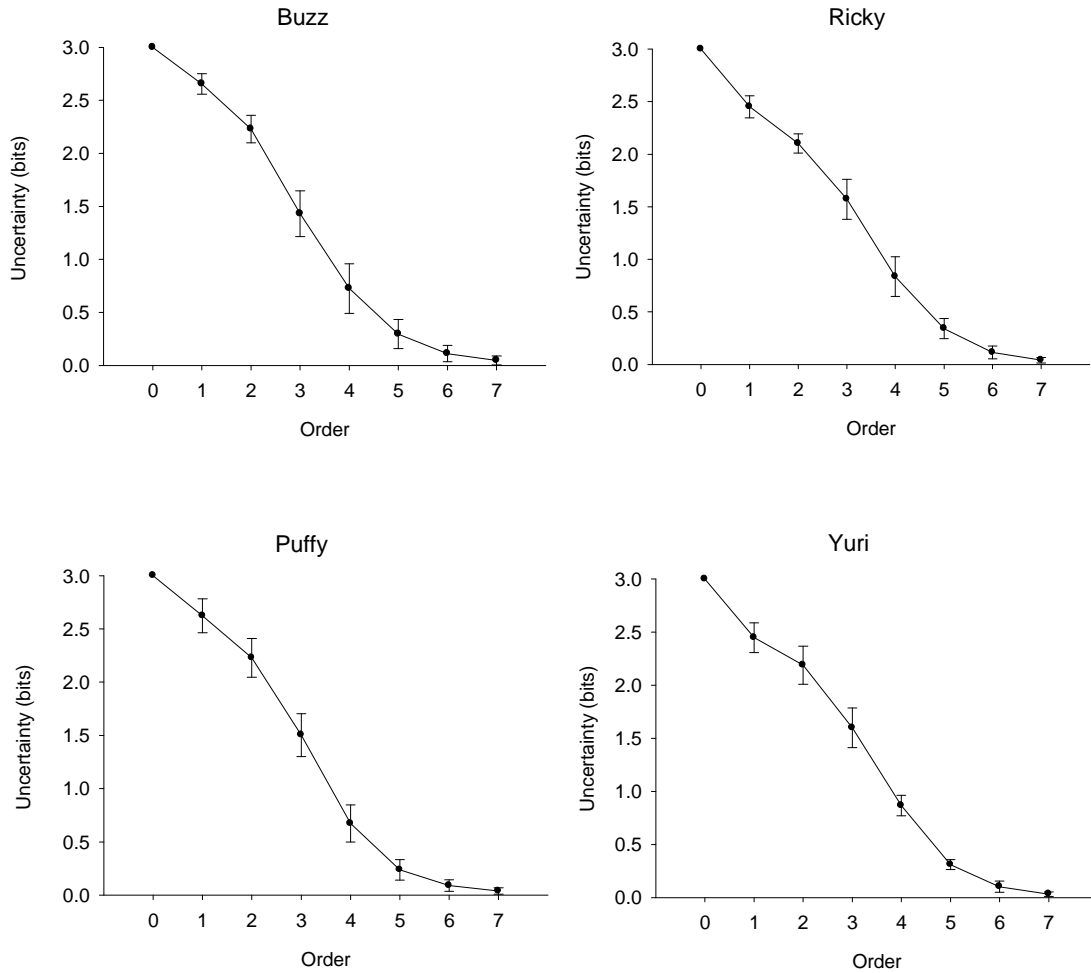


Figure 18: Conditional uncertainty plotted as a function of the order of Markov chain for each budgerigar. The line connects the mean of 5 sequences, and the error bars indicate the standard deviations.

## Discussion

The Markov chain is frequently used in modeling bird vocalization (e.g., cardinals (Lemon & Chatfield, 1971), American thrushes (Dobson & Lemon, 1979), solitary vireos (Martindale, 1980), chaffinches (Slater, 1983), American redstarts (Lemon, Dobson, & Clifton, 1993), and European starlings (Gentner & Hulse, 1998)) because its basic concept is easy to understand and its application in biology is straightforward. The Markov model is particularly suited to the analysis of budgerigar warble for several reasons. First, only a small number of rules are sufficient to create a fair amount of variation (Dobson & Lemon, 1979), making it relatively easy to apply to the complex warble song. Second, budgerigars are open-ended learners that have life-long ability to learn and incorporate both biological and non-biological novel sounds into their warble (Gramza, 1970). The Markov model may be especially appropriate in this case since it would only require minor changes when new elements are added to the repertoire throughout life (Dobson & Lemon, 1979).

Using the Markov model approach with information theory, the acoustic analysis of recorded warble sequences shows that male budgerigars follow certain rule(s) in combining warble elements when producing long, rambling songs. More specifically, warble can be nicely described as a 5<sup>th</sup> order Markov chain where a relatively large amount of information is gained within the first five warble elements in a sequence, and a relatively small amount of information is gained with the addition of the next (6<sup>th</sup>) element.

Compared to songbirds, whose vocalizations are usually described by a Markov chain lower than an order of three (e.g., American thrush: Dobson & Lemon, 1979;



European starling: Gentner & Hulse, 1998; cardinal: Lemon & Chatfield, 1971; chaffinch: Slater, 1983), budgerigars seem to be capable of creating variations in warble and encoding more information in the sequences of their complex repertoires. However, a 5<sup>th</sup> order Markov chain definitively shows that there is production limit on the number of elements budgerigars are able to manipulate once at a time during singing. The fact that the result is very similar among different individuals is strong evidence that this limit is species-specific.

The reason for this limit is a matter of speculation. Genetically, the ability to maneuver warble patterns may be simply hardwired. In zebra finch, knockdown of FoxP2, a protein implicated in song development, in Area X of the anterior forebrain pathway resulted in an incomplete and inaccurate imitation of tutor song that persisted into adulthood (Haesler, et al., 2007). Budgerigars also have FoxP2. In these open-ended learners, the expression of FoxP2 is related to the order of Markov model in their songs. Physiologically, budgerigars may not have a neural circuitry efficient enough to produce a Markovian unit higher than 5<sup>th</sup> order because the motor processing load would in some sense be too much. From an evolutionary perspective, whether and how the variations in warble affect its reproductive function is still unclear. It may be that changing the singing pattern too often might require so much energy that this cost outweighs the benefit of increased information exchange. In this case, the limit of five-element blocks may result from the best balance between the amount of information conveyed and the amount of energy used. Similarly, if the receiver can only process a short piece of warble at a time, there may be no added value in having a higher order Markovian structure in terms of

vocal production. However, this 5-element limit phenomenon may be a by-product of some other, unknown, evolutionary pressures and constraints.

Regardless of these possible explanations, since signalers and receivers usually evolve together (Seyfarth & Cheney, 2003), the results here suggest that budgerigars may have matching capabilities to perceive warble sequences, and this perceptual ability may also be one of the factors that regulate the motor control of warble production.

## Chapter 7: Perception of Sequences of Warble Elements

In the previous chapter, I have shown that budgerigars use 5<sup>th</sup> order Markovian units to construct short local structures in their warble. Whether this ability is unique or species-specific is not clear, but the budgerigar warble appears to be a higher order Markov model than do songbirds (Gentner & Hulse, 1998; Lemon & Chatfield, 1971; Lemon, et al., 1993; Martindale, 1980; Slater, 1983). As mentioned earlier, there is probably a neural processing load that is positively related to the number of transitions in the stream of warble to which the receiver needs to pay attention. If the benefit of attending to longer sequences does not outweigh the costs, selection pressure would constrain the evolution of higher Markov order sensitivity. Thus, it is reasonable to hypothesize that 5<sup>th</sup> order transition probabilities are not only the optimal length for the signalers to encode and produce information, but also the best strategy for the receivers to perceive and extract information.

In the following experiments, the perceptual capabilities for processing warble sequences are tested in budgerigars and compared with the result of the acoustic analysis from Chapter 6. Also discussed is whether the perceptual ability is natural and species-specific and whether it can be improved through learning.

### *Experiment 1: Budgerigars' sequencing ability on familiar sequences*

#### **Background and rationale**

Warble is mostly produced during courtship, when males sing to their mates and bring them to reproductive state (Brockway, 1964b, 1965, 1969). In theory, the goal of successful mating leads the evolutionary interests of signalers and receivers to overlap

and vary together. Selection should favor both signalers whose signals have easily extracted information and receivers who are skillful in extracting such information (Seyfarth & Cheney, 2003). Acoustic analysis of warble sequences showed that budgerigars use 5<sup>th</sup> order Markovian units to construct short local structures in their warble. However, there is no direct evidence indicating that budgerigars also perceive warble with a 5<sup>th</sup> order window that corresponds to the 5<sup>th</sup> order Markov chain during warble production.

In this experiment, budgerigars and other two songbirds (canaries and zebra finches) were tested on their ability to detect a change in the order of elements in a fixed sequence of four warble elements. The length of the sequence was then expanded by including new elements, making the task more difficult and testing the limits of the length of a perceivable sequence.

## **Method**

Four budgerigars (two males and two females) were used in the current experiment. Two zebra finches and two canaries were also used. However, both canaries and finches failed to learn the task even after intensive training and therefore were eliminated from the following experiments.

In each test session, the background was a continuous repetition of an artificial sequence composed of N warble elements from the same recording of the same individual. These were played with a constant silence interval of 150 ms. For example, if  $N = 4$ , the background sequence would be ...ABCDABCDABCD..., where A, B, C, and D are different warble elements. Elements of the same category were intentionally separated by at least one sound from another category. These N elements in the

background were also used as targets, each of which was inserted once in one trial. In other words, the same sound serves as both the target and part of the background in the same trial. Therefore, the only cue a subject can use to detect the target is a violation of the fixed ordering of elements. Using the above example, A, B, C, and D are all possible targets in a session of  $N = 4$ .

A trial involving target “A” proceeds as follows. The bird hears the sequence ABCD repeated continuously. The bird pecks the observation key until a random interval of 2-6 seconds times out. Then A is immediately inserted once as the next element in the background sequence, for example ...ABCDABACDABCD... There was a 0.5 second silence after each trial at the end of the response interval and then the repeating sequence ABCD begins again.

One test session was defined as approximately 100 trials, where about 25% are sham trials where no insertion occurs. The actual number of trials changes across sessions. When  $N$  varies across sessions, the number of possible targets ( $N$ ) and the number of sham trials (~25% of all trials) vary as well. To balance different levels of difficulties, one male and one female were tested on a set of sequences (see below) consisting of Ricky’s warble elements, and the other two subjects were tested on another set of sequences created by Puffy’s warble elements.

In the course of this experiment, subjects were tested against a background whose length gradually increased session by session starting with 4 elements ( $N = 4$ ). New elements were added one by one to the end of the original background until  $N = 14$ , and then six at a time thereafter until  $N = 50$ . For each of these seventeen conditions ( $N = 4, 5, 6 \dots 13, 14, 20, 26, 32, 38, 44$ , and 50), one session of data was collected.

Because both the hit rate and false alarm rate of the subjects could vary in this task,  $d'$  was calculated from the overall hit rate and false alarm rate of each session to indicate the subject's sensitivity of detecting an insertion in a repeating background. The 95% confidence interval (CI) of each  $d'$  was also calculated to show whether two  $d'$  values are significantly different (i.e., their CIs do not overlap) or not (Macmillan & Creelman, 2005).

## **Results**

After only four training sessions, in which the four elements that comprised the shortest sequence were introduced one by one, the budgerigars were immediately able to run successfully on the 4-element sequence task. The performance of all subjects is shown in Figure 19. The sensitivity for detecting an element insertion in an artificial warble sequence is approximately stabilized at  $d' = 1.5$  and is generally unrelated to the length of the sequence. These sequences were slowly built from a short sequence (4 elements) to a longer sequence (up to 50-element). On the average, it only took about 20 days for the animals to complete the experiment (each session was 20 to 40 minutes, and every bird ran two sessions per day).

Given the repeating background method used, one could image that birds would become more familiar with the beginning of the sequence compared to the newly-added pieces as they were tested over days and weeks. This might make them better at detecting targets inserted among the first few elements of the sequence compared to later ones. If this is the case, budgerigars' performance would decrease as the sequence becomes longer, given that the insertions were evenly distributed along the whole sequence. However, such trend was not observed (Figure 19).

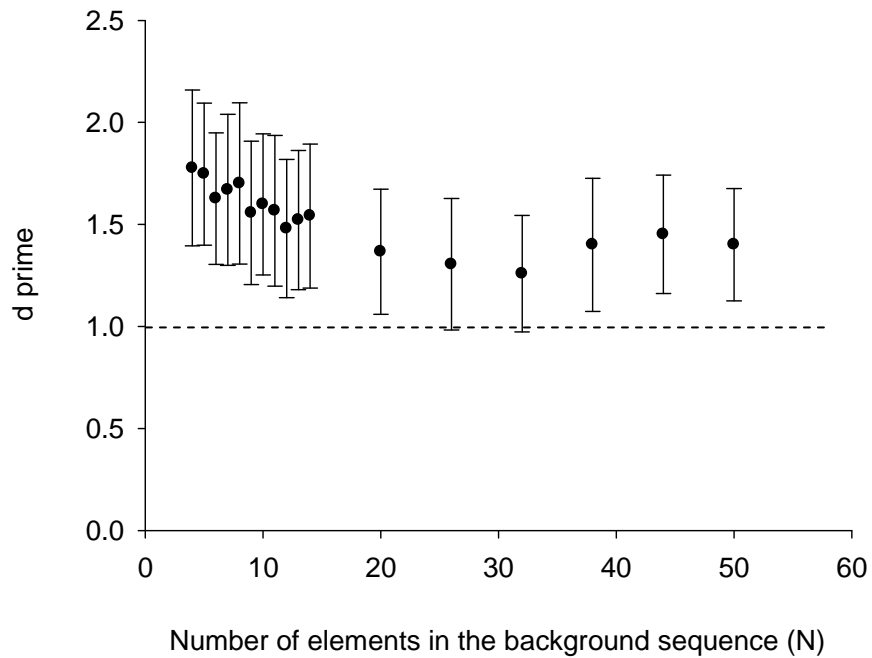


Figure 19: The performance of all subjects detecting insertions in a sequence. Shown as overall  $d'$  with 95% confidence interval in the error bars. Dashed line is at  $d' = 1$ .

## Discussion

Conventionally, chance performance was defined as a  $d'$  score of zero, and threshold discrimination was defined as a  $d'$  score greater than 1 (Macmillan & Creelman, 2005). The overall good performance ( $d'$  consistently greater than 1) showed that gradually increasing warble length may not be a difficult task for the birds (Figure 19). However, the Markov chain analysis in Chapter 6 suggests that the capacity of neural processing may restrict the immediate attention span to five or six elements in budgerigars. It seems that through continuous practice starting from a short sequence, budgerigars are able to manage small chunks of information while each chunk is still

within the limit of 5<sup>th</sup> order Markov model (less than 6 elements), store them as fixed but still parseable sequences, and recollect them as needed. Whether this only works when the birds have sufficient time to learn the short sequences and are repeatedly exposed to them is an open question, but it would be interesting to understand the limits of their ability.

There are several other interesting findings from this experiment. First, the data suggest that males may be slightly more sensitive than females (Figure 20A). It is also the case that males are easier to train than females (data not shown here). Intuitively, we could argue that males, as primary signalers, need to listen to themselves while warbling. We already know that auditory feedback is important in budgerigars during vocal learning (Heaton, et al., 1999). Males may need to carefully monitor what they are producing in order to quickly make changes in warble depending on the behavior of the female. Although females are primarily receivers, they may need to pay more attention to some other aspect of the content of warble (e.g., concomitant visual display and other intimate courtship behaviors) in addition to or instead of the overall sequential ordering.

Second, the composition of warble sequences does not seem to affect the ability of birds to detect insertions (Figure 20B). Birds tested on Ricky's elements and those tested on Puffy's elements showed similar patterns. This suggests that the ability to memorize a sequence is more related to biological constraints on the number of elements or the amount of practice rather than on the characteristics of the sound itself. In other words, this limit may be biologically hardwired regardless of what the information is.



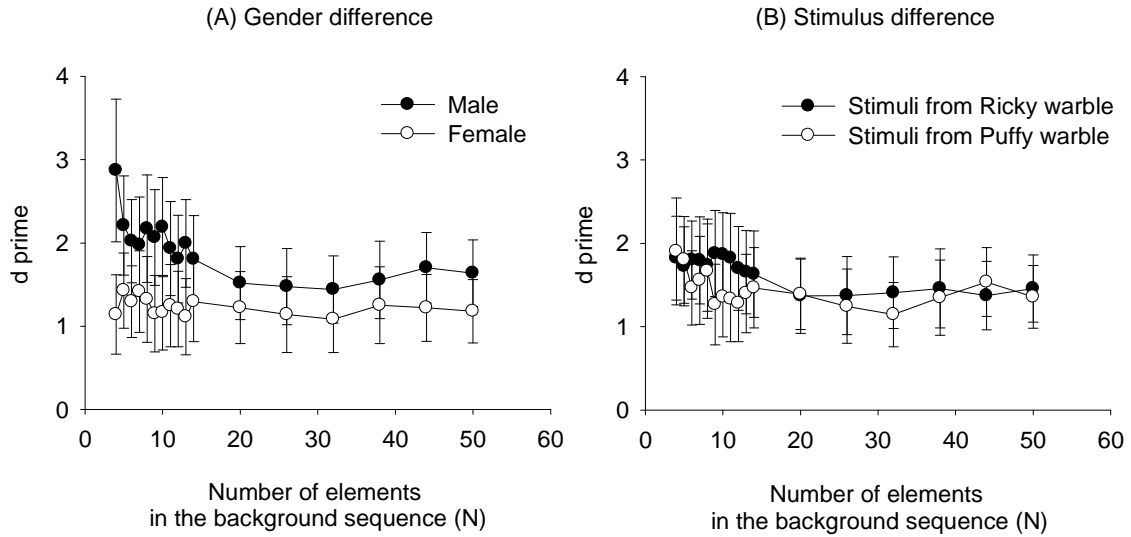


Figure 20: (A) Comparison of the sensitivity between males ( $n = 2$ ) and females ( $n = 2$ ). (B) Comparison of the sensitivity between different stimuli. Two birds were tested on Ricky's warble elements, and two birds were tested on Puffy's warble elements. Shown as overall  $d'$  with 95% confidence interval in the error bars.

### Experiment 2: Budgerigars' sequencing ability on novel sequences

#### **Background and rationale**

The conclusion of the test above suggests that when the sequence becomes longer gradually by adding short (less than 6 elements) new pieces, budgerigars are able to learn a sequence up to 50 elements long through continuous practice, even though their natural warble is only a 5<sup>th</sup> order Markov chain (see Chapter 6).

Here, a similar experiment was designed to test budgerigars' ability to detect a change in a novel sequence (i.e. one that they are not familiar with). Once birds finished the previous experiment, they were tested on five new sequences where the original warble elements were still used, but the order of these elements was scrambled.

## Method

The sound stimuli and testing paradigm were the same as Experiment 1. Five new tests were run ( $N = 4, 7, 14, 26$ , and  $50$ ) using a novel artificial sequence with the same component elements. These data were referred as “new” compared to the corresponding “original” sessions.  $D'$  and the standard error (se) were calculated to construct a 95% confidence interval (CI) in each session, which was used to evaluate the statistical difference between sessions (Macmillan & Creelman, 2005).

## Results

Figure 21 and Table 7 summarize the pooled data of all four subjects since their responses are quite similar. The performance drops significantly with the new sequences (new 50-element sequence:  $d' = 0.39$ ,  $se = 0.15$ ; original 50-element sequence:  $d' = 1.39$ ,  $se = 0.14$ ), even though the element compositions are identical to the original sequences, until the sequence contained less than 7 elements (new 4-element sequence:  $d' = 1.53$ ,  $se = 0.19$ ; original 4-element sequence:  $d' = 1.74$ ,  $se = 0.19$ ).

Table 7: Summary of  $d'$  values and standard errors of budgerigars' performance on the original sequences and the new sequences.

N	Original		New	
	$d'$	se	$d'$	se
4	1.74	0.19	1.53	0.19
7	1.67	0.19	0.98	0.17
14	1.53	0.18	0.55	0.16
26	1.29	0.16	0.46	0.14
50	1.39	0.14	0.39	0.15

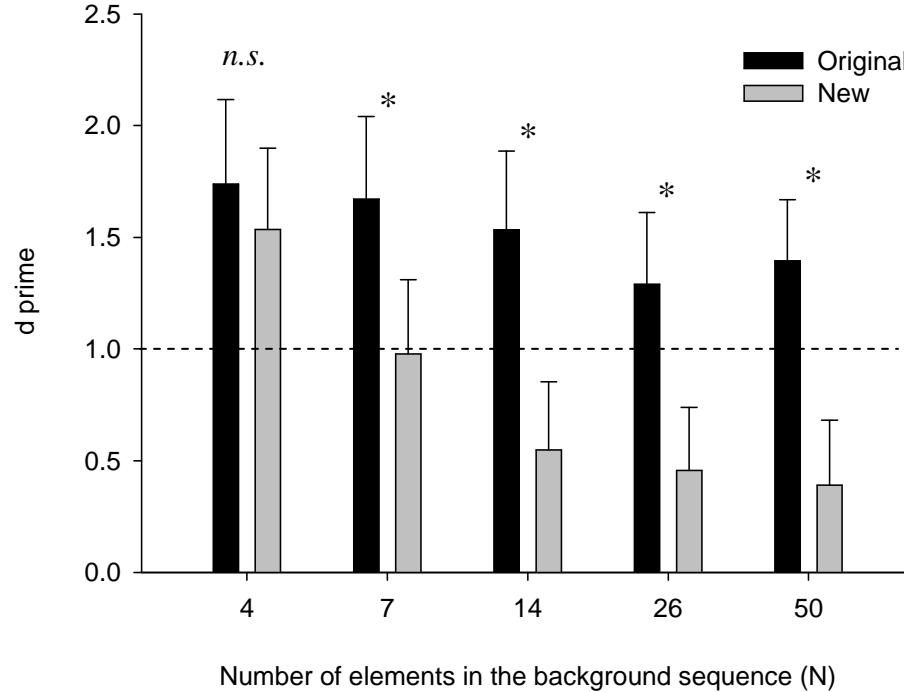


Figure 21: Comparison of budgerigars' sensitivity to detect a change in the original familiar sequences and the new sequences composed of the same elements. Shown as overall  $d'$  with 95% confidence interval in the error bars. Dashed line is at  $d' = 1$ .

## Discussion

The results show that if budgerigars are tested on a novel sequence longer than a 5<sup>th</sup> order Markovian unit (i.e., 6 elements), they are apparently unable to manage the information and therefore fail to detect an insertion. This suggests that budgerigars may be perceptually constrained to use building blocks that are less than 7 elements long (but definitely more than 4 elements long) in memorizing longer sequences.

This conclusion coincides well with the earlier finding that in vocal production, budgerigars can only manipulate no more than 5 transition probabilities while warbling. Evolutionarily, it makes sense that receivers covary with signalers, and this is reflected in

our findings that budgerigars may perceptually treat long warble sequences as multiple 6-element units (5<sup>th</sup> order transition probabilities) to facilitate memorization and information extraction.

The capacity of neural processing determines the span of immediate memory, which in turn imposes limitations on the amount of information one can receive, process, and remember. In humans, Miller (1956) found that we usually organize input stimuli into seven, plus or minus two, chunks, which is roughly the number of objects that we can focus on at once. Capacity can be increased if each chunk can be further parsed into seven smaller chunks, and each smaller chunk can further break down into seven even smaller chunks, and so on, forming an hierarchical structure that arranges incoming information into a sequence of chunks of seven. By this strategy, we can break (or at least stretch) the bottleneck of our immediate memory. The current results show that these same mechanisms may be operating in budgerigars as well, but with a slightly smaller “chunk size” of six warble elements.

Taken together, these two experiments suggest that budgerigars may build their memory representation from the bottom up, and it may only take a few trials for them to learn. Although they are not able to handle information immediately in a new, long sequence, they can strategically store what they have previously learned (at least up to 50 elements) and convert on-line immediate memory span to relatively long-term compilation. In terms of future experiments, aside from learning how much they can store this way, it would also be interesting to test for how long these stored items remain in memory. This might provide useful insight about different forms of memory involved in budgerigars.

## Chapter 8: Detection of Insertions in Warble Sequences

In human speech, each word has its own spectrotemporal features and meaning, while different combinations of words may provide different information. The set of ordering rules used to string words together into grammatical sentences is defined as syntax. It is required to truly understand and use language to communicate.

Like running speech, warble is long and complicated. Budgerigars form acoustic and perceptual categories of warble elements, depending on their spectrotemporal characteristics and potential functions (Chapter 3, 4, and 5). But there is scant evidence on whether important information is encoded in the order of warble elements, and whether budgerigars pay attention to the sequential cues and extract information from different combinations of warble elements.

Previous experiments have established that budgerigars are capable of memorizing a long, artificial sequence of warble elements through learning and can detect a change in a regular sequence of elements based only on ordering. It is important to determine whether this capability extends in any way to natural warble sequences.

This series of experiments was designed to examine budgerigars' sensitivity for detecting different types of insertions in a background of up to 1000 elements natural warble sequence. These experiments are not aimed at showing whether or not a human language like "syntax" exists in warble, but rather the more modest goal of showing whether there is any perceptual significance to the sequential order of warble elements to budgerigar perception.

## Experiment 1: Detection of non-budgerigar vocalizations in natural warble sequences

### **Background and rationale**

In order to ensure that the subjects learned the task, pure tones (non-biological sounds) and zebra finch song syllables (non-budgerigar vocalizations) were first used as targets and inserted into a natural warble sequence. These stimuli were easy for the budgerigars to detect and served as a simple training experiment that introduced subjects to the task.

### **Method**

#### *Subjects*

Four budgerigars, two zebra finches, and two canaries were tested.

#### *Procedure*

The basic paradigm is similar to what was used in the previous experiment (Chapter 7). Once a session starts, a sequence of warble elements is continuously played as background. A trial starts after the subject pecks the observation key, and after a random interval of 2-6 seconds a target is inserted into the sequence. This random interval is necessary to prevent the birds from establishing a rhythm of pecking instead of actually listening to and detecting the insertions. However, the random length of this interval makes it impossible for the program to control exactly where the insertion occurs in the continuous background of running stream of warble. Since we were interested in the birds' initial ability to detect insertions in a natural warble sequence without learning, all data were of interest regardless of the bird's percent correct level. Thus, the birds' initial response to these insertions was taken as final data rather than a stabilized correct

percentage that might occur after extensive training.  $D'$  prime was calculated and used to indicate the sensitivity of detecting a target. Standard error of  $d'$  was also calculated in order to provide a 95% confidence interval for statistical evaluation of two conditions.

### *Background sounds*

Three warble sessions were recorded from each of four male budgerigars (Buzz, Ricky, Puffy, and Yuri), making a total of 12 background warble sets. These recordings were segmented into individual elements by the procedure described earlier (Chapter 2 and 3). Each background contains more than 900 elements and was played back in its natural sequence with a constant 150 ms of silent inter-element interval. In other words, the order of the elements is preserved, but the “tempo” of the background is not perfectly natural. Nevertheless, it sounds like a natural warble to human ears.

For each subject, four sessions were run on each target set (see below). The background in each session was randomly chosen and arranged so that no two consecutive running sessions used a background from the same bird.

### *Targets*

Because the number of targets varies across sessions, the number of sham trials was also varied so that sham trials were approximately 20% of a session. The number of total trials per session varies accordingly.

#### ● ***Pure tone experiment***

One session includes a seven-sound duration gradient (25, 50, 100, 150, 300, 400, 670 ms) where the frequency is fixed at 1500 Hz, and a seven-sound frequency gradient (500, 1000, 1500, 2000, 2500, 3000, and 4000 Hz) where the duration is fixed at 150 ms, making a total of 13 stimuli (one stimulus, a 1500 Hz tone of 150

ms, is the same in these two gradients). These durations and frequencies cover the range of warble element characteristics. Every stimulus repeated 6 times (a total of 78 trials), plus 24 sham trials in each session.

- ***Zebra finch song syllable experiment***

One song from each of the six zebra finches recorded was segmented into syllables and used in one session. Zebra finch song syllables often have a fundamental frequency of approximately 500 Hz and harmonic structures. Their durations vary from 30 ms to 250 ms. One session includes a total of 42 targets, each of which repeats twice, and 27 sham trials.

## Results

Figure 22 summaries the results of all three species. They all show high sensitivity for detecting pure tones and zebra finch song syllables in warble. Budgerigars are clearly more sensitive than the other two species.

Table 8: Summary of  $d'$  values and standard errors of the subjects' sensitivity for detecting pure tones and zebra finch song syllables. Comparison between conditions in the same species.

	Detecting pure tones		Detecting ZF syllables		
	$d'$	se	$d'$	se	
Budgerigar	4.46	0.22	4.24	0.21	<i>n.s.</i>
Zebra finch	3.46	0.19	3.58	0.21	<i>n.s.</i>
Canary	2.88	0.19	2.85	0.18	<i>n.s.</i>



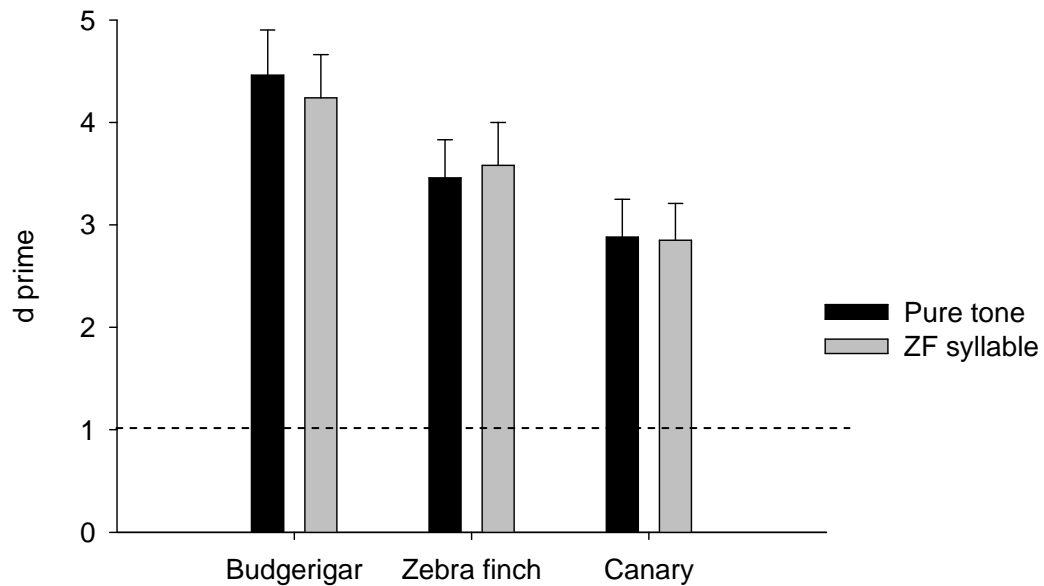


Figure 22: The performance of all three species detecting insertions of pure tones and zebra finch syllables. 95% confidence interval is presented in the error bars. Dashed line indicates  $d' = 1$ .

There is no significant difference in the discriminability of pure tones and that of zebra finch syllables for any of the species (Figure 22; Table 8). Zebra finches do not perform particularly better at detecting their own vocalizations, but their sensitivity for detecting zebra finch syllables is slightly higher than detecting pure tones and reach the level where it is significantly better than canaries but as good as budgerigars (Table 8).

## Discussion

All three species perform well at detecting pure tones and zebra finch song syllables in warble with false alarm rate lower than 20%. This result was expected since these are either non-biological sounds that are acoustically distinct from warble elements, or non-budgerigar vocalizations that are also acoustically distinct. Results show that all

three species can pick these sounds out of a very complex background. Budgerigars still have better sensitivity than the other two species, which indicates that conspecific warble makes other non-warble sounds especially easy to detect.

Most of the “misses” by zebra finches and canaries occur in low frequency pure tone targets which budgerigars do not have trouble with (data not shown here). This can be explained by their behavioral audiograms. Budgerigars have better hearing than finches and canaries at low frequencies. At 500 Hz, the difference in threshold is up to 30 dB SPL (about 20 dB SPL for budgerigars and about 50 dB SPL for finches and canaries) (Dooling & Okanoya, 1995a; Dooling & Saunders, 1975; Hashino & Okanoya, 1989; Okanoya & Dooling, 1987).

This experiment shows that budgerigars, zebra finches, and canaries can be trained to perform in the psychophysical chamber, while listening to long, natural sequences of warble and to detect “oddballs” inserted in it. This creates the foundation of subsequent experiments that were more focused on budgerigars’ sequencing ability.

### *Experiment 2: Detection of budgerigar calls in natural warble sequences*

#### **Background and rationale**

Having established that budgerigars, zebra finches, and canaries can be trained to pick out insertions within a natural warble sequence, the next step is to make the task more “natural” by using sounds that normally occur in budgerigar warble.

Two sets of vocalizations, contact calls and warble calls (contact call-like warble elements), have been extensively studied (Chapter 5). Budgerigars perceive them as belonging in two different categories, but zebra finches and canaries do not. Here contact calls and warble calls were used again as targets to be inserted in a natural warble

sequence, and birds' detection performance was compared with the perceptual results obtained in call discrimination tasks described in Chapter 5.

## **Method**

### *Subjects*

The same four budgerigars, two zebra finches, and two canaries were tested.

### *Procedure*

The same as Experiment 1.

### *Background sounds*

The same 12 background sets from Experiment 1 were used. For each bird, eight sessions were run on each target set. The background sequence was randomly chosen and arranged so that warble sequence from the same bird was not used in two sessions in a row to minimize the effect of memorization.

### *Targets*

To minimize the effect of memorization, "target sets" composed of multiple tokens of a call type (contact calls or warble calls) were used in one session instead of repeatedly presenting the same target in every trial.

- ***Contact calls***

One session contains 80 calls, 20 from each of the four birds (Buzz, Ricky, Puffy, and Cosmo), and 20 sham trials. Note that Cosmo's contact calls were used instead of Yuri's because Yuri died before any of his contact calls were recorded.

- ***Warble calls***

One session contains 72 calls, 18 from each of the four birds (Buzz, Ricky, Puffy, and Yuri), and 24 sham trials.

## Results

When detecting contact calls embedded in warble sequences, budgerigars remained very sensitive to the targets ( $d' = 3.86$ ), but somewhat less sensitive than detecting zebra finch syllables ( $d' = 4.24$ ). Zebra finches and canaries, on the other hand, were significantly less sensitive but still showed a high  $d'$  (Figure 23).  $D'$  dropped from 3.58 to 2.46 in zebra finches and from 2.85 to 1.69 in canaries.

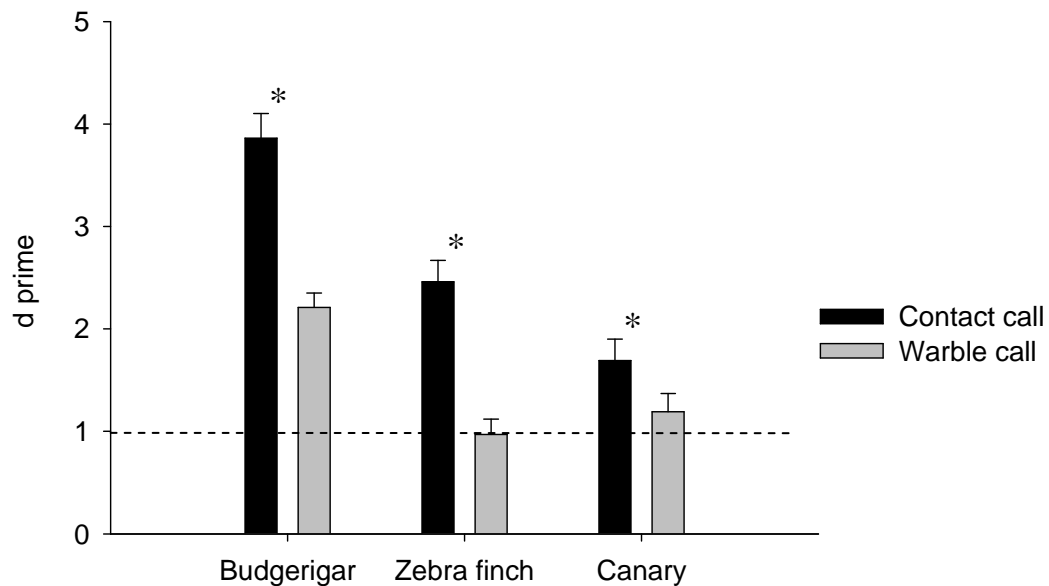


Figure 23: The performance of all three species detecting insertions of contact calls and warble calls. 95% confidence intervals were shown in error bars. Dashed line indicates  $d' = 1$ .

When detecting warble calls in warble sequences, budgerigars' performance decreased substantially compared to their sensitivity in detecting contact calls in warble sequences ( $d'$  dropped from 3.86 to 2.21). The sensitivity of zebra finches and canaries

also declined significantly ( $d'$  dropped from 2.46 to 0.97 in zebra finches, from 1.69 to 1.19 in canaries) (Figure 23; Table 9).

Table 9: Summary of  $d'$  values and standard errors of the subjects' sensitivity for detecting contact calls and warble calls. Comparison between conditions in the same species.

	Detecting contact calls		Detecting warble calls		
	$d'$	se	$d'$	se	
Budgerigar	3.86	0.12	2.21	0.07	*
Zebra finch	2.46	0.11	0.97	0.08	*
Canary	1.69	0.11	1.19	0.09	*

Comparing across species, the budgerigars' ability to detect a contact or warble call were significantly better than that of zebra finches and canaries (Table 9). However, the zebra finches' and canaries' performance was still significantly better than chance (Figure 23). This result is surprising. The expectation was that perception of warble elements ordering might be species-specific and that zebra finches and canaries would not be able to detect the targets above chance level. As a follow-up analysis to test the alternative hypothesis that calls from different individuals may have subtle individual identity cues that help the birds to detect a target, the target set was further divided into calls from the same individual as the one in the background (e.g., Buzz's calls tested against Buzz's warble background) and calls from a different individual than the one that provided the background warble (e.g., Buzz's calls tested against Ricky's warble background).

Table 10: Summary of  $d'$  values and standard errors of the subjects' sensitivity for detecting contact calls (A) and warble calls (B) from the same individual as the background and from different individuals than the background. Comparison between conditions in the same species.

(A) Results of detecting **contact calls**

Targets and background	From the same individual		From different individuals		
	$d'$	se	$d'$	se	
Budgerigar	3.48	0.15	3.98	0.13	<i>n.s.</i>
Zebra finch	2.17	0.13	2.64	0.11	<i>n.s.</i>
Canary	1.41	0.13	1.74	0.11	<i>n.s.</i>

(B) Results of detecting **warble calls**

Targets and background	From the same individual		From different individuals		
	$d'$	se	$d'$	se	
Budgerigar	1.11	0.09	2.77	0.08	*
Zebra finch	0.39	0.10	1.18	0.08	*
Canary	0.35	0.13	1.01	0.11	*

Analysis of the data based on the source of the target calls showed that the birds' responses were not significantly different when detecting contact calls from the same individual as the background warble and those of a different individual from the background warble (Figure 24A; Table 10A). The  $d'$  in all cases remained above 1. However, birds were better when detecting warble calls recorded from different birds than the bird that provided the background (Figure 24B; Table 10B). In particular, 64% of the correct responses of zebra finches and 71% of those of canaries when detecting warble calls seen earlier in fact only showed their sensitivity to different individuals. It is significantly more difficult for them to detect a warble call if it is from the same

individual that produced the background sequence (Table 10B) and the cues for a successful detection have been limited to the ordering of warble elements. Nevertheless, even though budgerigars' sensitivity also declined when the target warble calls were from the same individual in the background, it is still above  $d' = 1$ .

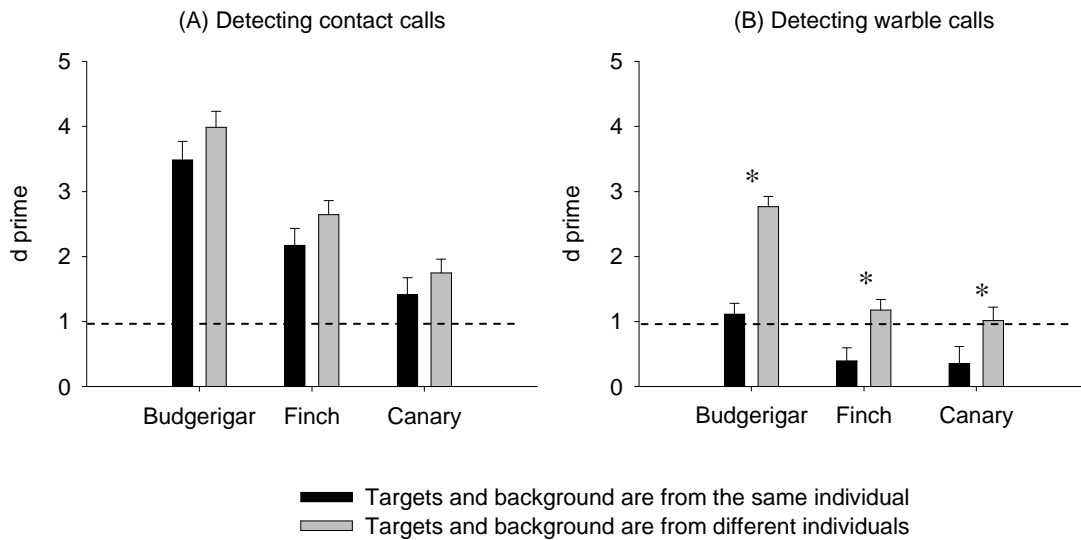


Figure 24: Comparison between targets from the same individual in the background and targets from a different individual than the one in the background. 95% confidence intervals are shown in error bars. Dashed line indicates  $d' = 1$ .

It is instructive to compare these results to those from Chapter 5. Here contact calls and warble calls were inserted in a natural warble sequence. In Chapter 5, contact calls and warble calls were discriminated from a pool of 70 independent warble calls (see Chapter 5 for more details). All three species showed significant improvement in detecting contact calls when the targets were embedded in a natural warble sequence compared to being played among a collection of unrelated single warble calls (Figure 25A). This effect was not observed when detecting warble calls (Figure 25B).

Table 11: Summary of  $d'$  values and standard errors of the subjects' sensitivity for detecting contact calls (A) and warble calls (B) from a collection of unrelated warble calls and against a natural warble sequence. Comparison between conditions in the same species.

(A) Results of detecting **contact calls**

Background sounds	A collection of unrelated warble calls		A natural warble sequence		
	$d'$	se	$d'$	se	
Budgerigar	2.69	0.13	3.86	0.12	*
Zebra finch	1.04	0.14	2.46	0.11	*
Canary	0.40	0.19	1.69	0.11	*

(B) Results of detecting **warble calls**

Background sounds	A collection of unrelated warble calls		A natural warble sequence		
	$d'$	se	$d'$	se	
Budgerigar	2.28	0.13	2.21	0.07	<i>n.s.</i>
Zebra finch	0.75	0.14	0.97	0.08	<i>n.s.</i>
Canary	0.85	0.18	1.19	0.09	<i>n.s.</i>



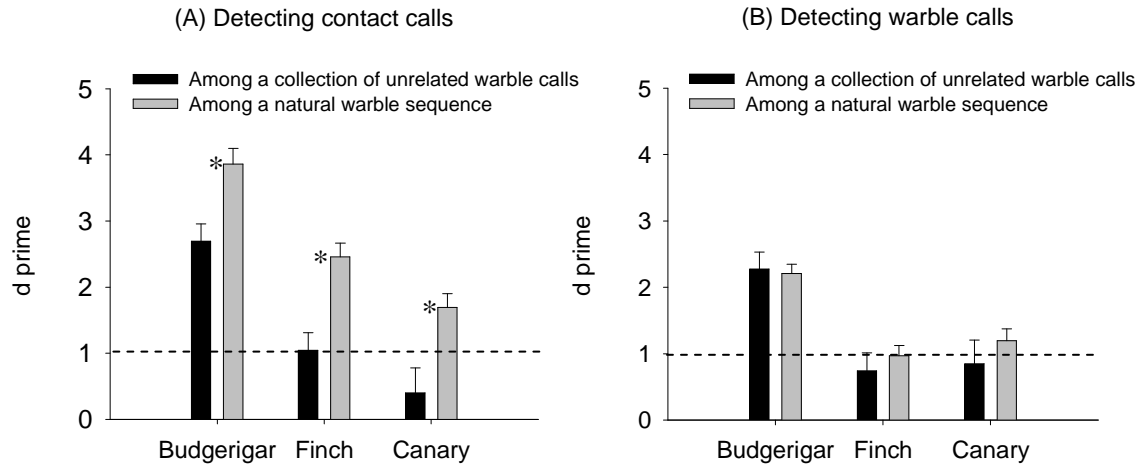


Figure 25: Cross-species comparison of the sensitivity of detecting a contact call (A) or a warble call (B) among a group of individual calls (Chapter 5) and in a continuous warble sequence (present experiment). 95% confidence intervals are shown in error bars. Dashed line indicates  $d' = 1$ .

## Discussion

All birds had more difficulty when the background and targets were both budgerigar vocalizations (and even from the same individual for some trials). From earlier experiments, we know that contact calls and warble calls are acoustically different in many aspects, and budgerigars (but not finches and canaries) are able to perceptually distinguish between these two types of vocalizations in their repertoire. Not surprisingly, the results here are along the same line.

The sensitivity of detecting warble calls significantly decreased when the target and the background were from the same individual, but the birds' performance in detecting contact calls remained at high levels (Table 10). Since contact calls and warble calls are obviously different in several acoustic aspects (see Chapter 5), contact calls should be relatively easy to detect in a warble sequence, regardless of their origin.

On the other hand, when a warble call is from the same individual that produced the background sequence, the same call served both as the target and part of the background in a running session, and the cues that birds' can use for detection have been restricted to the sequential ordering of warble elements in the background – whether a warble call is out of place or not. This significantly lowers the sensitivity in all three species, especially zebra finches and canaries, whose  $d'$  decreased significantly to almost chance level (Table 10). For budgerigars, detecting warble calls from one individual against the background warble of another individual is relatively easy because subjects may have the extra cues of “vocal characteristics” of a specific individual, but  $d'$  was still above 1 when detecting warble calls of the same individual from the background. This suggests that budgerigars do attend to the ordering of warble elements and can detect alterations in it.

A close comparison with past data showed that contact calls embedded in a warble sequence are significantly easier to be detected for all three species (Figure 25A). Natural warble sequences create a homogeneous flow of budgerigar vocalization in the background, possibly giving the birds a Gestalt perception that a collection of individual warble calls does not provide, and thus making the contact calls stand out as targets clearly. In other words, it may not necessarily relate to the sequential order per se, but the overall warble environment makes contact calls (vocalizations of a different type) more obvious and easy to detect.

However, a natural warble sequence is no longer advantageous when it comes to detecting warble calls. The discriminabilities are the same when targets were presented among a group of individual warble calls as when they were inserted in a natural warble

sequence (Figure 25B). Since the targets were the same type of vocalizations as the background, streams of warble do not provide a good contrast to the targets anymore. On the contrary, target warble calls could easily blend into the background and make them even harder to be detected, although we did not observe any notable interference that restrains detection.

### Experiment 3: Detection of warble calls in warble sequences

#### **Background and rationale**

From the result of Experiment 2 we know that budgerigars are able to detect a warble call inserted into a natural warble sequence. The performance was better when the insertion was a warble call from an individual other than the background warble singer, suggesting warble calls may contain information that allows for individual recognition. However, even when the inserted warble calls and the background warble were produced by the same individual, the performance of budgerigars was still above the discrimination threshold at  $d' = 1$ . In order to eliminate all possible cues that the subjects can use to detect an insertion except the ordering of warble elements, and to pinpoint the importance of sequential effect in budgerigars warble, only warble calls and background warble recorded from the same individual were used in this experiment.

#### **Method**

##### *Subjects*

The same four budgerigars were tested. Note that only budgerigars were used in this experiment. The task was too difficult for other species: zebra finches had very high

false alarm rates, and canaries had very low hit rates. Both situations indicate a very poor performance, and the birds could not be tested further.

### *Procedure*

The same as Experiment 1.

### *Background sounds*

One background set was randomly selected from each of the four individual (Buzz, Ricky, Puffy, and Yuri) (see Experiment 1). The resulting four background sets were used in three ways listed below in order to pinpoint the importance of ordering in warble elements on budgerigars' sensitivity of detecting insertions.

- ***Natural sequence experiment***

Here, the background elements were played in their natural sequence as in previous experiments. Birds were tested twice with a 4-months break in between. (The birds were run on a different task (Chapter 7) during these 4 months.) The results before 4 months and after 4 months were compared to test whether budgerigars depended on their long-term memory to solve the task.

- ***Randomized sequence experiment***

The same background warble sets were used, but the background warble elements were produced in a random sequence, not in their natural sequence.

- ***Natural sequence of reversed warble elements experiment***

The same background warble sets were used. Each background element was temporally reversed but the order of elements was still in natural sequence. In this experiment, the duration and overall spectrotemporal features of each warble

element remained the same, but the fine structure of the whole sequence has been changed.

### *Targets*

One target set of warble elements was used. This set was made up of warble calls from the same individual who produced the background sequence, selected from multiple recording sessions.

There were 100 trials in one session, 80 of which were test trials where one target was inserted, and 20 of which were sham trials where no insertion was presented. For each background warble set, each subject ran four sessions, a total of 400 trials.

### **Results**

Figure 26 presents the  $d'$  values of budgerigars detecting warble calls as targets inserted in different background warble sets. At first glance, budgerigars seemed to have sensitivity significantly higher than chance level ( $d' = 0$ ) in all four conditions (Figure 26), but it was an unexpected result. Since the inserted warble calls and the background warble sequence both came from the same individual, making the ordering of background warble elements the only cue to identify an insertion, there should be no way that budgerigars can detect a warble call out of a randomized background warble set. Therefore, in order to ensure that background warble element ordering was truly the only cue left for the subjects to use, the target sets were further divided into warble calls from the same recording session as the background warble sequence, and warble calls from a different recording session of the same individual who produced the background warble sequence. By doing so, we hope to eliminate artificial effects (e.g., minor changes during

recording) and biological fluctuations (e.g., the context in which the bird warbled) as much as possible.

Figure 27 and Table 12 summarize the results of the follow-up analysis. Warble calls from recording sessions different from the background warble sequences did provide extra cues to improve budgerigars' sensitivity of detecting the insertions (Figure 27B). When the targets were limited to the warble calls extracted from the background warble sequence, budgerigars' performance decreased in all of the conditions (Figure 27A). Thus, most of the performance shown in Figure 26 was in fact budgerigars being extremely sensitive to subtle cues other than the warble element ordering in the background.

Table 12: Summary of  $d'$  values and standard errors of the budgerigars' performance on detecting warble call insertions in different backgrounds.

Targets and background	From the same session		From different sessions	
	$d'$	se	$d'$	se
Sequential (a)	0.98	0.10	1.70	0.09
Sequential (b)	0.67	0.10	1.35	0.09
Randomized	0.18	0.10	1.28	0.09
Reversed	2.62	0.12	3.11	0.12

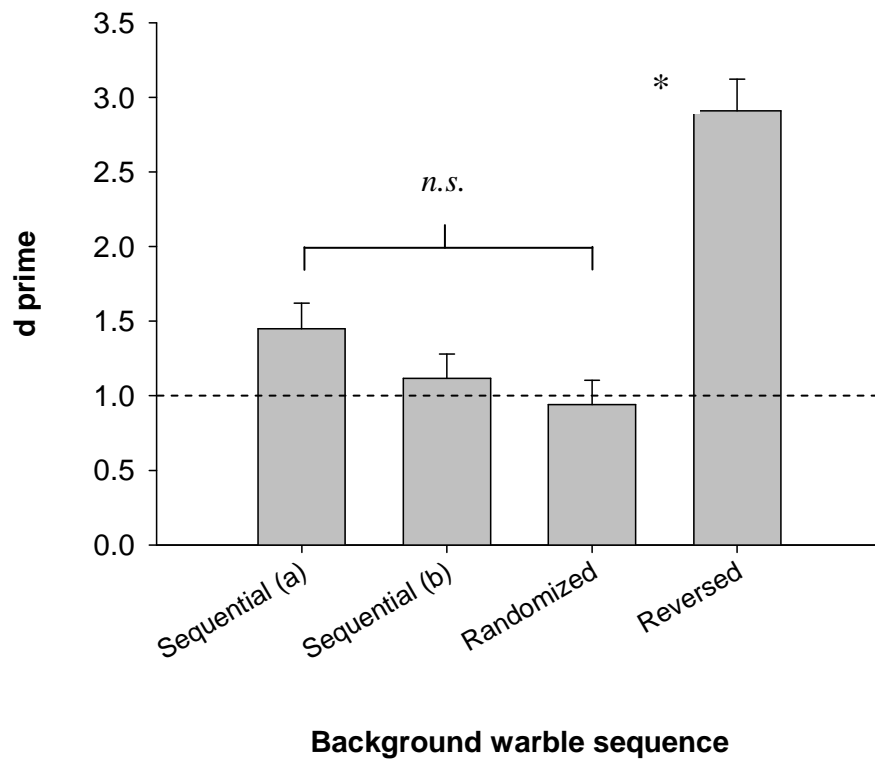
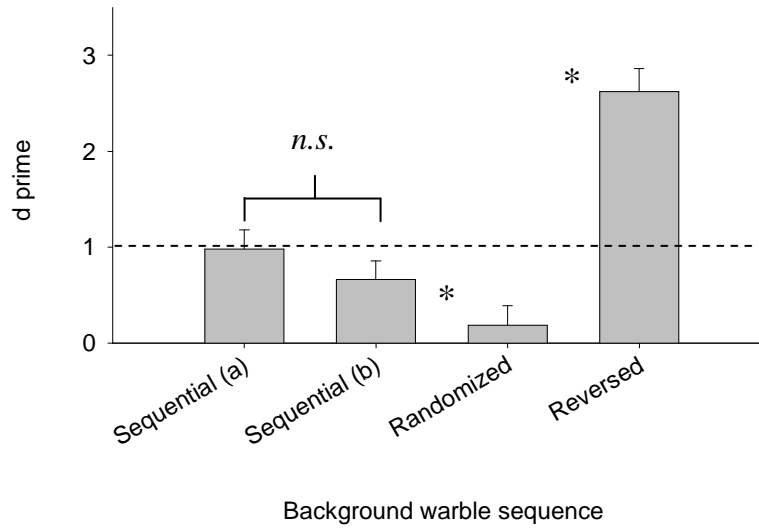


Figure 26: Comparison of budgerigars' ability to detect a warble call inserted in different background warble sequences. Sequential (a) shows the data when the birds were run four months before. Sequential (b) shows the data when the birds were run four months later. During these four months, the birds were run on a different task described in Chapter 7. 95% confidence intervals are shown in error bars. Dashed line indicates  $d' = 1$ .

**(A) Targets: warble calls from the same recording session as the background warble sequence**



**(B) Targets: warble calls from a different recording session than the background warble sequence**

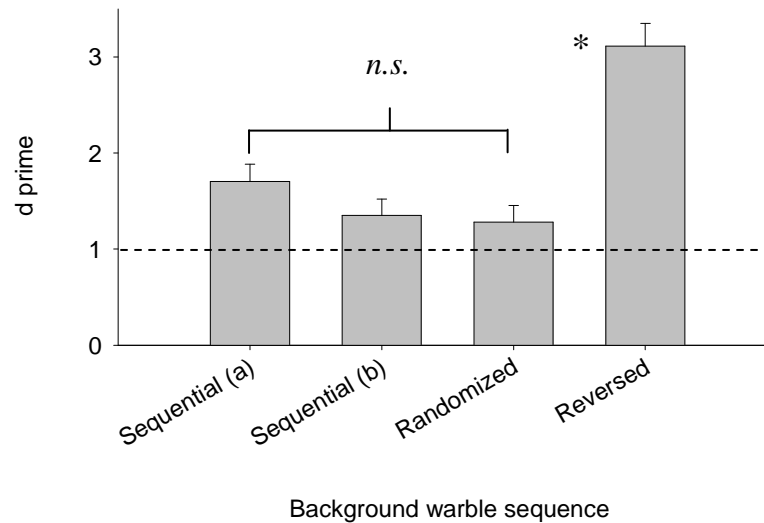


Figure 27: Comparison of budgerigars' ability to detect insertions of warble calls in different background warble sequences. (A) Analysis of inserted warble calls from the same recording session as the background warble sequence. (B) Analysis of inserted warble calls from a different recording session of the same individual being inserted in different background warble sequences. Sequential (a) shows the data when the birds were run four months before. Sequential (b) shows the data when the birds were run four months later. During these four months, the birds were run on a different task described in Chapter 7. 95% confidence intervals are shown in error bars. Dashed line indicates  $d' = 1$ .



When pursuing the effect of sequential ordering of warble elements on budgerigars' perception, we are more interested in the results shown in Figure 27A where the same warble call sometimes served as target and sometimes served a part of the background, leaving the ordering the only cue that budgerigars could use. Here, comparing "Sequential (a)" (when the birds were run four months before) with "Sequential (b)" (when the birds were run four months later), budgerigars' discriminability does not show any significant difference, but both of them are significantly higher than "Randomized" and significantly lower than "Reversed" (Table 12).

When the background was randomized, i.e., a new, unnatural sequence was created and presented to the birds,  $d'$  significantly declined to chance (not significantly different than  $d' = 0$ ). When reversed elements were used in the background sequence, although they were played in their natural sequence, budgerigars responded significantly faster and more accurately as if the targets were as easy to detect as pure tones (Table 12).

## **Discussion**

As mentioned earlier, the fact that budgerigars were able to detect a warble call insertion in a randomized warble sequence is probably due to some small acoustic variations that budgerigars, but not humans, are exquisitely sensitive to. Some of these variations, for instance, could be caused by subtle changes, such as the position of the singer's head relative to the microphone, the bird's level of arousal, etc.

To answer the question of how important the ordering of warble elements is to budgerigars, we should focus on the result of insertions of warble calls from the same

recording sessions as the background (Figure 27A). Budgerigars' performance did not change significantly over time (Table 12), suggesting that they indeed learned and used the ongoing sequential cue within a session to detect an out-of-order warble call, and this sensitivity neither deteriorates within 4 months nor improves through learning and practicing some other task (in Chapter 7). However, one may argue that it has nothing to do with the ordering of warble elements. The budgerigars could simply memorize every single target used in this experiment and pick them out by recognizing those sounds among the background, i.e., a group of other sounds. To minimize this possibility, in each running session, all targets in a target set were only presented once, making it a challenge for the bird to memorize individual targets. This was further shown in the condition where the background warble elements were randomized and the bird retested.

When the warble elements were played in random order as the background, budgerigars' sensitivity decreased significantly to chance ( $d' = 0$ ), indicating that they were no longer able to detect any insertion. Note that the randomized sequence in the background was composed entirely of previously heard elements. Also note that the randomized sequence had novel element ordering, with which the subjects were not familiar. Thus, the decrease in performance, compared to the result of detection against natural sequences, could be interpreted as the effect of ordering. Alternatively, the budgerigars may not necessarily learn the explicit ordering but just sense the holistic changes in the rhythmic or melodic "texture" of the background warble sequence. When the randomization destroyed the overall "harmony" of natural warble, the subjects were no longer able to detect any additional "discord."

Finally, individually reversed elements were played in their natural sequence in the background to serve as another comparison. Temporally reversed warble elements preserved budgerigars “vocal quality” in terms of physical complexity and acoustic characteristics such as overall spectral content, intensity, and duration, but distorted temporally-based properties. They are therefore ideal as controls for acoustic input.

In human speech, words played backwards violate several phonological properties that are universally observed (Binder, et al., 2000; Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002) and are perceived as unfamiliar and alien-sounding (Galbraith, et al., 2004). They also convey less phonetic information and presumably very little lexical or semantic information (Binder, et al., 2000).

Table 13: Summary of  $d'$  values and standard errors of the subjects’ sensitivity for detecting normal warble calls and reversed warble calls. Comparison between conditions in the same species.

Warble call Targets	Normal		Reversed		
	$d'$	se	$d'$	se	
Budgerigar	2.21	0.07	4.54	0.19	*
Zebra finch	0.97	0.08	1.24	0.11	<i>n.s.</i>
Canary	1.19	0.09	0.88	0.15	<i>n.s.</i>

Here I show that reversed warble elements have an effect very similar to reversed words; they sound so “unusual” to budgerigars that the birds immediately detect the difference. Discriminability is significantly better in the reversed background than any other background condition (Table 12) because of the clear contrast to the normal, forward playing targets. However, there is no difference in sensitivity for temporally

reversed targets or forward normal targets do not seem to make a difference in other species (Figure 28; Table 13), which further shows the species-specific significance of warble calls.

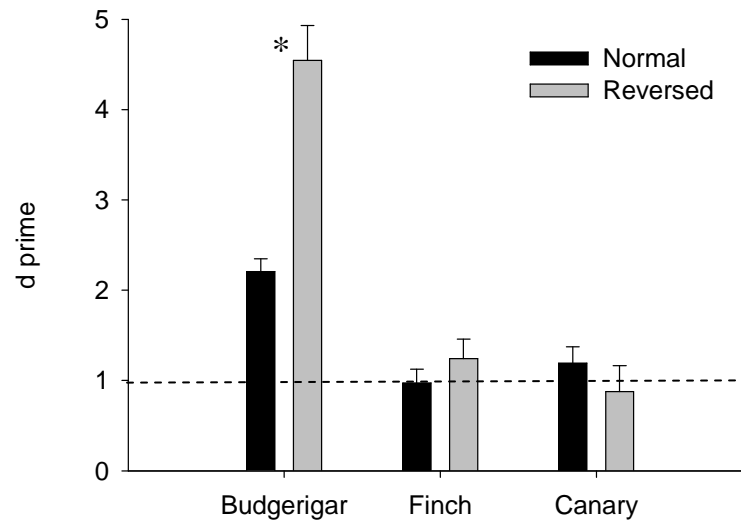


Figure 28: Cross-species comparison between detection of normal warble calls and temporally reversed warble calls. Error bars indicate the 95% confidence interval. Dashed line indicates  $d' = 1$ .

Budgerigar warble is so long and complicated that many complex features (e.g., spectrotemporal features of each element, the ordering of those elements, and the speed of vocal production) are delivered to the recipient at the same time. Information can be encoded in each of these variables and all of them conspire to make communication happen. Earlier experiments have shown that budgerigars perceive warble elements categorically (Chapter 4), and warble calls provide information of individual identity (Chapter 5). Here, this series of experiments further showed that the organization of warble elements alone can provide significant information to budgerigars ( $d'$  around 1)

but not other species ( $d'$  significantly less than 1). It remains to be seen what the relative role of the different acoustic characteristics of warble play in budgerigar vocal communication. While these experiments do not prove that budgerigars have “syntax” in their warble as humans do for speech, they do show that the warble song of this species certainly follows some patterns instead of being produced randomly, and individuals do pay attention to changes in the order of warble elements. The findings here open the door to deeper comparison between animal vocalization and human speech.

## Chapter 9: General Discussion

The long, rambling warble song of the budgerigar is an intriguing vocalization for its acoustic complexity, reproductive functions (Brockway, 1961, 1962, 1964b, 1965, 1969), and the association with intimate behaviors in very close quarters. However, very little is known about this vocalization. In fact, much of what we know is based on papers from more than thirty years ago (e.g., Brockway, 1964b, 1965, 1969; Gramza, 1970; Trillmich, 1976b; Wyndham, 1980).

Budgerigars are opportunistic breeders that reproduce whenever the condition is favorable. They form long-term pair bond that maintains in the flock throughout years so that once rainfall starts, creating suitable breeding conditions, they can begin to breed and reach reproductive peak only within a few days (Trillmich, 1976a, 1976c). Warble plays an important role in this scenario. Since both females and males themselves are sexually stimulated by the conspecific warble (Brockway, 1962, 1965, 1967a, 1969), it not only helps to maintain the pair bond, but also keeps the birds sexually ready to mate at any time. The unpredictability of rainfall generates the need to quickly coordinate the reproductive behaviors between pairs, which in turn puts selective pressure on warble, assuming the variability of warble is information-rich and can be modified readily according to the environment. More specifically, it leads to the speculation that the amount of information coded in the elements and arrangements of elements in this complex vocalization might be unusual for an animal vocalization.

Through a series of acoustic analyses and perceptual testing, I provide a comprehensive view of budgerigar warble, advancing our fundamental knowledge of this biologically important vocalization. Moreover, the selection of tests used in assessing the

perceptual relevance of element ordering in warble affords new points of contact with other acoustically complex vocal streams such as human speech.

*Acoustic perceptual categories for communication signals*

Categorization is defined as the process in which ideas or objects are sorted according to their perceived similarity (Horn & Falls, 1996; Pothos & Chater, 2002). This is considered by some to be a basic step in communication where senders and receivers share the code to exchange information (Horn & Falls, 1996; Seyfarth & Cheney, 2003; Smith, 1977).

Budgerigars perceive their extremely long and variable warble songs as discrete components belonging to seven basic acoustic-perceptual categories that exist across all warble recordings from different birds, establishing the foundation of budgerigar vocal communication. These categories are the building blocks of warble across individuals of this species, and they can be arranged in various ways to create all the variations seen in warble. These findings may provide a new animal parallel to human speech where words are strung together in different orders to make different sentences.

It is important to distinguish between the strict definition of categorical perception of speech sound and the perception of acoustic categories demonstrated here. Proof of categorical perception of speech sounds in speech requires a peak in discriminability at the category boundary separating two speech sound categories (e.g., /da/ vs. /ta/) as well a demonstration that listeners ignore variation within a category when labeling or identifying the same stimuli (Liberman, Harris, Hoffman, & Griffith, 1957; Macmillan, Kaplan, & Creelman, 1977). The perception of acoustic categories demonstrated here for budgerigars might more accurately parallel the perception of different vowels spoken by

different speakers in human speech. Early work with humans, chinchillas (Burdick & Miller, 1975), and even budgerigars (Dooling & Brown, 1990) has shown that these species can all focus on the many relevant acoustic features that define vowel categories in natural speech and learn to ignore the irrelevant acoustic features (e.g. voice characteristics) that distinguish one speaker from another. In other words, organisms can focus on relevant features that define acoustic categories and ignore irrelevant features that do not. However, unlike the demonstration using one dimension in human speech (e.g., /pa/-/ba/ distinctions along variations of voice onset time (Pickett, 1999)), the parsing of vocalizations by budgerigars cannot be simply explained by variations in a single acoustic feature. This is because the stimuli and the task adopted in this thesis did not restrict the variations to only one dimension, and the birds were allowed to discriminate the targets naturally. Different acoustic features are often not orthogonal but interact with each other (Horn & Falls, 1996). It is the gross combination of all dimensions that the birds were categorizing.

To truly investigate categorical perception in the budgerigars in a way that is analogous to that in human speech, one would have to construct a series of synthetic stimuli similar to natural warble elements but only differed along one acoustic dimension. Moreover, identification tasks would also have to be used to show that the birds not only have more difficulty discriminating stimuli from the same category, but are able to explicitly label two categories of sounds by variations of one single acoustic feature.

Behavioral assessments showing that animals categorize their own vocalizations are paralleled by physiological evidence for perceptual categories. For example, swamp sparrows have been observed to react categorically toward changes in the duration of a



certain note in their song. Recordings from a single sensorimotor neuron in HVC showed that those neurons respond categorically to changes in note duration as well, and the boundary coincides nicely with that found earlier through behavioral study (Prather, Nowicki, Anderson, Peters, & Mooney, 2009). Since now it has been shown that warble elements can be acoustically and perceptually categorized into seven basic groups, it would be an interesting future study to see if budgerigars have some “categorical neural pathways” corresponding to these groups.

#### Contact calls vs. warble calls

Because of their prevalence, particular attention was paid to the relation between the elements of the most common category (warble calls) and contact calls produced as single utterances. A close examination found that they are both perceptually and acoustically different from each other, supporting the notion that they are two different vocal systems.

Functionally, contact calls are mainly used for social coordination (Farabaugh, et al., 1994), whereas warble, as a whole vocalization, is important in budgerigar reproduction (Brockway, 1969). Detailed analysis has shown that certain low-pitched warble elements are especially effective in budgerigar courtship (Brockway, 1961, 1962, 1969), but whether each element category has its own specific function is still unclear. In humans, substituting words in a running sentence with the same words produced alone does not change the overall “meaning” of that sentence, even though the substitutions are obviously distinguishable. Knowing that contact calls can be easily detected in warble, it would be interesting to examine whether replacing all the warble calls in a warble

sequence with contact calls would change its significance or effectiveness in stimulating reproduction readiness by females.

This experiment also showed that budgerigars can extract the “individual voice characteristics” from each vocalization and can presumably use this information to identify individuals. Voice characteristics can be seen as unique spectral features that belong to one individual, probably resulted from individual variations in the vocal production apparatus. Great tits (Weary & Krebs, 1992) and European starlings (Gentner & Hulse, 1998) have been shown identify individuals by their voice characteristics, but there is not much evidence in the use of individual voice characteristics in animals other than humans. The finding here further show the deep parallels between budgerigar warble and human speech.

#### *The structure of budgerigar warble and its biological importance*

The Markov chain analysis and the perceptual follow-up experiments provided convergent evidence that the signalers (usually males) use at least the 5<sup>th</sup> order Markovian structures to construct their complex and variable warble song, and the receivers (usually females) have an attention span of approximately 5 warble elements. In some ways this may be analogous to the magic number seven, plus or minus two, used to describe human memory span (Miller, 1956).

In most songbirds studied so far, females tend to choose males with more complex vocalizations (Searcy & Yasukawa, 1996), but it is still unclear how females choose a mate or what features they prefer in such a long vocalization. Perhaps the magic number five in warble is the key.

Given the complex nature of warble, it is possible that female budgerigars become overwhelmed and lose track of the entire warble bouts. Therefore, they instead only focus on the occurrence of certain “attractive units,” probably in the order of 5 elements. To test this hypothesis, warble sequences of lower order ( $< 5$ ) Markov chain and higher order ( $> 5$ ) Markov chain can be artificially generated and played back to females. Their response, especially the tendency to copulate and/or the degree of ovarian development, can be recorded and compared to the effect of listening to the natural warble of 5<sup>th</sup> order Markov chain. Such an experiment may show that a 5<sup>th</sup> order Markovian chain may be the optimal structure that females favor, where either higher order or lower order is less attractive to them.

Alternatively, females may prefer males with warble of a Markov chain higher than 5<sup>th</sup> order. Male budgerigars may focus on improvising those “attractive units,” but the heavy neurological load constrains them from developing a higher order Markovian structure. How energetic it is to produce warble and whether some neurological bottleneck is indeed the limit factor on warble structure are unknown but can be approached by physiological methods. Since the neural pathway of warble production has been mapped out (Heaton & Brauth, 2000; Jarvis & Mello, 2000), it may be possible to find a brain region, a neural group, or even a neuron that actively controls the complexity of warble, particularly the order of the Markovian structures.

#### *The salience of sequence cues in warble*

The ability of budgerigars to detect insertions of warble calls from a background of natural warble sequences of the same individual established that the ordering of elements in warble is important. The paradigm used in this dissertation provides a chance

to approach interesting questions, such as budgerigars' sensitivity to sequential cues, but it also has some drawbacks. For example, the place where the insertion occurred cannot be controlled for, making it difficult to estimate how important the element ordering is. Some local structures of warble may be so "unbreakable" that any insertion of an element of any category is easy to detect. Some parts of the warble may be loosely constructed that all insertions are allowed or omitted and do not make any difference perceptually. This technical problem needs to be modified in order to generate more specific data in the future.

Moreover, it is still unclear whether the "sequential cues" that budgerigars used to detect the insertions are specific rules based on which warble elements are arranged, or more general melodic patterns that sound harmonic to budgerigars. Pilot data showed that budgerigars were not able to detect an out-of-order word in a repeating 4-word English speech ( $d' = 0.32$ ). Interestingly, when testing budgerigars on detecting an out-of-order note in a repeating 6-note pure tone melody (the first 6 notes of a happy birthday song), they did better than detecting English (intonation language) speech ( $d' = 0.67$ ). Like most bird songs, warble is rich in frequency modulation and sounds like a melody. Perhaps what budgerigars are learning is the pattern in the sequence as a whole, rather than the particular features of each sound and their relative ordering. Future tests on a tone language (like Chinese) and/or a song with lyrics may help clarify this interesting hypothesis.

Regardless of whether warble is rule-governed or pattern-based, budgerigars' sequencing ability has been shown to be highly species-specific. Zebra finches and canaries failed to detect same-individual warble call insertions showing both high false

alarm rates and low hit rates. Furthermore, the pilot data described above also indicate that this ability is highly warble-specific. Budgerigars fail to perform the task when the stimuli were human speech made up of English words. Moreover, budgerigars were extremely proficient at detecting insertions of normal warble calls in a naturally-sequenced background of temporally reversed warble elements. It is not hard to imagine a speech parallel to these experiments where humans are asked to pick a normal word out of reverse speech or a reversed word out of a normal speech sequence – both tasks should be quite easy if speech is in the listeners' native tongue. All of these results support the notion that warble is very unique to budgerigar vocal communication, and budgerigars have evolved a specialized system to perceive and process their own vocalization.

#### *Relevance of present studies for animal models of human speech communication*

Most approaches to understanding human speech using animals rely on non-natural stimuli, artificial training, and unusual task demands that fall well outside the animal's usual repertoire. For example, many studies have tested animals on human speech sounds (e.g., Burdick & Miller, 1975; Dooling & Brown, 1990; Kluender, et al., 1987) or trained them to learn unique features of human language (e.g., Fitch & Hauser, 2004; Gentner, et al., 2006) or human speech itself, usually through modalities other than vocalization, such as sign language (e.g., Herman, et al., 1984; Pepperberg, 1992; Savage-Rumbaugh, et al., 1986). Usually, intensive training is required before the subject animals start to perform as if they understand what the task is (Corballis, 2007), indicating that their response is an unusual demand away from their natural behavior.

The complexity and intimate delivery of budgerigar warble and the fact that it contains learned components, conjures up the parallels with aspects of human speech.

Without the issues of over-training or “forcing” the animals to perform unnaturally, warble offers a completely natural system for the study of animal vocal communication, especially serial order learning in non-human animals. I have showed that budgerigars form acoustic perceptual categories of warble elements, and that budgerigars not only combine their warble elements in a non-random fashion, but also attend to the sequential order of these elements, which reinforces the parallels between the vocal communication systems of humans and budgerigars. Moreover, the techniques used here are appropriate for investigating budgerigars’ warble. Artificial training was not necessary in most of the experiments, allowing budgerigars to respond with their instinctive ability.

### Conclusions

In conclusion, the work completed in this dissertation has established the foundation of research in budgerigar warble. This is the first study that systematically analyzed large amounts of warble. This resulted in the discovery of seven perceptual categories of warble elements that serve as the building blocks of warble across individuals and that are arranged in non-random ways to create variations in warble. This is also the first study that psychophysically tested budgerigars’ perception of the ordering of their own warble. Finally, our knowledge of this highly variable vocalization has now advanced considerably, but there are still many questions. Future studies should focus on creating a complete understanding of this natural system parallel to human speech, particularly on the mechanism of warble element categorization, the functional relevance of the structure of warble, and the rules, if any, or simply the patterns that budgerigars adopt to produce warble.

# Appendix I: Measures for the Classifier

*By Dr. E. W. Smith*

*(C-CEBH Research Engineer)*

## Introduction

The Multiple Neural Networks using Majority Vote Classifier (APPENDIX II) is a measure-based or feature-based classifier. To classify a warble segment into a group number, a warble segment is first analyzed and a numeric value for each of the twenty measures is computed. These twenty numbers are then input to the classifier. This appendix describes each of the measures in detail.

## Features

Because the features in warble to which birds attend is a subject of past and present research (Nelson, 1989), a minimally sufficient or otherwise optimal feature set is not known. For this dissertation research, members of the feature set were selected with the intention of having a broad extent in the space of plausibly appropriate features. Twenty different features were used for input to the classifier (Table 2). It is accepted that choosing measures this way leads to redundancy within the measures.

## **Spectral Roughness**

Spectral roughness is an attempt to capture the amount of variation in the spectral content of a segment. More specifically, it is an attempt to capture how often, in the frequency domain, the power spectrum exceeds a smoothed local power spectrum. First,

short-term power spectra (STPS) are calculated (footnote 1) for the segment, and a fifth order interpolating polynomial is made for each STPS. Residual power spectra (STPS minus the corresponding interpolating polynomial values) are computed. The number of frequency points at which each STPS exceeds its polynomial approximation is counted. That total count, divided by the number of five-millisecond subsegments in the segment, gives the Spectral Roughness for the segment. Computed this way, Spectral Roughness is unitless.

### **Tonality**

Tonality, or spectral purity, provides an indication of the extent to which a segment is a pure tone at each moment in time. More specifically, tonality is defined as a normalized sum of selected residuals. The Tonality measure makes use of the same STPSs and residual power spectra calculated for Spectral Roughness. In the Tonality measure, the single largest values in each residual power spectrum are added together. Dividing that sum by the number of STPSs in the segment gives the value for Tonality. Loosely, the units of Tonality are dBs.

### **Duration**

Duration is the temporal extent of a segment, expressed in milliseconds.

### **Harmonic Strength**

Harmonic Strength is a measure of the extent to which a segment is composed of tonal lines. An array of STPSs is constructed for the segment, and all of the peaks are identified in each STPS. Same-frequency-bin peaks and adjacent-frequency-bin peaks in



at least two temporally-adjacent STPSs are taken to be spectral lines. The total power in the identified spectral lines is summed, and then divided by the total power in the segment to give Harmonic Strength. Computed this way, Harmonic Strength is a number between zero and one.

### **Frequency Quartiles**

1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> Frequency Quartiles provide the classifier with a gross description of the spectrum of a segment. It is based on the power spectrum of a segment. First the integral of the power spectrum (starting at zero Hertz continuing to the Nyquist frequency) is computed, and then the total spectral power of the segment is computed. To identify the 1<sup>st</sup> Frequency Quartile, the first point of the integrated spectrum that equals or exceeds one fourth of the total spectral power is identified. The frequency associated with that point is the 1st Frequency Quartile. For the 2<sup>nd</sup> Frequency Quartile, the lowest frequency containing half or more than half of the total power is identified. For the 3<sup>rd</sup> Frequency Quartile, the lowest frequency containing at least three fourths of the total power is identified. The units for Frequency Quartiles are Hertz.

### **Skewness of Power**

Skewness is a common statistic. Here, it is based on the distribution of the power in a segment over time. It is the value of the skewness function (Matlab), applied to the power of each sample in the segment. Skewness is a unitless number.

### **Zero-Crossing Frequency**

Zero-Crossing Frequency is an estimate of the average frequency of a signal. It is somewhat abstract, because a segment may have very little power near the Zero-Crossing Frequency. To compute it, the number of times the segment crosses zero on the Y-axis is divided by twice the segment duration. The units of Zero-Crossing Frequency are Hertz.

### **Average Peak Spacing**

Average Peak Spacing provides information about how far apart local frequency maxima in a segment are. It is the average distance between local maxima in the STPSs. The units of Average Peak Spacing are Hertz.

### **Amplitude Modulation**

Amplitude Modulation provides a notion of the short-term temporal variation in a segment's amplitude. The segment is divided into twenty subsegments and the power in each segment is computed. The digital derivative of the twenty-element power sequence is computed, and the absolute values of the differences are averaged. The units of Amplitude Modulation measure are Hertz.

### **Number of Harmonic Lines**

Number of Harmonic Lines is an attempt to describe the tonal complexity of a segment. Tonal lines are identified with the same scheme as in the Harmonic Strength measure. The number total of lines is counted, and the Number of Harmonic Lines measure is taken to be the mode of these counts. This measure is unitless.

### **Frequency of Maximum Amplitude**

Frequency of Maximum Amplitude is based on the power spectrum of an entire segment. The power spectral bins are searched to find the single bin containing the most power, and the center frequency of this bin is the Frequency of Maximum Amplitude. There could be cases in which most of the total segment power is not close to this value. The units of Frequency of Maximum Amplitude are Hertz.

### **80% Bandwidth**

80% Bandwidth is based on the power spectrum of an entire segment. To compute it, the minimum number of power spectral bins which together contain 80% of the total segment power is counted. The spectral bins contributing to the 80% power count need not be contiguous. This bin count is then multiplied by the spectral width of one FFT bin, so the value of 80% Bandwidth is a frequency of less than the Nyquist frequency (24000 Hertz). The units of 80% Bandwidth are Hertz.

### **Entropy**

Entropy is based directly on the digital samples of the segment. First, the ratio of the geometric mean of the absolute values to the arithmetic mean of the absolute values is computed. Entropy is the base e logarithm of this ratio, and it is unitless.

### **Time to Peak Amplitude**

Time to Peak Amplitude is a temporal measure. The number of points coming before the largest (absolute value) sample is counted. That count is divided by the

sampling frequency to provide the value for the measure. After multiplying by 1000, the units are milliseconds.

### **Kurtosis of Power**

Like Skewness of Power, Kurtosis of Power is based on the distribution of the power in a segment. It is the value of the kurtosis function (Matlab), applied to the power of each sample in the segment. Kurtosis is unitless.

### **Frequency Modulation**

Frequency Modulation provides a feel for how much a signal varies in frequency over time. It is computed by dividing a segment into twenty equal-length subsegments, and identifying the single peak-amplitude frequency bin in the FFT of each subsegment. The absolute differences of the center frequencies of these peak bins are summed, and divided by twenty. Conceptually, the units are Hertz.

### **Standard Deviation of Power**

Standard Deviation of Power is based on the distribution of the power in a segment. It is the value of the std function (Matlab), applied to the power of each sample in the segment.

### **Average Power Per Sample**

Together with Standard Deviation of Power, Average Power Per Sample provides the classifier with basic statistics about the segment. It is the value of the mean function

(Matlab), applied to the power of each sample in the segment. The units are power (volts squared).

#### Footnote

### **Computing the Short-Term Power Spectrum (STPS)**

Several of the measures are based on arrays of short-term power spectra. One STPS represents the average spectrum of a segment for a five millisecond period. Given the sampling rate of 48,000 Hertz, five milliseconds contain 240 samples. With conventional FFT analysis, this provides short-term frequency resolution of 186 Hertz.

Arrays of STPSs are derived from a complete segment. To calculate the array of short-term power spectra for a segment, the segment is first divided into 50% overlapping, five millisecond subsegments. For example, a segment lasting for 200 milliseconds would be converted into an array of 79 STPSs.

## Appendix II: Multiple Neural Networks using Majority Vote Classifier

*By Dr. E. W. Smith*

*(C-CEBH Research Engineer)*

This appendix describes design and details of the neural network based automatic warble segment classifier.

### Objective

The objective is to construct an automatic classifier for warble segments, which can be trained using segments already classified into one of a handful of groups by humans. In AI, this is generally referred to as a problem in supervised machine learning.

### The classifier is based on neural networks (NNs)

For each segment, twenty features (APPENDIX I) are directly available to the classifier. NNs are a reasonable approach to this classification problem. The well-known back propagation scheme for training NNs is directly applicable. During classifier construction, classified segments (values of the twenty measures and the corresponding human-assigned group numbers) will serve as training cases for the classifier. During operation, the classifier will estimate which group a segment belongs to based on the values of the twenty measures.

### Large and redundant feature set

It is known that increasing the size of a feature set, or the presence of redundant features, can adversely affect classifier performance (Devijver & Kittler, 1982). The structure of the classifier and the way the classifier is used were formulated in an attempt to ameliorate performance.

### Approach

In AI terms, the classifier developed is referred to as a *Multiple Classifier System, using Majority Vote* (Kittler & Alkoot, 2003). Specifically, the classifier comprises twenty five separate, heterogeneous, feed-forward NNs. Together, the NNs are referred to as a *committee*. The potentially problematic feature set is addressed in three ways. First, the majority vote scheme described below is resilient to small numbers of misclassifications. Second, the structural diversity between the individual NNs acts to reduce bias potentially caused by NN structure. Third, in computing weights for the individual NNs makes use of over learning resistance provided by Matlab.

### Classifier construction

Construction of the classifier begins by defining the different NN structures. Each NN will have the same number of input nodes and output nodes; one input node for each feature value, and one output node for each of the eight mutually exclusive warble groups. The structure and size of the hidden layers is varied between the NNs. Thirteen NNs will have one hidden layer, and thirteen NNs have two hidden layers. Further, hidden layer size (the number of neurons in each hidden layer) is varied from eight to thirteen. This gives a total of twenty six different NN structures.

To assign the weights for links within each NN, the Matlab functions *dividevec()* and *train()*, are used to train each network using the *training set*, *validation set*, and *test set* approach (Matlab). In order to increase the likelihood of each NN being accurate, a number of separate attempts are made to train each NN. For each of the twenty six NN structures, the training process is performed eighty times, and the most accurate set of computed weights (accurate, in terms of training set classification accuracy) is kept and used for that NN. Lastly, the single NN with the lowest performance on the complete training set is completely removed from the committee. The remaining 25 NNs, each with a different structure, form the committee for the classifier.

#### Classifier operation

During use, the twenty feature values are computed for a segment to be classified. These twenty values are then used as input to each of the 25 NNs on the committee. Each NN produces one output – a preliminary classification.

Next, in the majority voting stage, the 25 preliminary classifications are interpreted as being votes, and the classification (the group number) with the most votes is output as the final classifier output for that warble segment. The developed software includes, as an input parameter, the minimum number of agreeing votes in order for a final classification to be considered valid. If the number votes for the winning classification does not equal or exceed that parameter, then the classifier outputs a “cannot classify” signal. In this dissertation research, however, the decision was made to set the limit to one, so the classifier would always output a valid group number.



### *One competing NN classifier approach*

One common strategy for dealing with large feature sets is to divide the feature set into a number of feature subsets. Then, separate NNs are built, each using only one feature subset. In some cases, this method provides a more computationally efficient majority vote classifier, and for some cases, it is shown to perform better than a committee of identical-structure NNs. The feature subset strategy was not appropriate for this dissertation research, because a secondary goal was to rank the relative merit of each feature (see the knock-out study section). In the feature subset approach, different NNs have different structures, so it is difficult or impossible to be confident that the resulting committee does not bias some features over other features. The approach taken here provides for a straightforward implementation and interpretation of the knock-out study.

## Bibliography

- Adret, P. (1993). Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Animal Behaviour*, 46, 149-159.
- Ali, N. J., Farabaugh, S. M., & Dooling, R. J. (1993). Recognition of contact calls by the budgerigar (*Melopsittacus undulatus*). *Bulletin of the Psychonomic Society*, 31, 468-470.
- Armstrong, T. A. (1992). Categorization of notes used by female Red-winged Blackbirds in composite vocalizations. *The Condor*, 94, 210-223.
- Balaban, E. (1988). Bird song syntax: Learned intraspecific variation is meaningful. *Proceedings of the National Academy of Sciences of the United States of America*, 85, 3657-3660.
- Balaban, E., Teillet, M.-A., & Douarin, N. L. (1988). Application of the quail-chick chimera system to the study of brain development and behavior. *Science*, 241, 1339-1342.
- Baptista, L. F. (1996). Nature and its nurturing in avian vocal development. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 39-60). Ithaca, NY: Cornell University Press.
- Baptista, L. F., & Petrinovich, L. (1986). Song development in the white-crowned sparrow: social factors and sex differences. *Animal Behaviour*, 34, 1359-1371.
- Bates, E. (2003). On the nature and nurture of language. In E. Bizzi, P. Calissano & V. Volterra (Eds.), *Frontiere della biologia [Frontiers of biology]. Il cervello di Homo sapiens [The brain of homo sapiens]*. (pp. 241-265). Rome: Giovanni Trecanni.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512-528.
- Bloomfield, L. L., Charrier, I., & Sturdy, C. B. (2004). Note types and coding in parid vocalizations. II: The chick-a-dee call of the mountain chickadee (*Poecile gambeli*). *Canadian Journal of Zoology*, 82, 780-793.
- Bloomfield, L. L., Phillmore, L. S., Weisman, R. G., & Sturdy, C. B. (2005). Note types and coding in parid vocalizations. III: The chick-a-dee call of the Carolina chickadee. *Canadian Journal of Zoology*, 83, 820-833.

- Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, 417, 351-358.
- Brenowitz, E. A., & Kroodsma, D. E. (1996). The neuroethology of birdsong. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 285-304). Ithaca, NY: Cornell University Press.
- Brittan-Powell, E. F. (2002). *Development of auditory sensitivity in budgerigars*. Unpublished Dissertation, University of Maryland, College Park.
- Brittan-Powell, E. F., Dooling, R. J., & Farabaugh, S. M. (1997). Vocal development in budgerigars (*Melopsittacus undulatus*): Contact calls. *Journal of Comparative Psychology*, 111, 226-241.
- Brittan-Powell, E. F., Dooling, R. J., Larsen, O. N., & Heaton, J. T. (1997). Mechanisms of vocal production in budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, 101, 578-589.
- Brockway, B. F. (1961). The effects of visual and vocal stimuli upon the reproductive ethology and physiology of paired Budgerigars (*Melopsittacus undulatus*). *American Zoologists*, 1, 346-347.
- Brockway, B. F. (1962). Investigations of the auditory stimuli for laying in Budgerigars (*Melopsittacus undulatus*). *American Zoologists*, 2, 508-509.
- Brockway, B. F. (1964a). Ethological studies of the budgerigar: non-reproductive behavior. *Behaviour*, 22, 193-222.
- Brockway, B. F. (1964b). Ethological studies of the budgerigar: reproductive behavior. *Behaviour*, 23, 294-324.
- Brockway, B. F. (1964c). Social influences on reproductive physiology and ethology of budgerigars (*Melopsittacus undulatus*). *Animal Behaviour*, 12, 493-501.
- Brockway, B. F. (1965). Stimulation of ovarian development and egg laying by male courtship vocalization in budgerigars (*Melopsittacus undulatus*). *Animal Behaviour*, 13, 575-578.
- Brockway, B. F. (1967a). The influence of vocal behavior on the performer's testicular activity in budgerigars (*Melopsittacus undulatus*). *Wilson Ornithology Society*, 79, 328-334.
- Brockway, B. F. (1967b). Interactions among male courtship warbling, photoperiodic and experiential factors in stimulating the reproductive activity of female budgerigars (*Melopsittacus undulatus*). *American Zoologists*, 7, 215-216.
- Brockway, B. F. (1968). Influences of sex hormones on the loud and soft warbles of male budgerigars. *Animal Behaviour*, 16, 5-12.

- Brockway, B. F. (1969). Roles of budgerigar vocalization in the integration of breeding behavior. In R. A. Hinde (Ed.), *Bird Vocalizations* (pp. 131-158). London: Cambridge University Press.
- Brown, S. D., Dooling, R. J., & O'Grady, K. (1988). Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): III. Contact calls. *Journal of Comparative Psychology*, 102, 236-247.
- Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*, 58, 415-427.
- Butler, A. B., & Hodos, W. (2005). *Comparative Vertebrate Neuroanatomy: Evolution and Adaptation* (2 ed.). Hoboken, NJ: John Wiley & Sons.
- Catchpole, C. K., & Slater, P. J. B. (2008). *Bird Song: Biological Themes and Variations* (2 ed.). New York: Cambridge University Press.
- Charrier, I., Bloomfield, L. L., & Sturdy, C. B. (2004). Note types and coding in parid vocalizations. I: The chick-a-dee call of the black-capped chickadee (*Poecile atricapillus*). *Canadian Journal of Zoology*, 82, 769-779.
- Chatfield, C., & Lemon, R. E. (1970). Analysing sequences of behavioural events. *Journal of Theoretical Biology*, 29, 427-445.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science*, 31, 697-704.
- Dawson, M. R. W., Charrier, I., & Sturdy, C. B. (2006). Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types. *Journal of the Acoustical Society of America*, 119, 3161-3172.
- Deecke, V. B., & Janik, V. M. (2006). Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls. *Journal of the Acoustical Society of America*, 119, 645-653.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298, 2013-2015.
- Devijver, P. A., & Kittler, J. (1982). *Pattern recognition: A statistical approach*. Englewood Cliffs, NJ: Prentice Hall International.
- Dobson, C. W., & Lemon, R. E. (1979). Markov sequences in songs of American thrushes. *Behaviour*, 68, 86-105.

- Dooling, R. J., Best, C. T., & Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taenipygia guttata*). *Journal of the Acoustical Society of America*, 97, 1839-1846.
- Dooling, R. J., & Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels. *Perception & Psychophysics*, 47, 568-574.
- Dooling, R. J., & Brown, S. D. (1992). Auditory perception of conspecific and heterospecific vocalizations in birds: Evidence for special processes. *Journal of Comparative Psychology*, 106, 20-28.
- Dooling, R. J., Brown, S. D., Park, T. J., Okanoya, K., & Soli, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): I. Pure tones. *Journal of Comparative Psychology*, 101, 139-149.
- Dooling, R. J., & Okanoya, K. (1995a). The method of constant stimuli in testing auditory sensitivity in small birds. In G. M. Klump, R. J. Dooling, R. R. Fay & W. C. Stebbins (Eds.), *Methods in Comparative Psychoacoustics* (pp. 161-169). Basel, Switzerland: Birkhäuser Verlag.
- Dooling, R. J., & Okanoya, K. (1995b). Psychophysical methods for assessing perceptual categories. In G. M. Klump, R. J. Dooling, R. R. Fay & W. C. Stebbins (Eds.), *Methods in Comparative Psychoacoustics* (pp. 307-318). Basel, Switzerland: Birkhäuser Verlag.
- Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced-voiceless distinction. *Perception & Psychophysics*, 46, 65-71.
- Dooling, R. J., Park, T. J., Brown, S. D., Okanoya, K., & Soli, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals. *Journal of Comparative Psychology*, 101, 367-381.
- Dooling, R. J., & Saunders, J. C. (1975). Hearing in the parakeet (*Melopsittacus undulatus*): Absolute thresholds, critical ratios, frequency difference limens, and vocalizations. *Journal of Comparative and Physiological Psychology*, 88, 1-20.
- Dooling, R. J., & Searcy, M. (1980). Early perceptual selectivity in the swamp sparrow. *Developmental Psychobiology*, 13, 499-506.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567-631.
- Eda-Fujiwara, H., & Okumura, H. (1992). The temporal pattern of vocalizations in the budgerigar *Melopsittacus undulatus*. *Journal of the Yamashina Institute for Ornithology*, 24, 18-30.

- Eens, M., Pinxten, R., & Verheyen, R. F. (1992). Song learning in captive European starlings, *Sturnus vulgaris*. *Animal Behaviour*, 44, 1131-1143.
- Farabaugh, S. M., Brown, E. D., & Dooling, R. J. (1992). Analysis of warble song of the budgerigar *Melopsittacus undulatus*. *Bioacoustics*, 4, 111-130.
- Farabaugh, S. M., & Dooling, R. J. (1996). Acoustic communication in parrots: Laboratory and field studies of budgerigars, *Melopsittacus undulatus*. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 97-120). Ithaca, NY: Cornell University Press.
- Farabaugh, S. M., Linzenbold, A., & Dooling, R. J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108, 81-92.
- Ficken, M. S., & Popp, J. W. (1992). Syntactical organization of the gargle vocalization of the black-capped Chickadee, *Parus atricapillus*. *Ethology*, 91, 156-168.
- Fischer, J. (1998). Barbary macaques categorize shrill barks into two call types. *Animal Behaviour*, 55, 799-807.
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 377-380.
- Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: a case of language acquisition beyond the "critical period". *Brian and Language*, 1, 81-107.
- Galbraith, G. C., Amaya, E. M., Rivera, J. M. D. d., Donan, N. M., Duong, M. T., Hsu, J. N., et al. (2004). Brain stem evoked response to forward and reversed speech in humans. *Neuroreport*, 15, 2057-2060.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440, 1204-1207.
- Gentner, T. Q., & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour*, 56, 579-594.
- Gil, D., & Slater, P. J. B. (2000). Song organisation and singing patterns of the willow warbler, *Phylloscopus trochilus*. *Behaviour*, 137, 759-782.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100, 8030-8035.
- Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178-200.

- Gottman, J. M., & Roy, A. K. (1990). *Sequential Analysis: A Guide for Behavioral Researchers*. New York: Cambridge University Press.
- Gramza, A. F. (1970). Vocal mimicry in captive budgerigars (*Melopsittacus undulatus*). *Zeitschrift für Tierpsychologie*, 27, 971-998.
- Haesler, S., Rochefort, C., Georgi, B., Licznarski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology*, 1-13.
- Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1985). The "chick-a-dee" calls of *Parus atricapillus*: a recombinant system of animal communication compared with written English. *Semiotica*, 56, 191-224.
- Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1987). Constraints on the structure of combinatorial "chick-a-dee" calls. *Ethology*, 75, 62-80.
- Hashino, E., & Okanoya, K. (1989). Auditory sensitivity in the zebra finch (*Poephila guttata castanotis*). *Journal of the Acoustical Society of Japan (E)*, 10, 51-52.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1569-1579.
- Heaton, J. T., & Brauth, S. E. (1999). Effects of deafening on the development of nestling and juvenile vocalizations in budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 113, 314-320.
- Heaton, J. T., & Brauth, S. E. (2000). Effects of lesions of the central nucleus of the anterior archistriatum on contact call and warble song production in the budgerigar (*Melopsittacus undulatus*). *Neurobiology of Learning and Memory*, 73, 207-242.
- Heaton, J. T., Dooling, R. J., & Farabaugh, S. M. (1999). Effects of deafening on the calls and warble song of adult budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, 105, 2010-2019.
- Heaton, J. T., Farabaugh, S. M., & Brauth, S. E. (1995). Effect of syringeal denervation in the budgerigar (*Melopsittacus undulatus*): The role of the syrinx in call production. *Neurobiology of Learning and Memory*, 64, 68-82.
- Herman, L. M., Kuczaj, S. A., & Holder, M. D. (1993). Responses to anomalous gestural sequences by a language-trained dolphin: evidence for processing of semantic relations and syntactic information. *Journal of Experimental Psychology: General*, 122, 184-194.
- Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, 16, 129-219.

- Hile, A. G., Plummer, T. K., & Striedter, G. F. (2000). Male vocal imitation produces call convergence during pair bonding in budgerigars, *Melopsittacus undulatus*. *Animal Behaviour*, 59, 1209-1218.
- Hile, A. G., & Striedter, G. F. (2000). Call convergence within groups of female budgerigars (*Melopsittacus undulatus*). *Ethology*, 106, 1105-1114.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099-3111.
- Horn, A. G., & Falls, J. B. (1996). Categorization and the design of signals: The case of song repertoires. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 121-135). Ithaca, NY: Cornell University Press.
- Hosino, T., & Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *Neuroreport*, 11, 2091-2095.
- Howell, P., & Archer, A. (1984). Susceptibility to the effects of delayed auditory feedback. *Perception & Psychophysics*, 36, 296-302.
- Hultsch, H., & Todt, D. (2004). Learning to sing. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 80-107). San Diego, CA: Elsevier Academic Press.
- Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016, 749-777.
- Jarvis, E. D., & Mello, C. V. (2000). Molecular mapping of brain areas involved in parrot vocal communication. *Journal of Comparative Neurology*, 419, 1-31.
- Jesteadt, W. (2005). The variance of d' estimates obtained in yes-no and two-interval forced choice procedures. *Perception & Psychophysics*, 67, 72-80.
- Jones, A. E., Ten Cate, C., & Bijleveld, C. C. J. H. (2001). The interobserver reliability of scoring sonagrams by eye: a study on methods, illustrated on zebra finch songs. *Animal Behaviour*, 62, 791-801.
- Kanwal, J. S., Matsumura, S., Ohlemiller, K., & Suga, N. (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *Journal of the Acoustical Society of America*, 96, 1229-1254.
- King, A. P., West, M. J., & Goldstein, M. H. (2005). Non-vocal shaping of avian song development: Parallels to human speech development. *Ethology*, 111, 110-117.
- King, A. S. (1989). Functional anatomy of the syrinx. In A. S. King & J. McLelland (Eds.), *Form and Function in Birds* (pp. 105-192). London: Academic Press.



- Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition* (pp. 173-204). Cambridge, UK: Cambridge University Press.
- Kittler, J., & Alkoot, F. M. (2003). Sum versus vote fusion in multiple classifier system. *IEEE*, 25, 110-115.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195-1197.
- Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift für Tierpsychologie*, 22, 770-783.
- Konishi, M., & Nottebohm, F. (1969). Experimental studies in the ontogeny of avian vocalizations. In R. A. Hinde (Ed.), *Bird Vocalizations: Their Relations to Current Problems in Biology and Psychology* (pp. 29-48). New York, NY: Cambridge University Press.
- Kroodsma, D. E. (1982). Learning and the ontogeny of sound signals in birds. In D. E. Kroodsma & E. H. Miller (Eds.), *Acoustic Communication in Birds* (Vol. 2, pp. 1-23). New York: Academic Press.
- Kuhl, P. K. (1994). Speech perception. In F. D. Minifie (Ed.), *Introduction to Communication Sciences and Disorders* (pp. 77-148). San Diego, CA: Singular Publishing Group.
- Kuhl, P. K., Kiritani, S., Deguchi, T., & Hayashi, A. (1997). Effects of language experience on speech perception: American and Japanese infants' perception of /ra/ and /la/. *Journal of the Acoustical Society of America*, 102, 3135.
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425-2438.
- Leger, D. W. (2005). First documentation of combinatorial song syntax in a suboscine passerine species. *The Condor*, 107, 765-774.
- Lemon, R. E., & Chatfield, C. (1971). Organization of song in cardinals. *Animal Behaviour*, 19, 1-17.
- Lemon, R. E., Dobson, C. W., & Clifton, P. G. (1993). Songs of American redstarts (*Setophaga ruticilla*): sequencing rules and their relationships to repertoire size. *Ethology*, 93, 198-210.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology: General*, 54, 358-368.

- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773-1781.
- Lohr, B., & Dooling, R. J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 112, 30-47.
- Lucas, J. R., & Freeberg, T. M. (2007). "Information" and the chick-a-dee call: Communicating with a complex vocal system. In K. A. Otter (Ed.), *The Ecology and Behavior of Chickadees and Titmice* (pp. 199-214). New York: Oxford University Press.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, 84, 452-471.
- Manabe, K., Dooling, R. J., & Brittan-Powell, E. F. (2008). Vocal learning in Budgerigars (*Melopsittacus undulatus*): Effects of an acoustic reference on vocal matching. *Journal of the Acoustical Society of America*, 123, 1729-1736.
- Manabe, K., Sadr, E. I., & Dooling, R. J. (1998). Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard effect. *Journal of the Acoustical Society of America*, 103, 1190-1198.
- Marler, P. (1970a). Birdsong and speech development: Could there be parallels? *American Scientist*, 58, 669-673.
- Marler, P. (1970b). A comparative approach to vocal learning: song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, 71, 1-25.
- Marler, P. (2000). Origins of music and speech: Insights from animals. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The Origins of Music* (pp. 31-48). USA: Massachusetts Institute of Technology.
- Marler, P. (2004). Science and birdsong: the good old days. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 1-38). San Diego, CA: Elsevier Academic Press.
- Marler, P., & Peters, S. (1981). Birdsong and speech: Evidence for special processing. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech* (pp. 75-112). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Marler, P., & Peters, S. (1987). A sensitive period for song acquisition in the song sparrow, *Melospiza melodia*, a case of age-limited learning. *Ethology*, 76, 89-100.

- Marler, P., & Peters, S. (1989). Species differences in auditory responsiveness in early vocal learning. In R. J. Dooling & S. H. Hulse (Eds.), *The Comparative Psychology of Audition: Perceiving Complex Sounds* (pp. 243-273). Hillsdale, NJ: Erlbaum.
- Marler, P., & Sherman, V. (1983). Song structure without auditory feedback: Emendations of the auditory template hypothesis. *Journal of Neuroscience*, 3, 517-531.
- Martindale, S. (1980). A numerical approach to the analysis of Solitary Vireo songs. *The Condor*, 82, 199-211.
- Matlab. Neural Network Toolbox™ Software (Version R2007b). Natick, MA: The Mathworks Inc.
- Matlab. Signal Processing Toolbox™ Software (Version R2007b). Natick, MA: The Mathworks Inc.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, 63, 81-97.
- Mindlin, G. B., & Laje, R. (2005). *The Physics of Birdsong*. The Netherlands: Springer.
- Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour*, 109, 20-45.
- Nelson, D. A. (1989). The importance of invariant and distinctive features in species recognition of bird song. *The Condor*, 91, 120-130.
- Nelson, D. A., & Marler, P. (1993). Innate recognition of song in white-crowned sparrow: a role in selective vocal learning? *Animal Behaviour*, 46, 806-808.
- Nottebohm, F. (1972). The origins of vocal learning. *The American Naturalist*, 106, 116-140.
- Nottebohm, F., Nottebohm, M. E., & Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral Neural Biology*, 46, 445-471.
- Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature*, 325, 53-55.
- Nowicki, S., & Marler, P. (1988). How do birds sing? *Music Perception*, 5, 391-426.
- Okanoya, K., & Dooling, R. J. (1987). Hearing in Passerine and Psittacine birds: A comparative study of absolute and masked auditory thresholds. *Journal of Comparative Psychology*, 101, 7-15.

- Okanoya, K., & Dooling, R. J. (1988). Obtaining acoustic similarity measures from animals: A method for species comparisons. *Journal of the Acoustical Society of America*, 83, 1690-1693.
- Okanoya, K., & Dooling, R. J. (1991). Perception of distance calls by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*): Assessing species-specific advantages. *Journal of Comparative Psychology*, 105, 60-72.
- Oller, D. K., Eilers, R. E., & Bull, D. H. (1985). Prespeech vocalizations of a deaf infant: a comparison with normal metaphonological development. *Journal of Speech and Hearing Research*, 28, 47-63.
- Osmanski, M. S., & Dooling, R. J. (2009). The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, 126, 911-919.
- Park, T. J., & Dooling, R. J. (1985). Perception of species-specific contact calls by budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 99, 391-402.
- Park, T. J., & Dooling, R. J. (1986). Perception of degraded vocalizations by budgerigars (*Melopsittacus undulatus*). *Animal Learning & Behavior*, 14, 359-364.
- Pepperberg, I. M. (1988). An interactive modeling technique for acquisition of communication skills: separation of "labeling" and "requesting" in a Psittacine subject. *Applied Psycholinguistics*, 9, 59-76.
- Pepperberg, I. M. (1990). Referential mapping: a technique for attaching functional significance to the innovative utterances of an African grey parrot (*Psittacus erithacus*). *Applied Psycholinguistics*, 11, 23-44.
- Pepperberg, I. M. (1992). Proficient performance of a conjunctive, recursive task by an African gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 106, 295-305.
- Pepperberg, I. M. (1997). Social influences on song acquisition of human-based codes in parrots and nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social Influences on Vocal Development* (pp. 157-177). New York: Cambridge University Press.
- Pepperberg, I. M. (2007). Grey parrots do not always "parrot": the roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Sciences*, 29, 1-13.
- Pickett, J. M. (1999). *The Acoustics of Speech Communication*. Needham Heights, MA: Allyn & Bacon.

- Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, 95, 201-236.
- Plummer, T. K., & Striedter, G. F. (2002). Brain lesions that impair vocal imitation in adult budgerigars. *Journal of Neurobiology*, 53, 413-428.
- Podos, J., & Nowicki, S. (2004). Performance limits on birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 318-342). San Diego, CA: Elsevier Academic Press.
- Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, 26, 303-343.
- Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience*, 12, 221-228.
- Ranjar, L., & Ross, H. A. (2008). Unsupervised bird song syllable classification using evolving neural networks. *Journal of the Acoustical Society of America*, 123, 4358-4368.
- Robinson, J. G. (1984). Syntactic structures in the vocalizations of wedge-capped capuchin monkey *Callicebus moloch*. *Behaviour*, 90, 46-79.
- Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, 115, 211-235.
- Savage-Rumbaugh, S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58(Nos. 3-4).
- Searcy, W. A., Podos, J., Peters, S., & Nowicki, S. (1995). Discrimination of song types and variants in song sparrow. *Animal Behaviour*, 49, 1219-1226.
- Searcy, W. A., & Yasukawa, K. (1996). Song and female choice. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds* (pp. 454-473). Ithaca: Cornell University Press.
- Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, 54, 145-173.
- Slater, P. J. B. (1983). Sequences of song in chaffinches. *Animal Behaviour*, 31, 272-281.
- Smith, W. J. (1977). *The behavior of communicating: an ethological approach*. Cambridge, MA: Harvard University Press.

- Snowdon, C. T. (1990). Language capacities of nonhuman animals. *American Journal of Physical Anthropology*, 33, 215-243.
- Soha, J. A., & Marler, P. (2001). Vocal syntax development in the white-crowned sparrow (*Zonotrichia leucophrys*). *Journal of Comparative Psychology*, 115, 172-180.
- Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6, 111-128.
- Stone, J. R., & Stoel-Gammon, C. (1994). Phonological development and disorders in children. In F. D. Minifie (Ed.), *Introduction to Communication Sciences and Disorders* (pp. 149-188). San Diego, CA: Singular Publishing Group.
- Striedter, G. F., Freibott, L., Hile, A. G., & Burley, N. T. (2003). For whom the male calls: an effect of audience on contact call rate and repertoire in budgerigars, *Melopsittacus undulatus*.
- Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, 291, 2564-2569.
- Todt, D. (2004). From birdsong to speech: a plea for comparative approaches. *Annals of the Brazilian Academy of Sciences*, 76, 201-208.
- Trillmich, F. (1976a). The influence of separation on the pair bond in budgerigars (*Melopsittacus undulatus*; Aves, Psittacidae). *Zeitschrift für Tierpsychologie*, 41, 396-408.
- Trillmich, F. (1976b). Learning experiments on individual recognition in budgerigars (*Melopsittacus undulatus*). *Zeitschrift für Tierpsychologie*, 41, 372-395.
- Trillmich, F. (1976c). Spatial proximity and mate-specific behaviour in a flock of budgerigars (*Melopsittacus undulatus*; Aves, Psittacidae). *Zeitschrift für Tierpsychologie*, 41, 307-331.
- Vu, E. T., Mazurek, M. E., & Kuo, Y.-C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. *Journal of Neuroscience*, 14, 6926-6934.
- Weary, D. M., & Krebs, J. R. (1992). Great tits classify songs by individual voice characteristics. *Animal Behaviour*, 43, 283-287.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.

- Werker, J. F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 25, 121-133.
- Westneat, M. W., Long, J. H., Jr., Hoese, W., & Nowicki, S. (1993). Kinematics of birdsong: Functional correlation of cranial movements and acoustic features in sparrows. *Journal of Experimental Biology*, 182, 147-171.
- Wilbrecht, L., & Nottebohm, F. (2003). Vocal learning in birds and humans. *Mental Retardation and Development Disabilities Research Reviews*, 9, 135-148.
- Wild, J. M. (1997). Neural pathways for the control of birdsong production. *Journal of Neurobiology*, 33, 653-670.
- Wyndham, E. (1980). Diurnal cycle, behavior, and social organization in the budgerigar (*Melopsittacus undulatus*). *Emu*, 80, 25-33.
- Yu, A. C., & Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science*, 273, 1871-1875.
- Zann, R. A. (1996). *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. New York, NY: Oxford University Press.
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63, 293-299.