Relative Salience of Envelope and Fine Structure Cues in Zebra Finch Song

Beth A. Vernaleo

Advisor:
Dr. Robert Dooling

March 2010

Abstract

Title of Dissertation: Relative Salience of Envelope and
Fine Structure Cues in Zebra Finch Song

Beth A. Vernaleo, Doctor of Philosophy, 2010

Dissertation directed by: Professor Robert J. Dooling
Neuroscience and Cognitive Science Program

This dissertation examines the perceptual salience of several acoustic cues in zebra finch song. Birdsong has long served as an animal model of speech development. Both are learned during a sensitive period, and require auditory feedback for learning and maintenance. Zebra finch song is commonly studied due to its stereotyped nature. Song syllables are complex, containing multiple cues that are modulated over millisecond time scales. Using psychoacoustic methods, male zebra finches were tested on discrimination of changes to their own and conspecific songs. Females and budgerigars were also tested, since they have auditory experience with song, but do not sing.

Three types of synthetic songs were created to determine which acoustic cues in song were most salient to birds. Same-seed noise songs were made of syllable envelopes filled with the same piece of random Gaussian noise. This removed spectral structure but kept song envelope cues intact. Random noise songs were made of each syllable envelope filled with a unique piece of noise. This provided more complex fine structure to the same song envelope. Lastly, Schroeder songs were made of Schroeder harmonic waveforms with the same duration as song syllables. In Schroeder waveforms, spectrum and envelope are constant, but phase changes occur across frequencies.

Two types of song changes were tested: single interval duration doublings and single syllable reversals. All birds were much more sensitive to syllable changes than to interval changes. For natural song, there was a duration effect on performance for male zebra finches only. Performance on syllable reversals shorter than 100 milliseconds was positively correlated with syllable duration. In Schroeder song, where only fine temporal structure changes with reversal, all three groups showed a duration effect. Thus, females and budgerigars may focus less on fine structure in natural song than males. In the absence of song spectral structure, birds relied on syllable envelope cues for reversal discrimination. Thus, removal of a single cue from song did not greatly affect reversal discrimination. However, birds performed best when all cues were present. This is reminiscent of human speech, in which multiple redundant cues are used for speech recognition.

Relative Salience of Envelope and Fine Structure Cues in Zebra Finch Song

by

Beth A. Vernaleo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland at College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:

Professor Robert J. Dooling, chair
Professor Catherine E. Carr
Professor Sandra Gordon-Salant
Professor Jens Herberholz
Professor Marjorie R. Leek

To all the birds in this world who gave their lives to science. Not only the ones who have died, but also the ones who never had the chance to fly free.

Acknowledgements

I would first like to thank my advisor, Bob Dooling. I came into Bob's lab halfway through my graduate career, and he had the difficult task of helping me find a project that both fit into the lab, and used my previous knowledge of zebra finch song. Not only did this knowledge not go to waste (if that is even possible), but I feel it has been enhanced a great deal in the time I have been in his lab.

Next, I would like to thank my committee members: Catherine Carr, Sandra Gordon-Salant, Jens Herberholz, and Marjorie Leek. Their advice, help, and support have been invaluable to me.

My labmates have provided a wonderful environment for me to do my work, and have acted not only as colleagues, but as friends. Thank you to Sandra Blumenrath, Kristina DeAngelo-Shea, Isa Noirot, Mike Osmanski, Yoshimasa Seki, Megan Shaw, and Vicky Tu. Vicky, having gone through this process just a few months ago has given me so much advice. I thank you for that. Thank you to Ed Smith for technical support,

and always finding interest in my adventures in signal processing. Beth Brittan-Powell has given me an immense amount of support through the most difficult times of my dissertation. Peter Marvit has always been there to lend an ear, whether it be for discussing logic, venting concerns, or literally lending an ear to listen to the quality of my stimuli.

Thank you to Steve Brauth, Sandy Davis, Chris Grogan, Bill Hodos, Pam Komarek, and Art Popper for helping me through things little and big, all through my time here at UMD. Thank you to Mia Bovill for being a great friend and neighbor. Thank you to Moonshine, Julep, Bear, Scotch, Freya, Merlot, Mudslide, Zoe, Jyoti, and Satchmo for all of your hard work pecking those keys! You guys are the best birds, and this dissertation would not be possible without you.

In addition to everyone mentioned above, I want to thank my amazing family. I am lucky enough to have two sets of parents: my biological parents (Mom and Dad Goldman), and my in-law parents (Mom and Dad Vernaleo). It's been a long road, and you've been there for it all. You always had faith that this day would finally come, and that means so much to me. Thank you to my brother-in-law Dennis for keeping me young with all the video games and those crazy nights out.

Lastly I want to thank my husband John, for being the only person who could always help me to see the end goal. These past 8 years have been some of the most difficult in my life, but also some of the best because they were with you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Birdsong has served as a model for human vocal development and communication for decades. Much of what is known about vocal development in Oscine songbirds has come from work with the zebra finch (*Taeniopygia guttata*). Zebra finches are closed-ended learners that have a single sensitive period for song learning, after which new song cannot be learned. The result of this sensitive period is a single, highly stereotyped song (Eales 1985) that is sung for both mating and territory display. While song motifs can range from three to eight syllables long (Sossinka and Böhner 1980), the single and repetitive nature of these songs allow for the study of normal song development. In addition, song learning and song production are subserved by two separate neural pathways in the avian forebrain. A great deal has been learned about neural mechanisms of song learning and production through lesions of the anterior forebrain pathway (AFP)(Bottjer et al. 1984; Nordeen and Nordeen 1993; Scharff and Nottebohm 1991), and electrophysiological recording in the AFP during singing and passive listening of song (Doupe and Solis 1997; Hahnloser et al. 2002; Solis and Doupe 1997). Thus, the song of the zebra finch has great importance for the understanding of vocal development and communication, both from an ethological and a neuroanatomical view.

While much is known about zebra finch song production, relatively little is known about how birds perceive song. Zebra finch vocalizations are quite complex, both spectrally and temporally. Given the importance of song to vocal communication in zebra finches, we sought to test the perceptual salience of three acoustic cues present in song syllables: syllable envelope, spectral fine structure, and temporal fine structure. In addition, song contains two scales of timing: fine structure changes that occur within syllables, and temporal envelope changes that occur more slowly across the entire song. Not only are these time scales seen in singing behavior, but temporal correlates are also seen in the avian forebrain. Specifically, the neurons in the nucleus HVC (proper name) seem to encode a representation of time for the bird's own song (BOS), and the neurons in the anterior forebrain pathway respond better to the BOS (and song in general) in the forward direction compared with song played in reverse. Thus, temporal information for fine structure and for overall song timing is represented in the avian forebrain as well as in singing behavior. For this reason, we will also specifically examine both global and local timing cues in song, to determine their salience in song perception.

The following sections review song and its development (Sections 1.1 and 1.2), timing in zebra finch song production (Section 1.3), anatomy of the zebra finch song system (Section 1.4), results of previous psychophysical experiments studying auditory perception in birds (Section 1.5), and perceptual phenomena which may explain why some cues in song are attended to more than others (Section 1.6).

1.1 Zebra finch song

Zebra finch song is composed of spectrally distinct acoustic elements separated by silence, such that each element is produced in a specific temporal order. These acoustic elements are termed syllables, and they are the smallest unit of song. The spaces of silence between adjacent syllables are referred to as intervals. Syllables sung in a specific order comprise a motif, which is repeated several times within the song bout. Motifs generally contain between three and eight syllables (Sossinka and Böhner 1980). Syllables are rich in harmonics, and can have rapid modulations in both amplitude and frequency. Even within single syllables, there is a multitude of timing information present. Figure 1.1 shows a typical zebra finch song, which begins with a few introductory notes, followed by several repetitions of the song motif. Figure 1.2 shows a single zebra finch song motif in detail, in which the individual syllables and intervals are labeled. The spectrograms show that most of the energy in zebra finch song is concentrated between 2 and 5 kHz. Zebra finches hear best in this frequency range. Zebra finch hearing begins to decline above 8 kHz, and thus there is very little energy in vocalizations above 10 kHz.

1.2 Song development in the zebra finch

Song learning in the zebra finch begins around 20 days post hatch (dph) with the *sensory acquisition phase*. In this phase of learning, birds listen to the tutor song and form an internal template. This phase lasts until $\sim$ 65 dph. Once birds begin to vocalize at 35 dph, an overlapping phase, the *sensorimotor phase*, begins. During this phase birds compare their own vocalizations to that of the tutor template, and modify their songs until they match the tutor song. This phase lasts until $\sim$ 90 dph

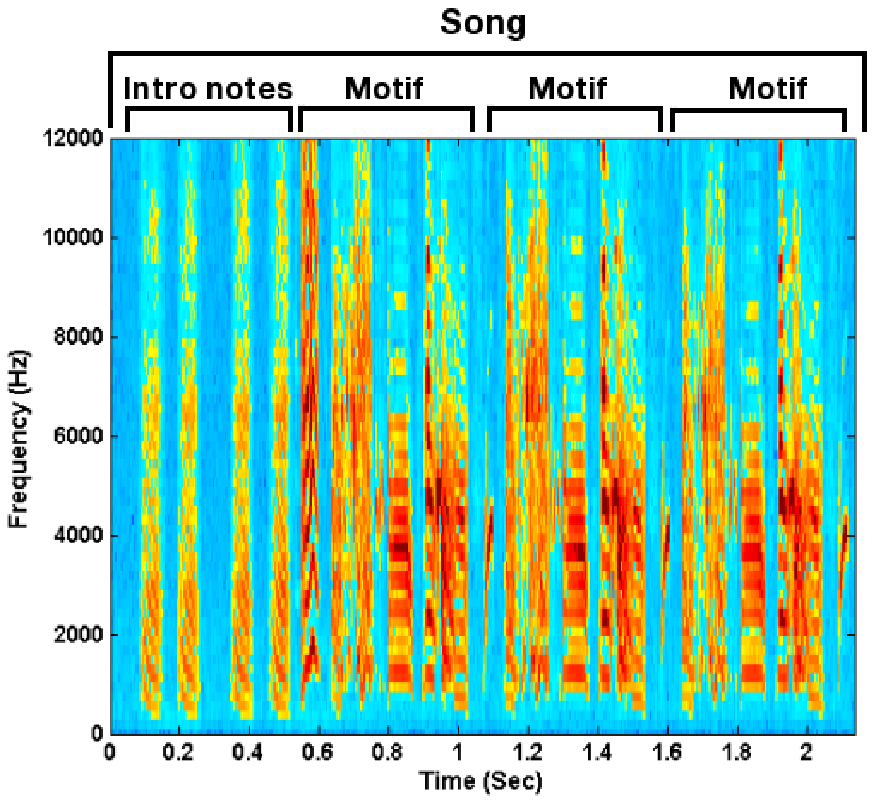


Figure 1.1: An example of typical zebra finch song. Each song bout generally begins with a few introductory notes, followed by a few renditions of the song motif.

when song crystallizes. At this point, song sequencing is stable (Immelmann 1969). These phases of song learning are illustrated in Figure 1.3.

Birds are able to hear sounds before they can vocalize. At ∼ 20 dph, hearing in juvenile zebra finches is mostly developed and audiograms resemble that of adults (Amin et al. 2007). At ∼ 30 dph, the synapses between the nucleus HVC and the robust nucleus of the archopallium (RA) (which sends input to the avian vocal organ, the syrinx) are formed. The first vocalizations at 35 dph are characterized by low

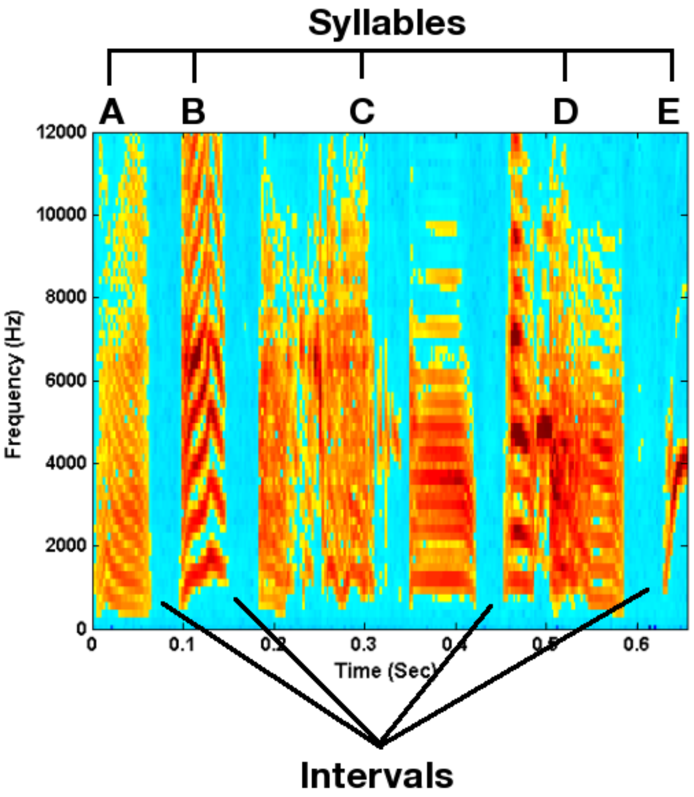


Figure 1.2: A single motif from the song shown in Figure 1.1. Syllables are labeled as A, B, C, D, and E. Intervals of silence separate adjacent syllables.

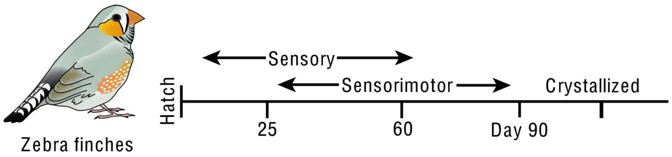


Figure 1.3: Time course for song learning in the zebra finch. Sensory acquisition begins at 20 dph and ends at 65 dph. The overlapping sensorimotor phase begins at 35 dph and ends around 90 dph. After 90 dph, song is crystallized and further tutoring does not result in addition of new song elements. Originally from Brainard and Doupe, 2002.

amplitude syllables which are noisy and lack the structure seen in adult zebra finch song (Immelmann 1969). Harmonics are present somewhat, but are highly variable and wavering rather than tonal. These vocalizations, termed *subsong*, acoustically resemble the variability seen in human babbling during vocal development. Subsong lasts until $\sim$ 50 dph, when plastic song begins. Early plastic song contains syllable repetitions that generally decline around 55 dph (Liu et al. 2004) when early syllable sequencing emerges. Late plastic song resembles adult syllable sequencing; however it is more variable. Song is stable at $\sim$ 90 dph, but small modifications are made over the next few months. For example, the tempo of singing increases, and song becomes more stereotyped.

1.2.1 Auditory feedback and song learning and maintenance

As song is practiced and modified throughout the sensorimotor phase, it is clear that auditory feedback is necessary. It was shown by Price (1979) that the songs of birds deafened between 63 and 84 dph showed a substantial degradation within a few months. Mean syllable frequency was lower than in normal birds, and most of the frequency patterning was lost, such that call and non-call syllables could not be distinguished. Syllable structure was much less stable than in hearing birds. While syllable durations remained the same, the length of intervals between syllables increased. The degree of degeneration was so severe that most post-operative songs could not be matched to their corresponding pre-operative songs by a naive observer. Somewhat smaller changes were found in most of the songs of adult birds deafened at 300 dph. However, recordings were done 6 months after deafening. It was later shown by Brainard and Doupe (2000, 2001) that birds deafened in adulthood do show a degradation of song, but the time course for this degradation is much slower than in birds deafened as juveniles or young adults (100 dph). Changes to song occur up to 1 year post-deafening. Thus, while real-time auditory feedback is not necessary for zebra finches to produce song, long-term maintenance of song does require auditory feedback (Nordeen and Nordeen 1992). Interestingly, the bengalese finch (a related species) is much more sensitive to the effects of deafening, and does requires real-time auditory feedback for song maintenance (Okanoya and Yamaguchi 1997).

1.2.2 Sensitive period for song learning

Tutoring and isolation studies have shown that there is a sensitive period for song learning (Eales 1985). Juvenile birds that are raised in isolation fail to develop normal song. Price (1979) showed that birds that were taken from their parents 9-12 dph, and housed together or separately, developed abnormal songs that had less distinct syllable types, and longer syllables with higher mean frequencies. Birds were isolated until 120 dph and then placed back into community cages. There were no changes to song when it was recorded 6-12 months later, confirming that birds did not incorporate their cage mates' song into their own during this time. However, birds that have been isolated do show an extended sensitive period for learning song, such that they can learn new song elements at later ages than birds reared normally (Eales 1985). This phenomenon also exists in birds that have been visually isolated from adult birds, but are still able to hear normal song (Morrison and Nottebohm 1993). Birds that were visually isolated developed more normal song, compared to birds that were both visually and acoustically isolated from tutors. However, visual isolates were still able to learn new song at 87 dph, suggesting that the sensitive period has remained open for longer in these birds. The results of this study also provide evidence that song learning is facilitated by social interaction in addition to auditory experience.

1.3 Scales of timing in song production

Two scales of timing exist within zebra finch song. *Temporal envelope cues* refer to the global timing of song, such as the duration of syllables and inter-syllable intervals. Modulations of overall envelope structure occur over fairly slow time scales. Envelope cues always refer to the amplitude envelope. Syllable envelope in particular refers to the amplitude envelope occurring over a single syllable. *Fine structure* is concerned with local timing over much faster time scales (milliseconds) within individual syllables. Fine structure encompasses both *temporal fine structure* such as phase and harmonic structure that vary within syllables, as well as *spectral fine structure* such as variations in frequency that occur over time.

Temporal envelope cues play an important role in song production, as they give song an overall rhythm. Song is highly stereotyped not only in syllable sequencing, but also in the timing of syllable production. It has been shown by Glaze and Troyer (2006) that for song not directed to females (undirected song), the length remains fairly stable, and deviations in song length are usually less than 1.5%. However, when there are deviations in song length, intervals tend to be more variable than syllables. Thus, the overall temporal patterning of song remains fairly stable from rendition to rendition, especially among the elements that are produced (i.e. syllables produced from expirations rather than silent intervals that are the result of inspirations).

Within the song bout, there is a tendency for each successive motif to become longer in duration (Chi and Margoliash 2001), and this occurs in both directed and undirected song (Cooper and Goller 2006). The mechanism appears to be similar in both types of song, as both syllables and intervals tend to increase in duration. This may reflect respiratory constraints or changes in motivation associated with prolonged singing. Cooper and Goller (2006) also compared motif and song lengths between directed and undirected song. Similar to other studies such as Kao and Brainard (2006) they found that overall, directed song is shorter in duration than undirected song, despite the increasing duration of successive motifs that is seen in both types of song. Upon measuring syllable and interval length in directed song, the authors found that only syllable lengths shortened, and interval lengths remained the same as in undirected song. This demonstrates two different mechanisms for the control of timing in song, one that is sensitive to social context, and one that is not.

These studies show the control of timing in zebra finch song to be systematic and deliberate, providing a precise and reproducible (i.e., stereotyped) behavior. Further evidence through anatomical studies suggests that a representation of timing within zebra finch song is encoded in the brain, both for temporal envelope and fine structure.

1.4 Anatomy of the zebra finch song system

The zebra finch song system consists of two separate neural pathways, one for song learning (the anterior forebrain pathway, or AFP), and the other for song production (the vocal motor pathway). One nucleus, HVC, sends input into both the AFP and the motor pathway. The vocal motor pathway consists of the following projections:

HVC → RA →NXIIts → Syrinx

where NXIIts is the *twelfth cranial nerve, the tracheosyringeal nerve.*

The AFP consists of the following projections:

HVC → Area X → DLM → LMAN → RA

where DLM is the *medial portion of the dorsolateral nucleus of the anterior thalamus* and LMAN is the *lateral magnocellular nucleus of the anterior nidopallium.*

These two neural pathways are illustrated in Figure 1.4. Separate populations of neurons in HVC project to the AFP (via the nucleus Area X) and the motor production pathway (via the nucleus RA). HVC receives auditory input from nucleus interfacialis (Nif), as well as the nucleus uvaeformis (Uva). Thus, HVC receives auditory feedback from the bird's singing, and can relay this information to both pathways in order to modify song during learning.

1.4.1 The anterior forebrain pathway (AFP)

The anterior forebrain pathway is considered a specialized basal ganglia loop, specifically the connections from Area X to the medial portion of the dorsolateral nucleus of the anterior thalamus (DLM), and from DLM to the lateral magnocellular nucleus of the anterior nidopallium (LMAN) (Luo et al. 2004). Electrophysiological studies have shown Area X contains striatal interneurons that receive input from HVC, and pallidal inhibitory neurons that project to the thalamic nucleus DLM (Farries and Perkel 2002). DLM projects to LMAN, which resembles cortex in this basal ganglia loop. As the output nucleus of the AFP, LMAN has been the focus of many studies investigating how the AFP shapes song production. Based on these findings, several hypotheses have been proposed concerning the functional role of the AFP(Brainard and Doupe 2000; Scharff and Nottebohm 1991). One hypothesis suggests that the AFP allows for plasticity during song learning (and extended plasticity during iso-

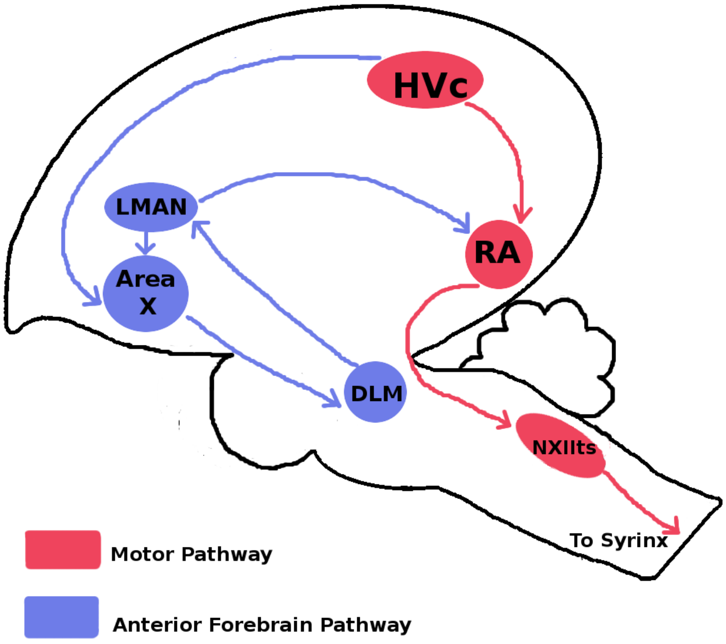


Figure 1.4: The two neural pathways involved in the learning and production of song. The nucleus HVC projects to both the anterior forebrain pathway (blue) and the vocal motor pathway (red).

lation) via trophic connections between LMAN and RA (Kittelberger and Mooney 1999; Morrison and Nottebohm 1993). Another possible role of the AFP is to provide variability in song that is necessary for exploration during reinforcement-based learning (Kao and Brainard 2006; Olveczky et al. 2005). A third hypothesis suggests that LMAN may provide an error correction signal when there is a mismatch between song output and the memorized tutor template. This last hypothesis is not limited to song learning and isolation; song degradation that takes place after deafening in adults is prevented when LMAN is lesioned (Brainard and Doupe 2000, 2001).

1.4.2 Representation of timing in the motor and anterior forebrain pathways

The motor production pathway has been implicated in representing timing information related to song. Electrophysiological studies have shown that RA-projecting HVC neurons ($\mathrm{HVC}_{\mathrm{RA}}$) are highly time-locked to the birds own song motif. $\mathrm{HVC}_{\mathrm{RA}}$ neurons fire only once during a song motif, and they reliably fire at the same time during each rendition of the motif (Hahnloser et al. 2002). This phenomenon is illustrated in Figure 1.5 and strongly suggests that there is a representation of time related to the birds own song in the brain. In addition, recent studies by Prather et al. (2008) have shown area X-projecting HVC neurons ($\mathrm{HVC}_{\mathrm{X}}$) in Swamp Sparrows *(Melospiza georgiana)* to have mirror properties such that neurons fire once at the same time during each motif rendition, both when the bird is singing, and when the bird is listening to his own song. Furthermore, auditory activity in single $\mathrm{HVC}_{\mathrm{X}}$ neurons is suppressed briefly before birds began to sing (for example, when the birds countersing to the song to which they were listening). Conversely, auditory-evoked activity in $\mathrm{HVC}_{\mathrm{X}}$ to songs played immediately after the bird finished singing remain suppressed for a short period of time. This suggests that $\mathrm{HVC}_{\mathrm{X}}$ neurons switch freely between auditory and motor production modes, and are important to both song perception and song production.

This representation of time is preserved in the firing of RA neurons, resulting in stereotyped song production. Unilateral lesions of HVC result in increased variability in the timing of when syllables are sung (Williams et al. 1992). The timing of syllable delivery after HVC lesions was so variable that the authors could not compare pre and post-surgery syllables for spectral changes, because they could not be matched through timing cues. As mentioned previously, HVC receives auditory

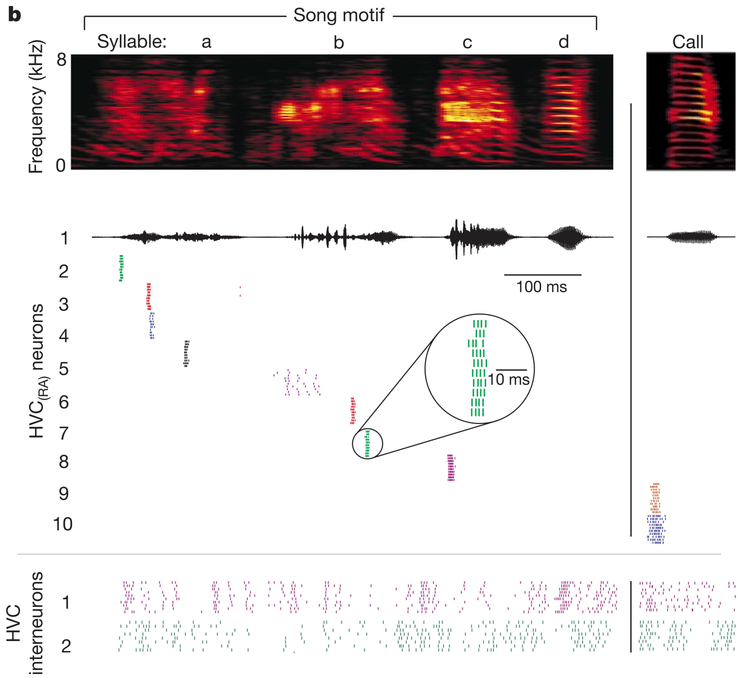


Figure 1.5: Sparse coding of timing within the nucleus HVC. Individual neurons are denoted by different colors in the raster plot. Single neurons in HVC fire during a specific time within the motif, every time the motif is sung. Neuronal firing to syllable A indicates that fairly close time points are coded by separate neurons, suggesting that the timing information in HVC is somewhat fine-grained. Originally from Hahnloser et al., 2002.

input from both Nif and Uva. Auditory feedback may be important in maintaining this precise timing in song production.

Evidence for this is seen in birds that have been deafened prior to 85 dph (Price 1979). Zebra finches that were deafened had longer interval durations compared to normally reared controls. However, syllable durations were the same as that of normal birds. Brainard and Doupe (2001) showed that birds deafened as young adults ($\sim$ 100 dph) and old adults had shorter syllable durations, and longer interval durations, compared to control birds. Lesions to LMAN, which are known to prevent spectral degradation to song associated with deafening (Brainard and Doupe 2000, 2001), failed to prevent the decrease in duration of syllables. However, LMAN lesions did prevent the increase seen in interval durations. From these studies, it appears that syllable and interval durations are in part regulated by auditory feedback. Peripheral damage such as unilateral tracheosyringeal nerve cuts did not reliably affect the duration or separation of syllables (Floody and Arnold 1997) although a slight lengthening of syllable duration was seen by Simpson and Vicario (1990) with bilateral nerve cuts. However, in both cases song remained stereotyped.

In addition to neural correlates of temporal envelope cues, neurons in the AFP may encode information regarding fine structure cues within individual syllables. AFP neurons have a preference for forward song compared to reversed song, which contains the same spectrum but differs in the timing (Doupe 1997). Neurons responded significantly less to song in which the entire song was reversed, but still responded to song in which only syllable order was reversed (CBA instead of ABC, which keeps syllable fine structure intact and only affects global temporal structure). Thus, while certain sequences produce the most neural activity (as would be expected, given the importance of sequencing in zebra finch song), local or fine structure within individual syllables is also important.

1.5 Auditory perception in birds

Auditory perception is the interpretation of a sound stimulus in order to form an auditory object that has meaning to the listener. Although the stimulus enters receptors located peripherally, perception is a central process. Psychophysical methods provide a way to rigorously investigate the perception and processing of sound. The following section reviews what has been learned thus far about song perception using psychophysical methods. This was done with several avian species, to gain an overall understanding of the hearing and perceptual capabilities of birds in general, and specifically zebra finches. The later sections present a few key psychophysical experiments that specifically studied perception of temporal envelope cues, and temporal fine structure in synthetic stimuli. From these results, one can ask whether similar temporal processing occurs when listening to song.

1.5.1 Perception of natural song

Zebra finch song contains both envelope and fine structure cues, both of which contribute to its characteristic sound quality. Temporal envelope cues provide the overall stereotypy and rhythm of song, while fine structure provides the modulations heard in individual syllables. Psychoacoustical experiments using natural song provide a way in which to ask whether birds perceive song similar to how they perceive non-song stimuli. Is song a special category of sound and perceived differently? Or are the hearing capabilities of zebra finches shaped by their auditory experience with song?

One classification experiment by Braaten et al. (2006) showed that zebra finches, regardless of age or auditory experience with song, used local timing information within syllables rather than global ordering of syllables to classify song. In a Go/No-go paradigm, birds were trained to discriminate between forward and reversed versions of a song. Birds pecked a key when they heard forward song and withheld pecking when they heard reversed song. On probe trials, birds were either presented with song in which the syllables themselves were reversed but remained in the same order (AᗺƆᗡƎ; syllable-reversed song), or song in which the syllables were played forward but their order was reversed (EDCBA; order-reversed song). Birds withheld pecking when they heard syllable-reversed song, which suggests that they treated this type of song like reversed song. When they heard order-reversed song, they pecked the key, suggesting that they treated this type of song like forward song. Thus, it was the local timing within the syllables and not the overall ordering that determined whether the birds categorized the probes as forward song, or reversed song. Since forward song is the only type of song encountered by zebra finches in the wild, it is possible that forward and reverse songs are categorized by these birds as natural and unnatural song. If this is the case, then songs are categorized as natural or normal based on the local timing of syllables rather than the global ordering of syllables.

Cynx (1993) also used a Go/No-go paradigm to test whether certain features in song were necessary for discriminating between two distinct songs. The birds used in these experiments all had different levels of familiarity with these songs (BOS, familiar song heard by male, and song heard by females). In the first experiment, birds were trained to discriminate between two distinct songs, and probe trials were given in which single syllables were deleted from these songs. Single syllable deletion only affected the discrimination task when the song being tested was the BOS. For females and male birds that were familiar with both songs, syllable deletion did not affect discrimination performance. Similar results were seen when probe stimuli were songs in which either the first half or the second half of song was deleted. Discrimination performance was only affected when the song tested was the BOS. In a third experiment, probe trials presented the time reversed versions of two songs. In this case, birds were not able to discriminate between the two songs and performance was at chance for both males and females. In a transfer test in which only the reversed songs were used, males performed significantly better than females in discriminating between these two songs. Thus, males were able to learn these discriminations, whereas females were not. It is unclear what cues in song males were using in order to make this discrimination. Due to the many acoustic cues present in song, more psychophysical experiments using natural song and song-like stimuli will provide insight into how each cue contributes to the perception of song.

1.5.2 Perception of temporal envelope cues in synthetic stimuli

Given that vocalizations in many species of birds can be very complex, with spectral and temporal modulations occurring within single syllables and over entire songs, it was previously thought that hearing in birds must be superior to that of other vertebrates. However, standard psychophysical studies involving temporal processing of envelope cues have shown birds to have similar processing to humans and other vertebrates, which was unexpected.

Duration discrimination experiments with tonal signals showed budgerigars to have similar thresholds to humans in detecting increases in duration, which is about 10-20% of the stimulus length, regardless of frequency (Dooling and Haskell 1978). Work in European Starlings *(Sturnus vulgaris)* by Maier and Klump (1990) showed that detection of an increase in duration has a lower threshold than detection of a decrease in duration. However, both thresholds for increments and decrements of tone duration ranged from 10-25% of the reference stimulus length. This same pattern was seen across all frequencies tested. In addition, gap detection thresholds are also similar in birds and humans. When House Finches *(Carpodacus mexicanus)* and humans were tested on detection of gaps in 500 msec noise bursts, they had similar thresholds (Dooling et al., 1978) ranging from 3.6-5.2 msec.

The threshold for detection of sounds depends upon the duration of the stimulus. This relationship between duration and threshold is the result of the temporal integration of sound energy. The time constant of the temporal summation function in zebra finches for a 3-kHz pure tone is around 230 msec, which is similar to the

$\sim$ 200 msec time constant seen in budgerigars, field sparrows, starlings, and humans (Okanoya and Dooling 1990). Together, these results suggest that birds have similar temporal processing abilities to humans and other vertebrates, with respect to temporal envelope characteristics of simple stimuli.

1.5.3 Perception of temporal fine structure in synthetic stimuli

Recent studies have shown that zebra finches excel in fine temporal discrimination tasks in which changes occur over periods as short as 1-2 msec. Zebra finches produce harmonic contact calls, as well as harmonic stacks and sweeps within the song motif. The following experiment tested perception of similar harmonic stimuli in a few different species.

Dooling et al. (2002) compared zebra finches, budgerigars, canaries, and humans on a discrimination task involving Schroeder-phase harmonic stimuli. Schroeder-phase harmonics are harmonic complexes in which the component phases either increase monotonically, or decrease monotonically. Regardless of the Schroeder phase (increasing or decreasing), the amplitude envelope and long-term spectra remain the same. Thus, positive and negative phase Schroeder harmonics differ only on temporal fine structure. Birds and humans were trained to discriminate between positive and negative phase Schroeder harmonic complexes.

The fundamental frequencies of these harmonic complexes ranged from 150-1000 Hz, which means the periods ranged from 1.0-6.6 msec. For 150 Hz Schroeder complexes (6.6 msec period), all birds and humans performed similarly well, and could discriminate between a positive and negative phase Schroeder complex. However, as the periods of the harmonic complexes decreased in length, performance dropped for all but the zebra finches. For the shortest period of 1.0 msec (1000 Hz fundamental frequency) performance for the budgerigars, canaries, and humans was at or below chance, while zebra finches discriminated approximately 90% of the stimulus pairs. Thus, the temporal resolution of the zebra finch auditory system is better than that of humans, and other avian species.

Further testing the limits of temporal resolution in zebra finches, Lohr et al. (2006) trained zebra finches to discriminate between normal zebra finch calls, and calls in which single periods of the call had been reversed. Finches were tested on how many time-reversed periods of a zebra finch call were necessary in order to discriminate the target from the normal call background. Zebra finch call periods were grouped together (in groups of 3) and the two end periods were ramped to create a pulse. Multiple pulses were concatenated to create a 200-msec stimulus, which served as the background. The targets consisted of the same stimulus, but with the beginning pulses reversed in time. Each target had a different number of pulses reversed, and the threshold for 50% discrimination was determined. The average number of pulses necessary for 50% discrimination was 7.29 pulses, or 31.71 msec (16% of the total stimulus length). Not only can zebra finches detect small changes in temporal structure when the amplitude envelope and frequency spectra remain the same, but they also detect these changes over a short time period, perhaps a similar time course to the modulations seen in their own song syllables.

1.6 Psychophysical approaches to song perception

There is no question that birdsong, specifically zebra finch song, is a complex sound source. One motif of song contains temporal envelope cues relating to the overall rhythm and prosody, fine temporal modulations seen in individual syllables, amplitude envelope cues, and spectral information that changes over millisecond time scales. Results of the experiments presented in this dissertation will show that certain changes to song are more salient, and thus more easily discriminated, than others. The following sections discuss two perceptual phenomena (*Informational masking* and *figure-ground perception*) that we may use as an interpretive framework to better understand why some changes to song are more salient than others. Informational masking and figure-ground perception have been widely studied in both the auditory and visual domains, using psychophysical techniques.

1.6.1 Informational masking

Masking in general (in the auditory realm) is defined as degradation of sensory processing of a signal in the presence of a masking stimulus, compared to processing of that signal in quiet or optimal conditions (Watson 2005). This degradation usually presents itself as an increased threshold for detection, or increased difference limen for discriminating changes in frequency, duration, or level. What constitutes a "masking stimulus" depends upon the type of masking, of which there are two. **Energetic masking (EM)** refers to masking that occurs when the masker and the signal both have energy in the same frequency bandwidth. Thus, there is competition at the auditory periphery such that the same response is given to the signal and masker together as to just the masker alone (Watson 2005). The result is that the signal and masker appear intertwined, and the signal cannot be resolved. **Informational masking (IM)** refers to all other masking that is not accounted for by energetic masking. Typically this occurs when the signal and masker are sufficiently similar (as in human speech), or when there is uncertainty in the stimulus.

Stimulus uncertainty can refer to trial to trial variability, or uncertainty as to what to attend to in a complex sound. While the term *complex sound* typically refers to multiple sounds either played sequentially or simultaneously, Kidd et al. (2007)

suggest that this could also refer to a single sound that contains multiple cues. In this case, failure to attend to a certain feature of a sound could also be considered IM, as interference is present that is not attributable to EM (Kidd et al. 2007). While this second definition is not the typical definition of IM, the idea of stimulus uncertainty affecting perception of certain features in a sound may still apply to birdsong. In this case, the techniques used to provide a release from informational masking may also be used to affect the perception of certain features in song, by drawing attention to them. Several studies testing the release from informational masking are presented below.

With regard to IM resulting from stimulus uncertainty, release from masking generally occurs when more information about the stimulus is provided to the listener. In a two-part paper, Charles S. Watson and colleagues examined frequency discrimination within tonal patterns under varying levels of stimulus uncertainty. The first paper looked at frequency discrimination of a single tone in a complex of ten 40-msec tones in which each tone was a different frequency. Subjects first heard a pattern of ten tones. Then they heard a second pattern of ten tones which was either the same, or had the frequency shifted for one tone in the pattern. Subjects had to indicate whether the two patterns were the same, or different. In one experiment, the same pattern was used every time, and the location of the frequency shift was changed from trial to trial (i.e. tone 1 was shifted on the first trial, tone 8 was shifted on the second trial, etc). In a second experiment, 50 different patterns were used, and the pattern changed on each trial in addition to the changing location of the frequency shift within the pattern. Watson et al. (1975) found that just detectable frequency differences (Δf/f) were smaller for the last tone in the sequence than for prior tones, which is explained by the recency effect. When looking at frequency discrimination of the last tone in the sequence, subjects had better frequency resolution when the same pattern was used for each trial than when the pattern changed from trial to trial. This suggests that uncertainty in the stimulus set somehow interferes with the task such that subjects have worse frequency resolution in these listening conditions.

In the second paper of the series, Watson et al. (1976) examined the results of fifteen frequency discrimination studies in which the same ten-tone paradigm was used. The main difference between each of the studies was the level of stimulus uncertainty. Subjects had the best frequency resolution (Δf/f=0.014) when only one frequency was used for the ten tones, and only one component of the pattern was shifted in frequency. Thus, there was no frequency or temporal uncertainty. Subjects had the worst frequency resolution (Δf/f=0.50) when the pattern changed each trial, there were 540 possible frequencies, and one of four temporal components (first, fourth, seventh, or last) were subject to frequency shift. While it is difficult to parse out which factors contributed to decreased frequency resolution (uncertainty in the tone pattern, or uncertainty of the location of the frequency shift), it is clear that this added uncertainty contributed to a degradation in performance. In one case, the listener becomes uncertain about which frequency pattern to expect from trial to trial, and in the other the listener becomes uncertain as to the temporal location of the frequency change, or "where to listen for the change".

It has been postulated that another reason why informational masking occurs is that the listener fails to focus attention to the relevant parts of the stimulus. Some studies have shown that making parts of the stimulus more salient to the listener results in a release from masking. In particular, Leek and Watson (1984) showed that for frequency discrimination of individual tones within a ten-tone sequence, some tones were more difficult to learn than others. In this task, subjects must discriminate changes to frequency in one tone that was embedded within a ten-tone pattern. The target tone was played out at different sound pressure levels, and the level at which frequency discrimination was correct 71% of the time was considered the threshold. All target tones started out with a high discrimination threshold. Target tones that were high frequency reached low thresholds the fastest. Some of the target tones that were low frequency reached low thresholds much slower, and in a few cases thresholds were high throughout the experiment. For these cases, the target tones were made more perceptually salient by increasing the duration of just the target tones. With this paradigm, listeners showed an immediate improvement in performance, in which thresholds lowered. When the duration was gradually shortened to the original duration of 45 msec, performance remained stable. Once the listeners had learned to "hear out" the tone of interest, they no longer needed the increased duration for the task.

The subject of informational masking and auditory attention was further examined by Leek et al. (1991) in a study in which they measured frequency resolution of single tones within a nine-tone sequence, as the target was increasingly isolated from the rest of the sequence. Target tones were isolated either by frequency or intensity, while the remaining tones in the sequence had similar frequencies and intensities. Two ranges of frequencies were tested, for a total of four conditions. For both frequency and intensity, increasing isolation resulted in a release from masking and improved frequency resolution. This relationship was linear up to a point, after which performance became asymptotic, and more isolation from the sequence did not improve performance. Thus, as the listener's attention was drawn to the target tone, a release from masking occurred. Once the listener knew which tone to attend to, frequency resolution approached that of the same tone presented singly.

1.6.2 Figure-ground perception

Figure-ground organization, exists in both visual and auditory perception, and determines how various portions of a complex stimulus are attended to. In vision, this tends to occur at luminance edges, where one region is considered the main figure, and the rest is considered the ground, or background. Even though much of the work on figure-ground perception has been done using visual stimuli, the same principles can be applied to auditory stimuli, whether they contain multiple or single sound sources. Thus, a review of figure-ground organization and how it affects attention and processing of stimuli may provide insight into how zebra finch song, a complex single source stimulus, is perceived.

A study by Nelson and Palmer (2007) suggests that when viewing a stimulus that contains classical figure-ground organization, attention is drawn to the figural side. In turn, this may result in more rapid and accurate processing of targets that appear on the figural side, compared to those that appear on the ground side. In their experiment, subjects viewed a face in profile (figure) against background (ground). The face and background were equal in area. The task was to press a button when an "X" appeared on the screen. The target "X" could appear on the face or the background with equal probability. Subjects detected the presence of the target faster when the target appeared on the face, or figural side of the screen. Reaction times for targets on the figural side were approximately 100 msec faster for targets on the figural side (~ 550 msec) than for targets on the ground side (~ 650 msec).

In a second experiment, the accuracy of processing targets on the figural vs. ground side was tested by presenting either an "X" or a "Y" as the target and asking subjects to indicate which target was presented. Both types of targets appeared an equal number of times, and the both locations of the targets (figure or ground) had an equal probability of occurring. On probe trials, subjects were also asked to indicate which side of the screen they judged to be the figure, and which side they judged to be the ground. Targets that appeared on the side that subjects judged to be figural were named more accurately than targets that appeared on the side judged to be the ground. The results of these two experiments provide evidence that processing of targets that appear on the figural side of a stimulus are detected faster, and more accurately than targets that appear on the ground side, due to differing levels of attention.

With regard to song, it is possible that syllables may be perceived as the figural part of the stimulus, whereas the intervals, or rate of delivery of syllables is perceived as the ground. Evidence for this is seen in Experiment 1, in which changes to syllables and changes to intervals show vast differences in discrimination performance.

1.7 Relative salience of acoustic cues in song

In this dissertation, we examine the relative salience of three acoustic cues to the perception of zebra finch song: syllable envelope, spectral fine structure, and temporal fine structure. In addition, we test perception of the two scales of timing in song: temporal envelope (global timing) and fine structure (local timing). Three groups are tested (male zebra finches, female zebra finches, and female budgerigars) in order to determine whether vocal or auditory experience with song affects which cues are most salient. There are a few possibilities for how these acoustic cues are weighted. One is that a single cue stands out from the rest, and is the main cue that birds attend to when listening to song. If this is the case, then removing this cue should greatly affect song perception, whereas removing all other cues should not affect perception. However, this seems unlikely since song is a biologically relevant stimulus, and its transmission is very important to species survival. Zebra finch song is mainly produced in grasslands and forests, and is subject to background noise. The more likely possibility is that birds use a combination of acoustic cues when listening to song, similar to the way in which speech is perceived. With speech, single cues can be removed without affecting intelligibility, but speech perception is best when all acoustic cues are present. This is especially true when listening to speech in situations where there is noise in the background.

Experiments presented in Chapter 3 involve testing discrimination of changes made to natural song. The main question asked here is: *Which time scale in song is more salient to birds: temporal envelope or fine structure?* Results from the first experiment showed that fine structure changes within syllables are much more salient than temporal envelope changes to intervals. Subsequent experiments in this chapter asked related questions: *Is fine structure discrimination dependent upon the syllable type, syllable duration, or syllable position within the motif?* (Experiment 2); *Does interval discrimination improve when only interval changes are presented?* (Experiment 3); *Does the presence of intervals aid in the perception of fine structure within syllables?* (Experiment 4); and *How does familiarity with the background stimulus affect discrimination of temporal reversals?* (Experiment 5).

Experiments presented in Chapter 4 involve testing discrimination of changes made to synthetic noise songs in which the syllable envelope is filled with random noise. Since fine structure in song syllables consists of both spectral and temporal fine structure, this chapter asks: *What is the role of spectral structure in syllable discrimination? Can changes to syllables be discriminated in the absence of spectral structure?* (Experiment 6). Results from this experiment showed that birds are able to discriminate changes to syllables when spectral structure is removed. Further experiments in this chapter asked: *Do birds rely more on syllable envelope cues or fine structure in noise to discriminate reversals of noise syllables?* (Experiment 7); *Does removing spectral structure from song syllables affect perception of temporal envelope cues?* (Experiment 8); and *What is the robustness with which birds can discriminate changes to the fine structure of random noise?* (Experiment 9).

Lastly, experiments presented in Chapter 5 involve discrimination of changes made to synthetic songs made of Schroeder harmonic complexes in order to ask: *Is temporal fine structure alone sufficient for syllable discrimination?* (Experiment 10). Schroeder harmonic complexes have a constant amplitude envelope and spectrum across time, but phase sweeps occur across frequencies. Thus, the only acoustic cue that changes when a Schroeder harmonic is reversed is temporal fine structure. The final experiment in this chapter (Experiment 11) utilizes the Schroeder harmonic to ask *Is the threshold for frequency discrimination in a syllable the same for a single syllable as for a syllable embedded in a song motif?* Since much of the previous research on zebra finch hearing and perception used single stimuli, it is important to know whether these same findings extend to a more complex acoustic environment, such as song.

Chapter 2

General methods

2.1 Subjects

Four male zebra finches (*Taeniopygia guttata*), three female zebra finches, and three female budgerigars (*Melopsittacus undulatus*) were used. Birds were housed at the University of Maryland in an avian vivarium, and kept on the photoperiod corresponding to the current season. Birds were maintained at 85-90% of their free feeding weights and were given free access to water. All procedures were in accordance with the University of Maryland Institutional Animal Care and Use Committee (IACUC) under the protocol,"Biological foundations of vocal learning" (07-10). For the four male zebra finches, thresholds were measured for the best frequency of hearing (2.86 kHz), and all four birds had normal thresholds (ranging from 1.75-4.29 dB) at this frequency.

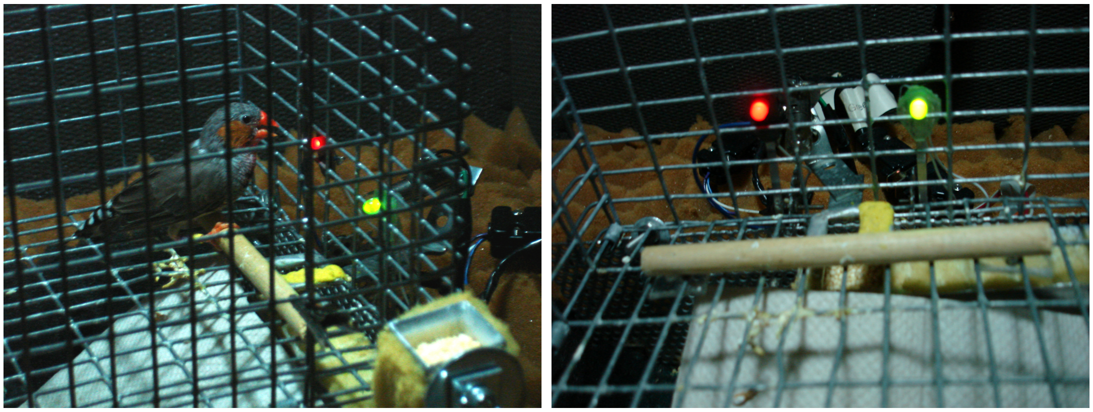
Two human subjects were used in a subset of the experiments. Human subjects were undergraduates at the University of Maryland, 21 years of age, and had normal hearing. All procedures for human testing were in accordance with the University of Maryland Institutional Review Board (IRB) under the protocol, "Auditory Perception in Humans and Birds" (09-0393). The experimenter completed all necessary training under the Collaborative Institutional Training Initiative (CITI) in order to work with human subjects.

2.2 Psychoacoustics apparatus

All psychoacoustic experiments involving birds took place in a wire cage anchored inside of a sound-attenuated chamber (Industrial Acoustics Company; Bronx, NY). The chamber was lined with acoustic foam to absorb reverberations of the sounds that are presented to the bird. Inside the cage a perch was fixed to the floor. The bird sat on this perch and had access to food through an opening in the floor of the cage. Millet was delivered through a food hopper, which was brought up to the food opening through activation of a solenoid. Without activation (via hitting the report key), the hopper remained in the down position, and could not be accessed through the food opening. The two response keys were mounted to the wall of the cage, directly in front of the perch. The keys consisted of 8mm LEDs separated by 5 cm, each attached to a micro switch. The left (red) LED served as the observation key, and the right (green) LED served as the report key. The chamber was illuminated from above with a 60-watt light bulb, and monitored from the outside via an overhead video camera. Water was available via a bottle attached to the side of the cage. Figure 2.1(a) shows a view of the entire psychoacoustics setup (minus the speaker, which is above the cage), and figure 2.1(b) shows a closeup view of a bird in position to begin a trial. Figure 2.1(c) shows the observation and report keys in detail.

Sounds were presented via a speaker (KEF Model 80C; Kent, UK) mounted to the chamber ceiling, angled 45° downward toward the cage. The speaker was approximately 40 cm away from the bird's head. Stimuli were generated in Matlab

(a) The psychoacoustics setup consists of a cage which houses a perch and two LEDs, which are the observation and report keys. The food hopper (on the right) is brought up via a solenoid, and allows access to food through a hole in the floor of the cage.



(b) A closeup of the setup, in which a bird is sitting on the perch ready to begin a trial.

(c) The observation (red) and report (green) keys from the bird's point of view, and the opening in the floor through which the bird is fed.

Figure 2.1: A detailed view of the psychoacoustics setup used for discrimination experiments

(Mathworks; Natick, MA) or Adobe Audition (Adobe; San Jose, CA) as wav files (48,000 Hz sampling rate) and stored on an Intel Core 2 Duo computer (Mid Atlantic Data Systems; Gaithersburg, MD), which controlled all experiments. The computer operated a Tucker Davis Technologies System 3 module (TDT; Gainsville, FL), in which sounds were sent to a D/A converter (TDT RX-6), then output to a programmable attenuator (TDT PA-5), and a signal mixer (SM-5). Then sounds were played out of the speaker in the psychoacoustics chamber at a sampling rate of 24,414 Hz.

2.3 Calibration, training, and testing procedures

2.3.1 Calibration of stimuli

Stimuli were calibrated regularly to ensure that the maximum sound pressure level (dB SPL) remained stable over time. Attenuation took place online during the experiment so that backgrounds and targets containing song were played out at 70 dB SPL. Calibration was done using a Larson-Davis sound level meter (Model 824; Provo, UT), placing a 1/2 in. microphone at the approximate location of the bird's head. Fast, maximum dBA measurements were recorded in a calibration book each time stimuli were calibrated.

2.3.2 Training

Once birds were 85-90% of their free feeding weights, they were trained to peck the keys for a food reward. Training was done using an operant auto-shaping program written in Matlab. Training consisted of five phases, which are described below.

Hopper Training: The food hopper is propped in the "up" position so that birds have free access to millet. Once the birds are acclimated to the chamber and eat from the hopper, Phase 1 begins.

Phase 1: The hopper begins in the down position, and at predetermined intervals (usually 45 or 60 seconds), the observation key LED blinks, a tone is played, and the hopper is raised. The bird must peck the observation key 10 times to raise the hopper in order to move onto Phase 2.

Phase 2: The observation key LED remains on and the hopper is no longer raised at the predetermined intervals as in Phase 1. The bird must peck the observation key 10 times to raise the hopper in order to move onto Phase 3.

Phase 3: The observation key LED remains on, and once the bird pecks it, the report key blinks and a tone is played. The bird must then peck the report key to raise the hopper. The bird must perform this sequence of actions 10 times in order to move onto Phase 4.

Phase 4: Both the observation key and report key LEDs remain on. The bird must first peck the observation key and then peck the report key to raise the hopper. After the bird pecks the observation key, a tone is played, signaling the bird to peck the report key. The bird must perform this sequence of actions 10 times in order to move onto Phase 5.

Phase 5: This final phase is identical to Phase 4, except that random sham trials (in which no tone is played while pecking the observation key) are introduced. Birds must withhold pecking the report key in the absence of a tone. Birds remain in Phase 5 until they are able to correctly perform sham trials.

After birds completed all five phases of training, the variable stimulus presentation interval was gradually increased to 2-6 seconds. This value was 0 seconds during training, meaning that a tone is presented immediately after the observation key peck. In addition, the maximum response interval was decreased from 3 to 2 seconds, meaning the bird must peck the report key within two seconds of the target being played. Lastly, food reinforcement was decreased from 100% to 80-90%. On the trials in which there is no food reinforcement, a light near the hopper came on, indicating a correct response.

2.3.3 Testing procedures

During training, birds must detect a tone in silence. For all experiments in this dissertation, birds must discriminate a change from a repeating background. The task is for birds to peck the observation key for a random interval (between 2 and 6 seconds) while listening to a repeating background. After the random interval, the background is alternated with a target sound. Once the bird hears this alternation between background and target, he must peck the report key within 2 seconds. If the bird performs this task correctly, this is recorded as a hit and he receives a food reward. If the bird fails to peck the report key within two seconds of the target presentation, this is recorded as a miss. In any given session, 30% of all trials are sham trials in which no target is alternated with background. If the bird correctly withholds pecking the report key, this is recorded as a correct rejection, but is not rewarded with access to the hopper. If the bird pecks the report key during a sham trial, this is recorded as a false alarm, and is punished with a 5-second time out period in which all lights in the chamber are extinguished. In most cases, any sessions in which the false alarm rate exceeds 20% are not included in analysis. Figure 2.2 shows a single trial, and the possible outcomes.

Birds ran 100 trial sessions in the morning and afternoon, for 5 days a week. Each 100 trials session contained 10 blocks of 10 trials, in which 7 trials contained targets, and 3 trials contained shams. For experiments involving natural song, birds ran no more than 300 trials for a given stimulus set, with the last 200 continuous valid trials used for analysis. Trials in which the false alarm rate is above 20% were thrown out and not counted towards the 300 trials. This limit on the number of trials was to prevent over-training, as the goal is to study how birds naturally perceive birdsong. In these experiments, we ask whether birds normally make these discriminations, and not whether they can make them after many days of training. For experiments that do not involve natural birdsong, birds ran until they could complete 300 trials with a false alarm that did not exceed 20%. In most cases, birds were run for less than 600 trials.

2.4 Stimuli

For all experiments using natural birdsong, vocalizations were recorded from birds in a foam-lined acoustic chamber, using a Marantz portable solid state recorder (Model PMD670) at a sampling rate of 48,000 Hz. Individual song motifs were extracted from song recordings using Adobe Audition, and high-pass filtered with a cutoff frequency of 350 Hz using Raven Pro 1.3 (Cornell Lab of Ornithhology; Cornell, NY). Inter-syllable intervals were band reject filtered at all frequencies to produce pure silence between syllables. Motifs were given a 5 msec cosine rise and fall time to prevent clipping artifacts. The original song motifs were saved, and copies were manipulated in Adobe Audition for single syllable reversals, addition of duration to

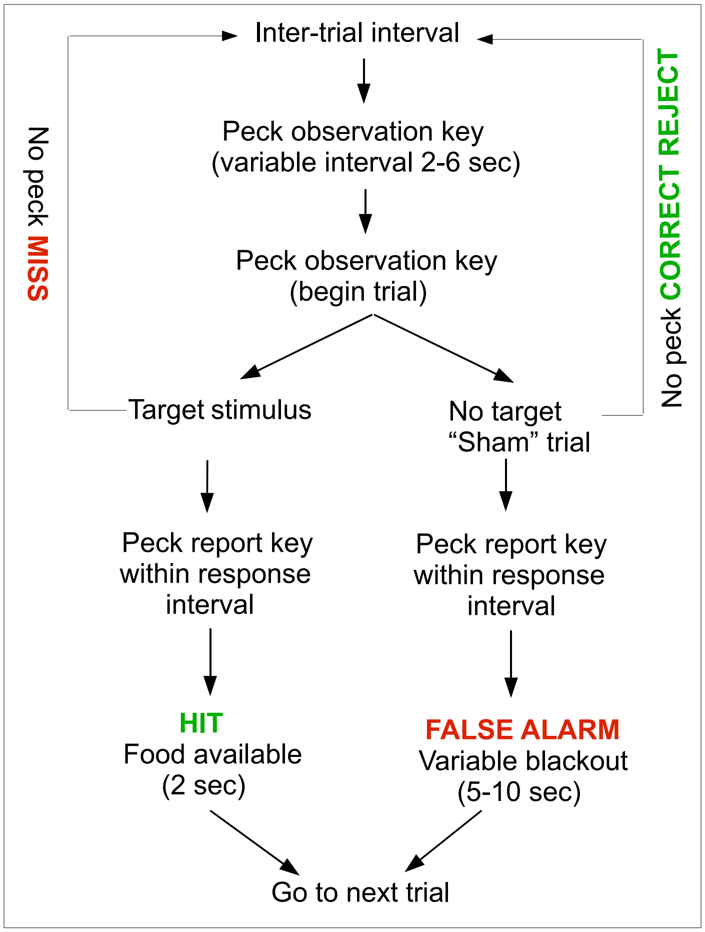


Figure 2.2: A flow chart of a single trial in the psychoacoustics testing procedure, and the possible outcomes. The bird pecks the observation key while listening to a repeating background, and after a variable interval (2-6 seconds) the trial begins. Either a target stimulus is alternated with the background (target trial) or the same background sound is alternated with the background (sham trial). During target trials, if the bird pecks the report key during the response interval, this is recorded as a hit and is rewarded with access to food. If the bird fails to peck the report key, this is recorded as a miss but is not punished. During sham trials, if the bird pecks the report key during the response interval, this is recorded as a false alarm and is punished with a blackout period. If the bird withholds pecking during the response interval, this is recorded as a correct rejection but is not rewarded.

intervals, etc. (see specific methods sections for each experiment).

For all experiments using noise bursts, stimuli were created in Matlab, either using a random seed (so that each noise burst was completely different from one another), or using the same seed (so that the beginning of each noise burst was the same and only the end of the burst differed based on duration). Bursts were sampled at 48,000 Hz, and individual bursts were given a 5 msec cosine rise and fall time to prevent clipping artifacts.

2.5 Human testing

For tests involving human subjects, humans were tested outside of the psychoacoustics chamber, and stimuli were presented over headphones at 70 dB SPL. The same procedure was used for testing humans as for birds. Humans were given a hand-held control for the observation and report keys, and used this to respond.

2.6 Analysis

Analysis was done using custom scripts that were written in Matlab. For each set of 100 trials, the percent correct hit rate was calculated for each target. For each trial, the response latency was recorded. For response latencies, missed targets in which birds failed to peck the report key were assigned a latency of 2500 msec, which corresponded to the maximum response interval. This assumes that birds would respond to the target, but could not do so within the allotted time. Raw latencies as well as latencies corrected for position were calculated. Corrected latencies were used in order to normalize for the location of the change within the song. If a change occurred in the last syllable of the motif, the response latency should not include the time before the change actually occurred. These latencies were calculated such that the duration of song that occurred before the start of each reversed syllable was subtracted from the raw response latency of that target. Thus, the resulting latency was the latency from the start of the individual reversed syllable. Any response latencies in which the response occurred before the start of the syllable reversal were thrown out.

Statistical tests were performed in SigmaStat (Systat Software; Chicago, IL) as well as Matlab. Because the majority of the data violated assumptions of normality and equal variance, non-parametric tests were used. In most cases the tests used were Mann-Whitney rank sum tests, Kruskal-Wallis ANOVA on ranks, and Friedman tests. These tests are the non-parametric equivalents of t-tests, one way ANOVA, and two-way repeated measures ANOVA, respectively. Parametric tests were used whenever data passed normality and equal variance tests.

2.7 Terminology

All stimuli that are based on songs (i.e., natural songs, noise songs, and Schroeder songs) are comprised of a single motif (see Figure 1.1) taken from the entire song. In this dissertation, song and motif are used interchangeably when referring to the stimuli. This is because we are inferring how birds perceive song, by testing discrimination of changes within single motifs.

Since the birds that provided the motifs are the same birds being tested, the birdsongs will be referred to as MoonBOS, BearBOS, ScotchBOS, and JulepBOS in data tables to avoid confusion with the birds Moonshine, Bear, Scotch, and Julep.

Chapter 3

Perception of envelope and fine structure cues in natural zebra finch song

3.1 Introduction

Two scales of timing exist in zebra finch song: global and local timing. Global timing occurs over many milliseconds to seconds, and accounts for the overall rhythm and timing of song. Global timing cues are considered *temporal envelope cues*, and include duration of syllables, duration of inter-syllable intervals, and the order in which syllables are sung. Local timing, on the other hand, takes place over the course of a few milliseconds. Local timing cues are considered *fine structure cues*, and include all of the changes occurring within single syllables or portions of syllables. Fine structure cues include both changes to the spectral structure of syllables (spectral fine structure), and changes to the phase and harmonic structure of syllables (temporal fine structure). While both temporal envelope and fine structure cues appear to be equally represented when zebra finches sing their song, it is unclear how salient each type of cue is to zebra finches when they hear song.

The following experiments were designed to ask whether zebra finches could discriminate changes made to their song, and which types of changes (temporal envelope or fine structure) were easier for them to discriminate. To ask whether vocal and auditory experience with song plays a role in perception, experiments were done with male zebra finches, female zebra finches, and budgerigars. Male zebra finches sing, whereas females do not. Thus, males have both auditory and vocal experience with song, while females only have auditory experience. Budgerigars lack vocal experience with zebra finch song. However, they are housed in the same room as zebra finches, so they have some auditory experience with song.

3.2 Experiment 1: Relative salience of envelope and fine structure cues within the song motif

3.2.1 Introduction

This experiment tests discrimination performance on changes to temporal envelope and changes to syllable fine structure, in a single test session. In this way, these two types of changes can be directly compared. Further experiments in which only one of these cues is presented in a test session will be discussed later.

3.2.2 Methods

Subjects

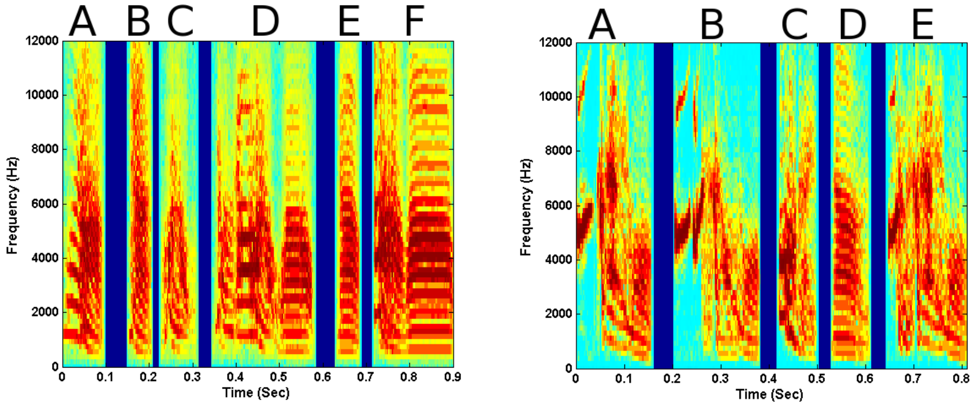
Four male zebra finches, three female zebra finches, three female budgerigars, and two human subjects (1 male, 1 female) were used in this experiment. Humans were used for comparison, since speech generally does not contain rhythm cues as zebra finch song does. Words in sentences are spoken at the same rate throughout the sentence. This contrasts zebra finch song, in which intervals separating syllables are all distinct durations.

Stimuli

All stimuli were created from the four male zebra finch motifs previously recorded (see general methods). The four original song motifs that were used in Experiments 1- 5 are shown in Figure 3.1.

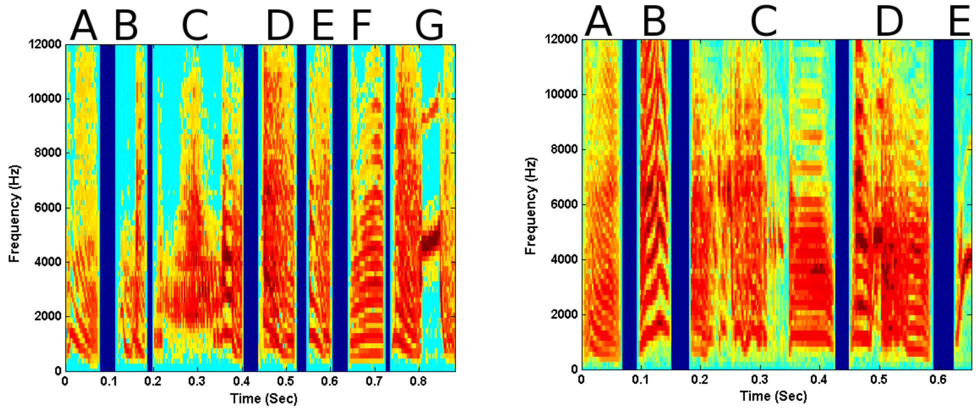
Temporal envelope changes consisted of doubling single inter-syllable intervals in the motif. This was done by adding the same amount of silence to the interval as the length of the interval itself. In generating the original motif, the inter-syllable intervals were band-reject filtered at all frequencies to produce pure silence between the syllables. This was done primarily to remove background recording noise, which can provide an additional cue when doubling the interval duration. Specific targets consisted of the original motif, with only one interval doubled in duration at a time. This allowed individual intervals in the song to be tested and compared to one another.

Changes to syllable fine structure in the motif consisted of reversing single syllables in time, while keeping the order of syllables in the motif intact. Thus, syllable order remains stable and only local timing within the syllable is changed. In revers-



(a) Moonshine's song motif

(b) Bear's song motif



(c) Scotch's song motif

(d) Julep's song motif

Figure 3.1: Spectrograms of the four birdsongs used in Experiments 1- 5. The x-axis represents time, the y-axis represents frequency, and amplitude is represented by the color map. The dark blue lines between syllables represent the inter-syllable intervals, which have been band-rejected at all frequencies to produce pure silence between syllables.

ing syllables, the overall spectral content remains the same, whereas fine structure (i.e. small scale timing) is changed. Targets consisted of the original motif, with only one syllable revered at a time. This allowed individual syllables in the song to be tested and compared to one another. Examples of these two types of changes to song are illustrated in Figures 3.2 and 3.3.

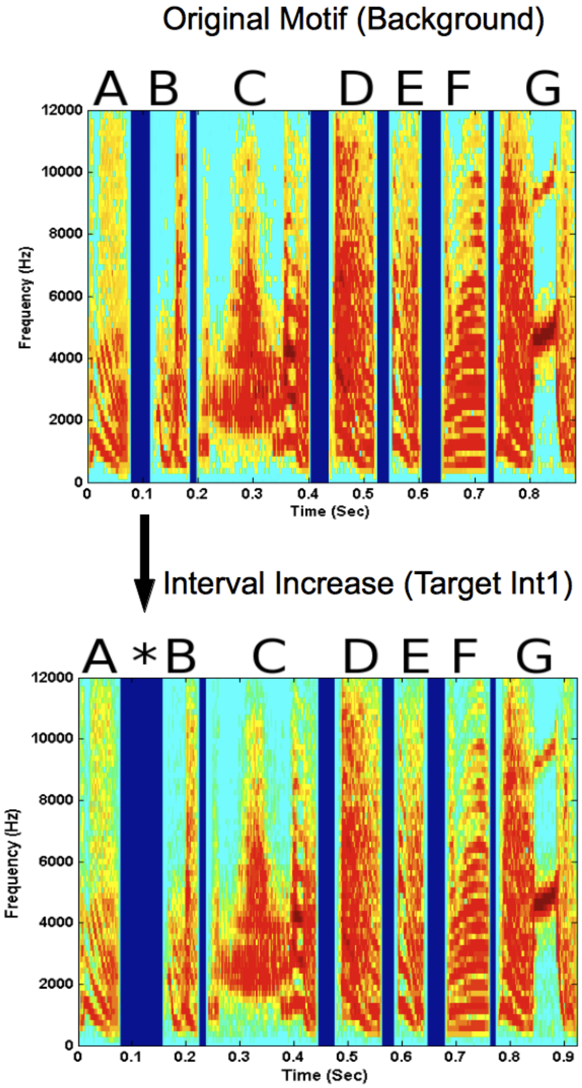


Figure 3.2: Original song motif (top) that serves as the repeating background, and the same motif with the first interval doubled (bottom) that serves as a target. Dark blue in the spectrogram indicates pure silence between the syllables.

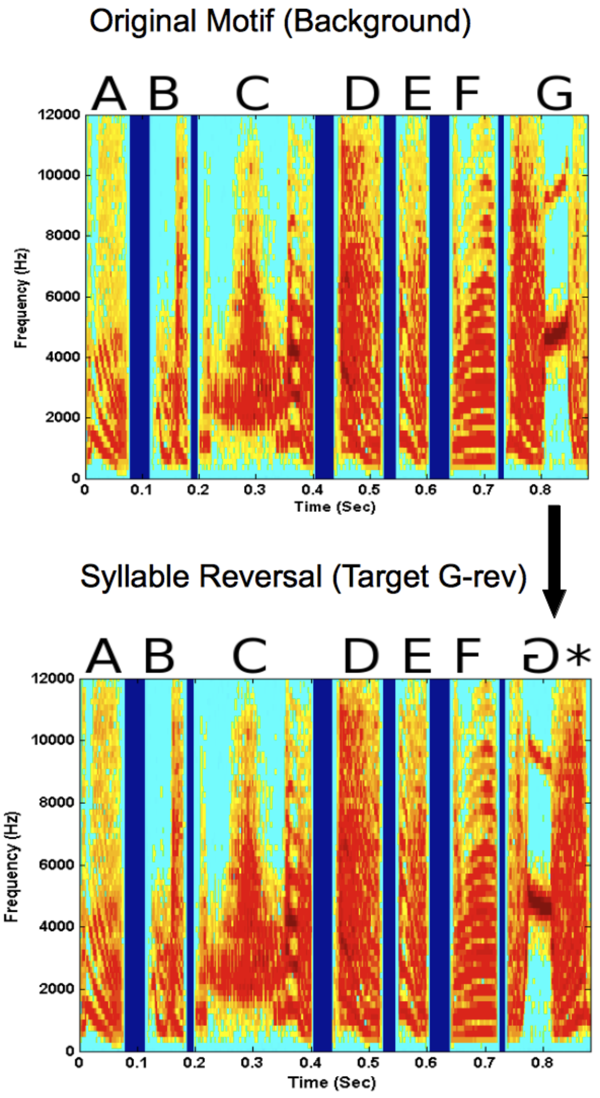


Figure 3.3: Original song motif (top) that serves as the repeating background, and the same motif with the last syllable reversed in time (bottom) that serves as a target. Dark blue in the spectrogram indicates pure silence between the syllables.

Table 3.1: Specific targets for Experiment 1: Natural song

Moonshine's song (A,B,C,D,E,F)	**Bear's song (A,B,C,D,E)**	**Scotch's song (A,B,C,D,E,F,G)**	**Julep's song (A,B,C,D,E)**
1. Interval 1	1. Interval 1	1. Interval 1	1. Interval 1
2. Interval 3	2. Interval 2	2. Interval 4	2. Interval 2
3. Interval 5	3. Interval 4	3. Interval 6	3. Interval 4
4. A reversed	4. A reversed	4. A reversed	4. A reversed
5. C reversed	5. B reversed	5. C reversed	5. B reversed
6. E reversed	6. D reversed	6. E reversed	6. D reversed
7. F reversed	7. E reversed	7. G reversed	7. E reversed

Experimental esign

All subjects were tested on the same four stimulus sets. Since the same four zebra finches that provided the songs were also tested, each male was tested on the BOS and three conspecific songs. To prevent practice effects, the order in which stimulus sets were tested was randomized. Because the motifs differed in the number of syllables and intervals, the first, middle, and last positions within the motif were tested. For syllable reversals, two syllables in the middle position were tested, as well as the first and last syllables. For interval doublings, one middle interval was tested, as well as the first and last intervals. The set of 7 targets for each of the four songs is listed in Table 3.1.

Training

After initial training using pure tones, birds were acclimated to song in a training session. A song motif from a different bird was used specifically for training, and not for data collection. For this stimulus set, exaggerated changes were made to the song. There were two targets in this session. One target consisted of the second inter-syllable interval quadrupled in duration, and the other target consisted of the entire song reversed in time. The interval target was presented on 30% of trials, and the reversal target was presented on 40% of trials. The remaining 30% of trials in the session are sham trials in which no target was alternated with the background. Birds ran on 2 sessions, or 200 trials of the training set before running on an experimental stimulus set.

Testing procedures

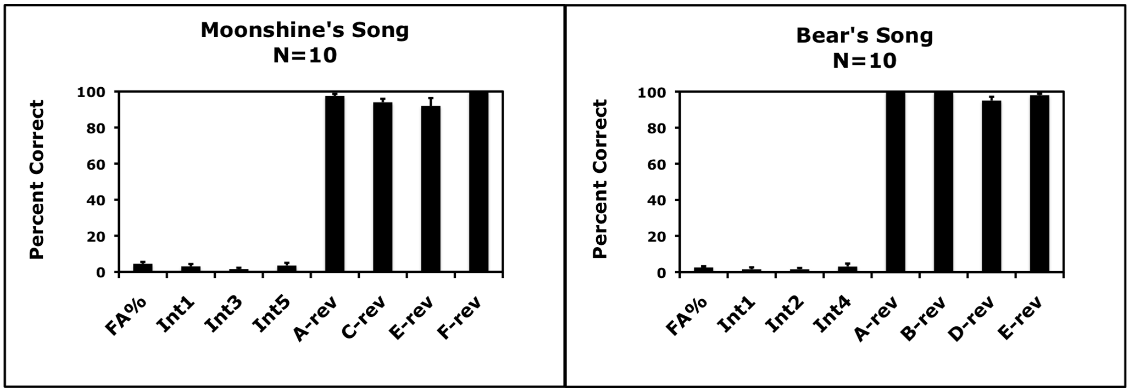
Song motifs were presented at 70 dB SPL. Motif durations ranged from 658-902 msec long. For each session, the background motif was presented at a rate of once per 1500 msec. After a variable amount of time (2-6 seconds) during which the bird pecked the observation key, a target motif was inserted into the background. After target insertion, the bird had to peck the report key within the 2500 msec response interval for food reward. This was recorded as a "hit". If the bird failed to peck the report key, this was recorded as a "miss". Sham trials in which there was no target insertion occurred on 30% of trials. If the bird pecked the report key during the sham trial, this was recorded as a "false alarm". If the bird withheld pecking during a sham trial, this was recorded as a "correct rejection". The order in which target and sham trials were presented was randomized from block to block, with each block of 10 trials containing the 7 target trials and 3 sham trials. Birds were run on 300 trials for each stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate did not exceed 20%. One bird (female) had a false alarm rate of 31.67% for Julep's song, and was not included in the analysis for this stimulus set.

3.2.3 Results

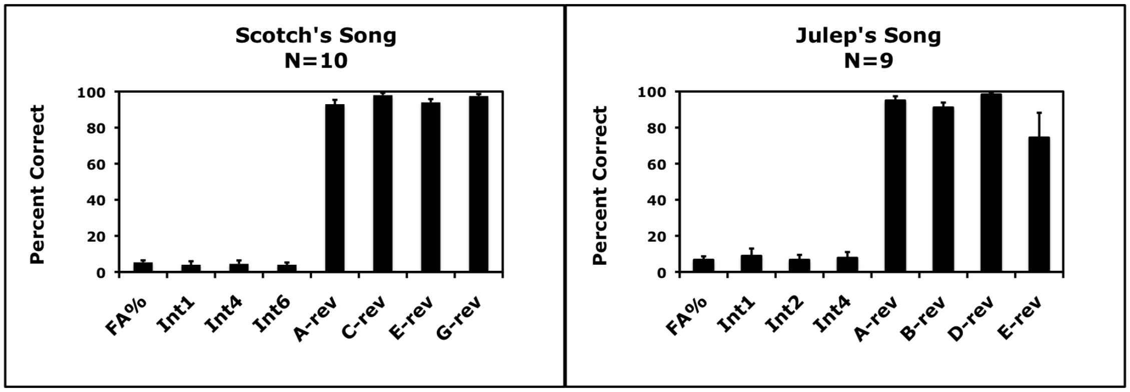
The average performance for all ten birds tested is shown in Figure 3.4. Birds discriminated single interval doublings on average $1.50 - 8.89\%$ of the time, and single syllable reversals on average $74.40 - 100.00\%$ of the time. In other words, all birds tested, regardless of gender or species, were much better at discriminating single syllable reversals then they were at discriminating single interval doublings. For each song, hit rates for syllable reversal targets were significantly higher than for interval doubling targets (Mann-Whitney Rank Sum Test, $p < 0.001$; for specific U values, see Table A.1 in Appendix A). Since this was true for all songs tested, this suggests that syllables in general are more salient than intervals, and that this is song independent. Because of this, most subsequent analyses will be across all songs tested, instead of within individual songs.

Separate plots for average performance in males, females, and budgerigars are shown in Figures B.1, B.2, and B.3 of Appendix B. All three groups performed similarly, and there were no differences in discrimination rates between the groups for intervals (Kruskal-Wallis ANOVA; H=5.651, df=2, $p > 0.05$), across all songs tested. However, for syllable reversals, male zebra finches had significantly lower discrimination rates compared with budgerigars (Kruskal-Wallis ANOVA, H=11.906, df=2, p=0.003). Discrimination performance on single syllable reversals will be examined further in Experiment 2, in which all three groups were tested on reversals of all syllables in all song sets.

Human subjects showed a much different pattern of performance, in which they were able to discriminate changes to some intervals, but could not discriminate changes to all syllables tested. Results are shown in Figure 3.5. In general, humans performed best for the first and middle intervals, but could not discriminate changes



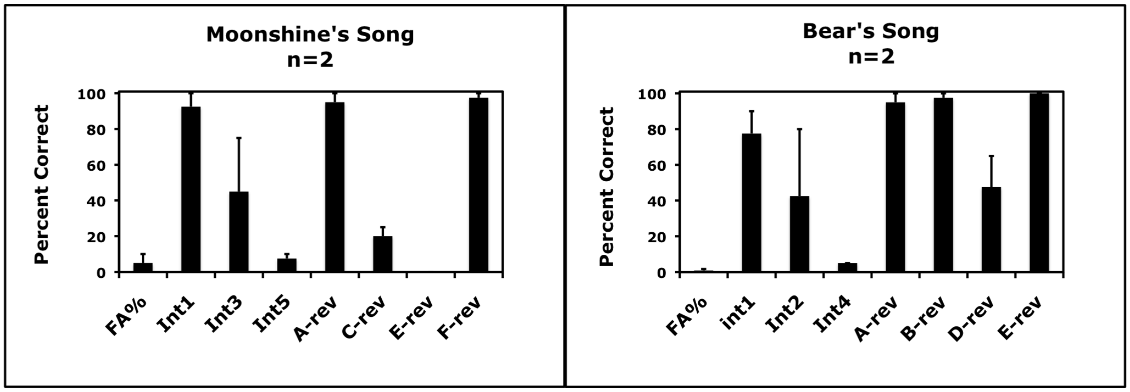
(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



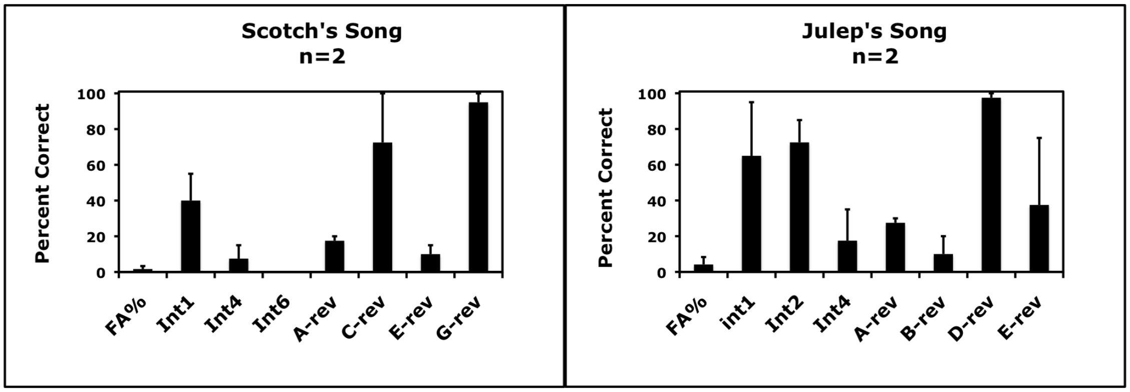
(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 3.4: Average discrimination performance of all birds (4 males, 3 females, 3 budgerigars) on single interval doublings and single syllable reversals in natural song, presented in the same testing session. Error bars show standard error of the mean. One bird (female) that had a false alarm of 31.67% for Julep's song (Figure 3.4(d)) is not shown.

to the last interval in song. Average hit rates for interval changes ranged from 40.00 – 92.50% for the first interval, 7.50 – 72.50% for the middle interval, and 0.00 – 17.50% for the last interval. For single syllable reversals, human subjects were only able to reliably discriminate 8/16 syllable reversals. All other syllable reversals were discriminated less than 50% of the time.



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 3.5: Average discrimination performance of human subjects on single interval doublings and single syllable reversals in natural song, presented in the same testing session. Error bars show standard error of the mean.

3.2.4 Discussion

All birds were substantially better at discriminating single syllable reversals compared with single interval doublings, suggesting this type of change to song is more salient. This result is universal across all birds, regardless of gender or species. Thus, vocal experience with song does not affect performance, as female zebra finches and budgerigars also showed the same pattern of performance as male zebra finches. Results were similar for the four acoustically distinct songs tested, suggesting fine structure salience is a general mechanism rather than a song-specific one. Human subjects were able to discriminate changes to the first and middle intervals in most songs, but could only discriminate half of the single syllable reversals presented. This suggests that birds and humans differ in how they perceive song, with humans paying attention to global temporal structure, whereas birds focus on local temporal structure. A previous study by Nespor and Dooling (1997) found that zebra finches and budgerigars were able to discriminate duration changes to intervals in one song tested. Birds were able to discriminate changes to the first, third, and sixth intervals as short as 10 msec in duration. This result is vastly different from our result that no interval doublings could be discriminated greater than 20% of the time, regardless of the song tested. It is possible that this study did not filter the recording noise between syllables, and this provided an additional cue when adding duration to the intervals. Our pilot work in which recording noise was not filtered from intervals agrees with this possibility, as birds had higher hit rates for interval doublings that contained recording noise, than interval doublings in which there was pure silence between syllables.

While intervals and syllables occur together in song, they are quite different from one another. One difference between intervals and syllables is their naturally occurring durations, and we wondered whether this played a role in discriminability of changes to intervals and syllables. The original durations of the intervals tested in this experiment ranged from 15-57 msec, with an average duration of 36.08 ± 10.53 msec. The durations of the syllables tested in this experiment ranged from 30-203 msec, with an average duration of 108.63 ± 55.30 msec. While syllables are generally much longer in duration than intervals, there is still some overlap in this stimulus set. Reversals of shorter syllables with durations of 52, 53, and 55 msec were easily discriminated whereas doublings of intervals with durations of 48 and 57 msec were not. In addition, doubling the interval durations resulted in intervals that ranged from 30-114 msec, which is most certainly within the range of syllable durations. Thus, it seems likely that it is the type of change, and not the duration over which the change occurs, that determines discriminability.

Another difference between intervals and syllables that may explain these results is the variation with which they are produced. Glaze and Troyer (2006) examined the durations of intervals and syllables in song, and found that the coefficient of variation is about 1.5 times greater for intervals than it is for syllables. In addition, tempo changes in song affect the durations of intervals more than syllables. In other words, when songs are sped up or slowed down, the intervals tend to stretch and compress whereas syllable durations are more stable. It is possible that because intervals are normally sung with some amount of variability, changes to interval duration are not particularly salient.

The following two experiments examine discrimination performance when only one type of change is presented in a test session. In this experiment, syllable and interval changes to song were directly compared with one another. One additional hypothesis as to why interval changes were not discriminated is that syllable changes are so salient for birds that interval changes are overlooked when tested in the same test session. This still leaves the question of whether birds are able to discriminate interval changes at all. To answer this question, Experiment 3 examines the ability of birds to discriminate changes to interval duration when the stimulus set only contains interval changes. Experiment 2 further examines bird's ability to discriminate syllable reversals by testing single reversals of all syllables in all four songs.

3.3 Experiment 2: Discriminability of syllable reversals at all locations within the song motif

3.3.1 Introduction

This experiment is an extension of Experiment 1, further examining the ability of zebra finches to discriminate changes to individual syllables in the song motif. Zebra finch song typically contains 3-8 syllables in a motif, and each syllable has unique acoustic properties, due to the spectral and temporal fine structure in the syllables. Since birds have been shown to be quite sensitive to fine structure in syllables, we wanted to know whether this was the case for all syllables. The goal of this experiment was to determine if all syllable reversals are equally discriminable, and whether syllable type, syllable duration, or syllable location within the motif is a determinant of discriminability.

3.3.2 Methods

Subjects

Four male zebra finches, three female zebra finches, and three female budgerigars were used in this experiment.

Stimuli

Targets consisted of single syllable reversals (example shown in Figure 3.3) at each location within the song motif. In two songs that contain especially complex syllables (Julep and Bear's songs), one additional target was a motif in which just a portion of the entire syllable (termed a sub-syllable) was reversed. For this experiment, all

Table 3.2: Specific targets for Experiment 2: Natural song

Moonshine's song (A,B,C,D,E,F)	**Bear's song (A,B,C,D,E)**	**Scotch's song (A,B,C,D,E,F,G)**	**Julep's song (A,B,C,D,E)**
1. A reversed	1. A reversed	1. A reversed	1. A reversed
2. B reversed	2. A2 reversed	2. B reversed	2. B reversed
3. C reversed	3. B reversed	3. C reversed	3. C reversed
4. D reversed	4. C reversed	4. D reversed	4. C2 reversed
5. E reversed	5. D reversed	5. E reversed	5. D reversed
6. F reversed	6. E reversed	6. F reversed	6. E reversed
7. All reversed	7. All reversed	7. G reversed	7. All reversed

target motifs had the silence separating the two sub-syllables band-reject filtered so that there was pure silence between them. This provides a boundary between the sub-syllables and prevents acoustic artifacts when reversing a single sub-syllable. This type of sub-syllable target is illustrated in Figure 3.6. Performance on these sub-syllable reversals may provide insight into whether birds attend to all of the syllable, or only certain portions when listening to song. For the three songs that contain less than 7 syllables, the remaining targets are motifs in which all syllables were reversed in time, but remained in the same sequential order (termed syllable reversed songs). This type of target is illustrated in Figure 3.7.

Experimental Design

All subjects were tested on the same four stimulus sets. For male zebra finches, birds were tested on the (BOS), and the other three conspecific songs. The target set for each of the birdsongs is listed in Table 3.2.

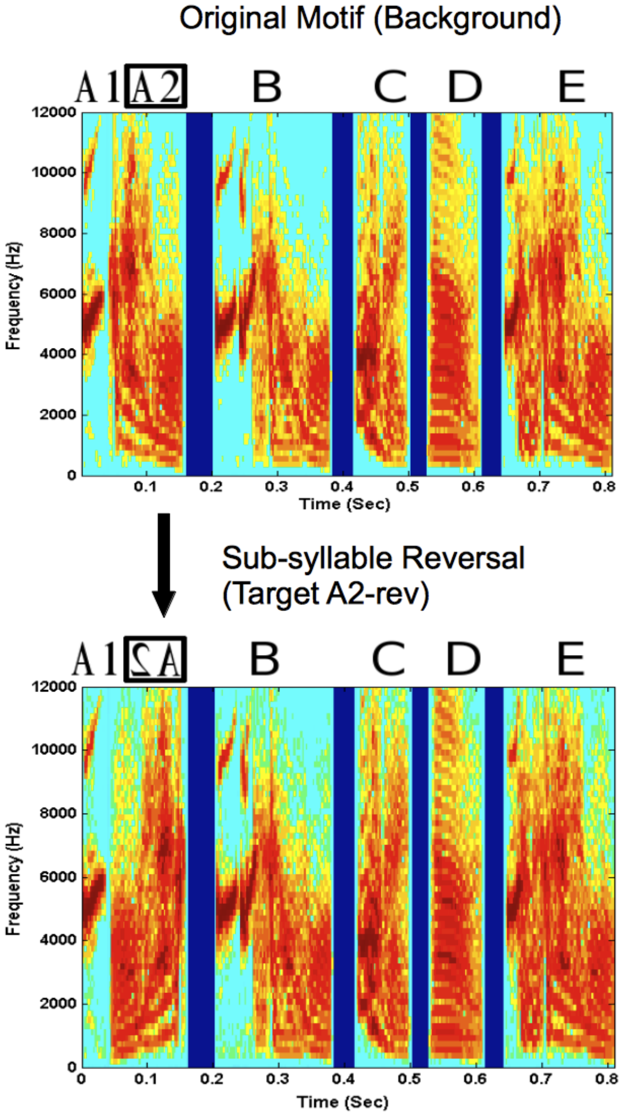


Figure 3.6: Original song motif (top) that serves as the repeating background, and the same motif with sub-syllable A2 reversed in time (bottom) that serves as the target. Dark blue in the spectrogram indicates pure silence between the syllables. In all motifs in this stimulus set (including the background motif) the interval between A1 and A2 has been filtered so that there is also pure silence between the sub-syllables. This boundary between A1 and A2 allows for the reversal of just A2 without acoustic artifacts.

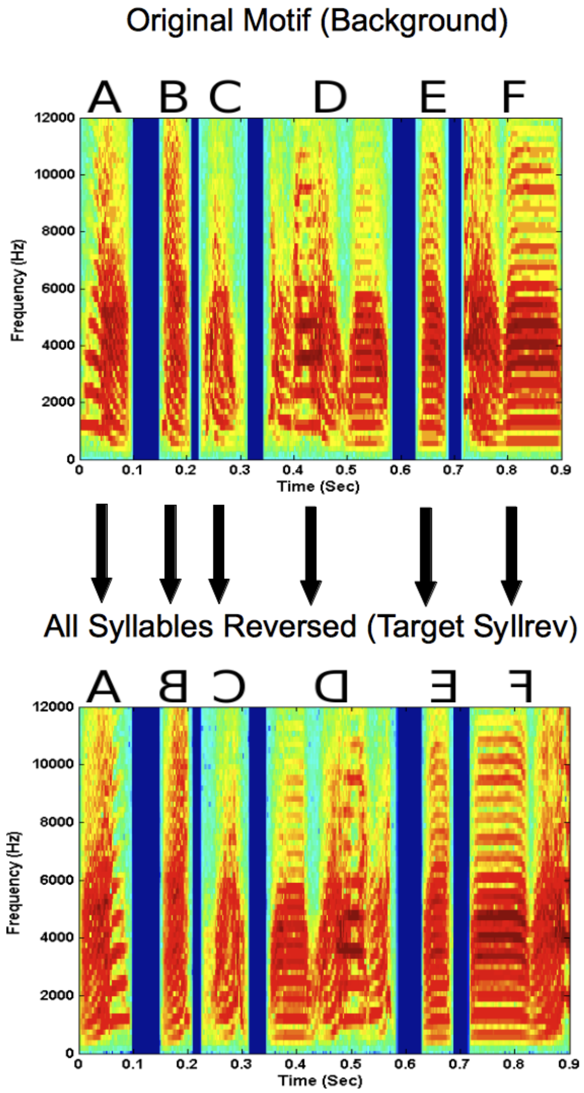


Figure 3.7: Original song motif (top) that serves as the repeating background, and the same motif with each individual syllable reversed in time (bottom) that serves as the target. Note that syllables remain in the original sequential order. Dark blue in the spectrogram indicates pure silence between the syllables.

Training

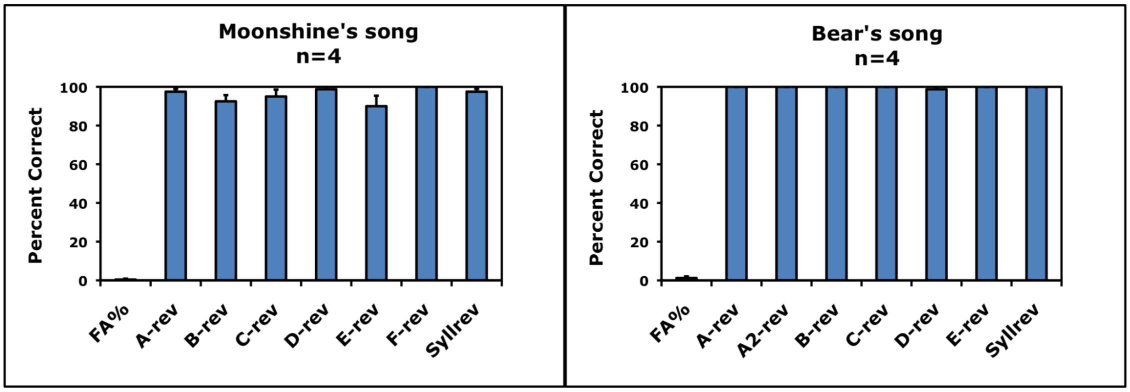
Since birds had already been acclimated to song previously, there were no training procedures for this experiment.

Testing Procedures

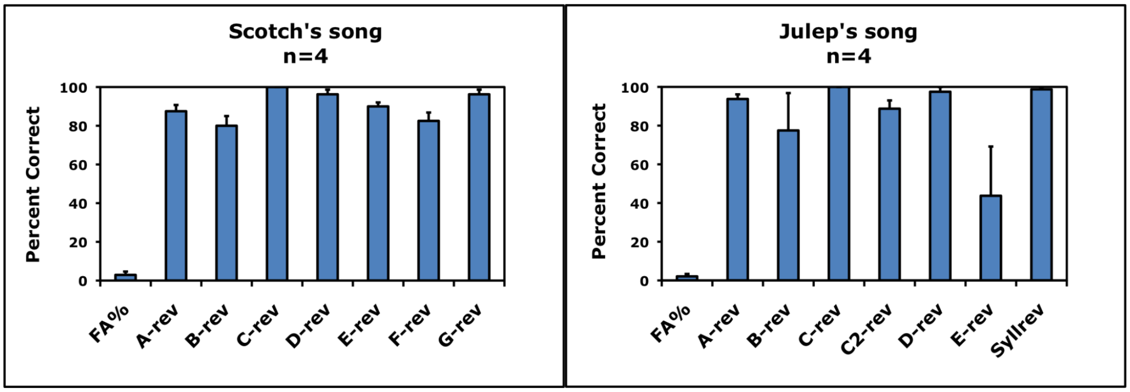
The background motifs were presented at 70 dB SPL, at a rate of once per 1500 msec, as in Experiments 1. After a variable amount of time (2-6 msec) a target was presented. After target presentation, birds had to peck the report key within 2.5 seconds for food reward. Failure to peck within 2.5 seconds resulted in a “miss”. For each 10-trial block, 7 targets (see Table 3.2) and 3 sham trials were presented in random order. Birds were run on 300 trials for each stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate did not exceed 20%.

3.3.3 Results

Average hit rates ranged from $77.50 - 100.00\%$ for males, $85.00 - 100.00\%$ for females, and $71.67 - 100.00\%$ for budgerigars, across all songs tested. Performance on each song is shown in Figure 3.8 for males, Figure 3.9 for females, and Figure 3.10 for budgerigars. These results show that birds are quite good at discriminating single syllable reversals within song motifs. For all birds, hit rates were well above the false alarm rate, indicating that this task is relatively easy and birds are performing above chance, and in fact near perfect on many syllables.



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

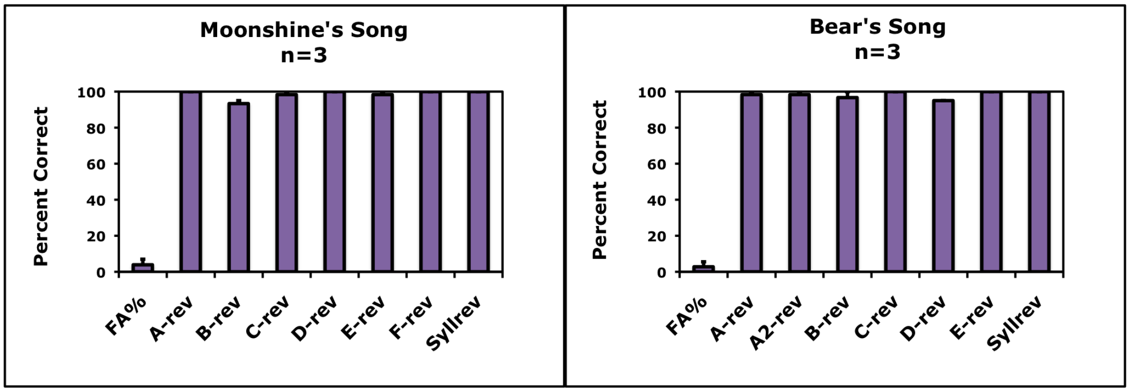


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

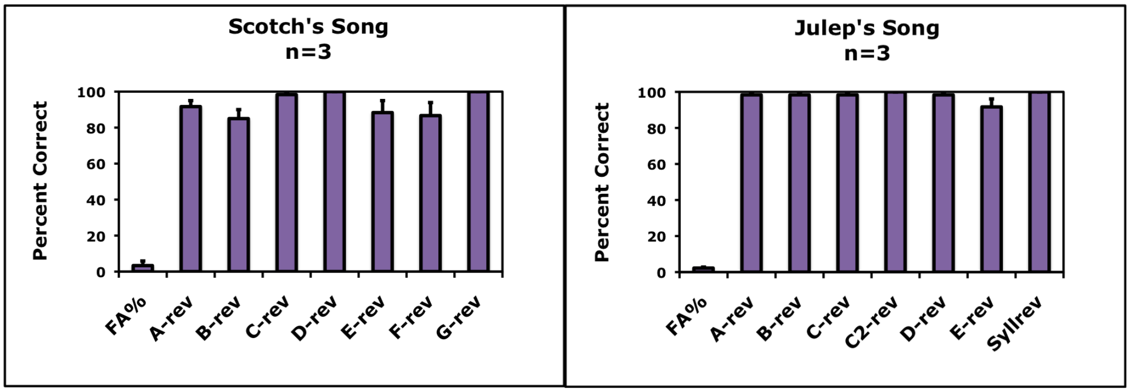
Figure 3.8: Average discrimination performance of male zebra finches on single syllable reversals at all locations within the natural song motif. Error bars show standard error of the mean.

Bird's Own Song Effect

It has been suggested that male zebra finches are most sensitive to their own songs, as they have specific neurons in the avian forebrain that respond best to the bird's own song, and very little to conspecific songs (Solis and Doupe 1997; Theunissen and Doupe 1998). This might display behaviorally as higher hit rates, or shorter response latencies for discrimination of changes to the bird's own song compared with conspecific songs. Since hit rates for most syllable reversals were very similar, response latencies were analyzed to determine if a BOS effect existed.



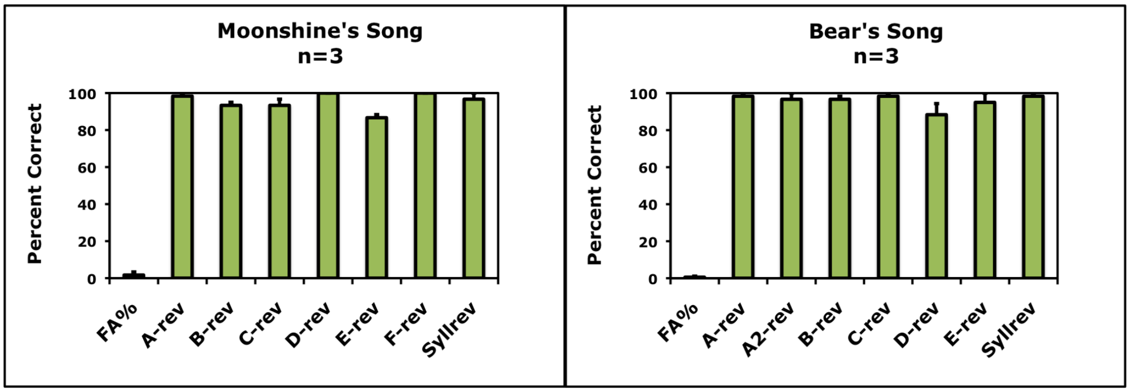
(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



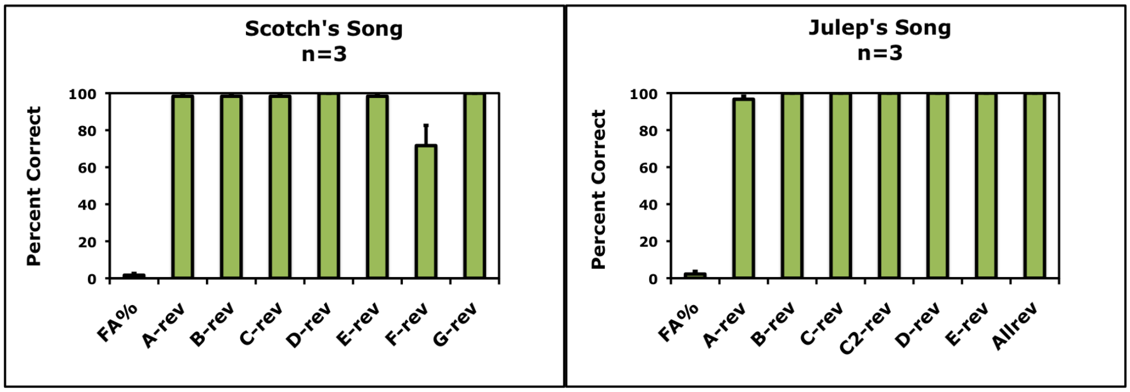
(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 3.9: Average discrimination performance of female zebra finches on single syllable reversals at all locations within the natural song motif. Error bars show standard error of the mean.

There was no BOS effect for male zebra finch performance on single syllable reversals. For each bird, average response latencies were shortest for Bear's song. Median latencies for Bear's song were almost always the shortest, and were significantly shorter than the longest median response latency (Kruskall-Wallace ANOVA, post-hoc Dunn's, for specific values see Table A.2 in Appendix A). This means that birds are attending to the acoustic properties of songs, rather than song identity. If birds were attending to song identity, then we woud expect each bird's response latencies to be shortest for his own song, and longer for the conspecific songs. Instead, each bird's response latencies are shortest for the same song, indicating that



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 3.10: Average discrimination performance of budgerigars on single syllable reversals at all locations within the natural song motif. Error bars show standard error of the mean.

birds are attending to the same features in this song.

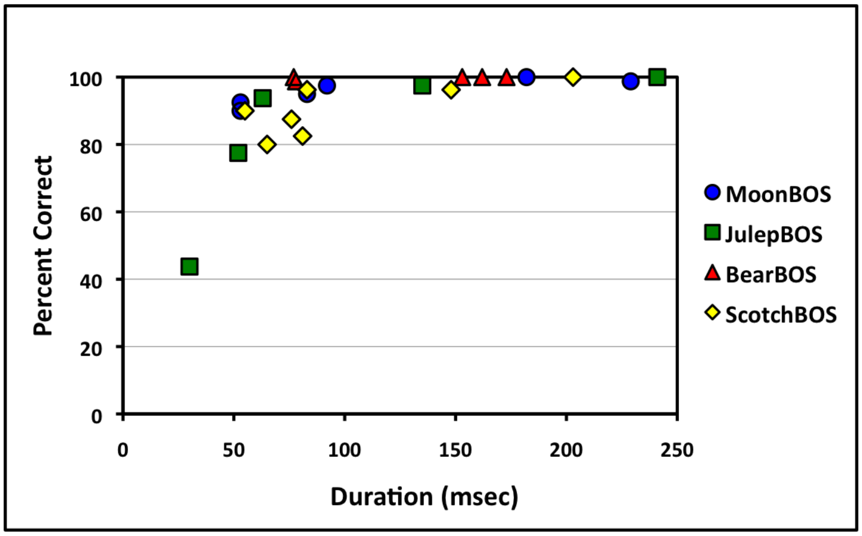
Syllable Effects

We wanted to know whether discrimination performance was the same for all syllables in the set, or whether there were systematic differences in perception of the syllables. For each of the three groups, a Friedman test was performed across the 25 syllables (23 syllables and 2 sub-syllables), taking into account individual performance. All three groups showed a syllable effect in which there were significant differences in discrimination performance between individual syllables

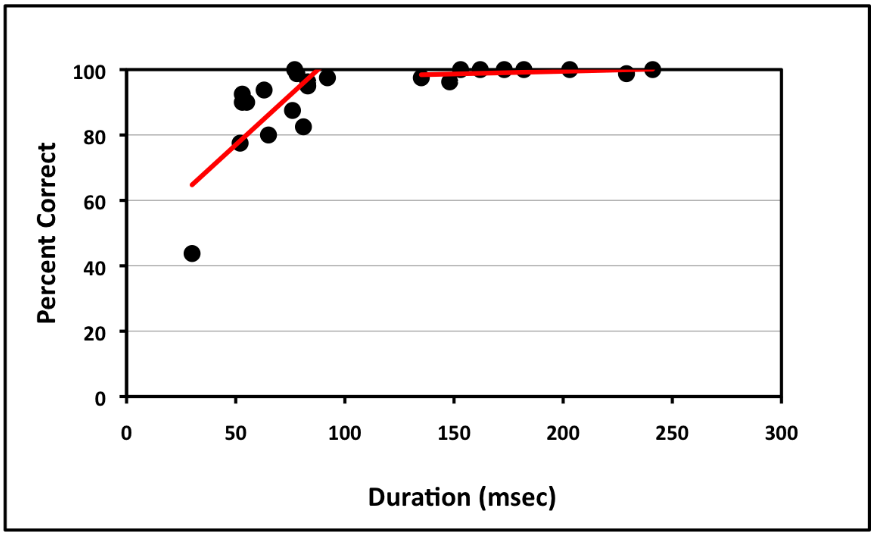
(Males: Chi-Square=64.33, df=24, p=0.000; Females: Chi-Square=41.87, df=24, p=0.013; Budgerigars: Chi-Square=44.86, df=24, p=0.006). Pairwise comparisons using Tukey's least significant difference method showed all three groups to have different patterns of performance. To explain these patterns of performance, we examined three qualities of syllables: duration, position, and syllable type.

A duration effect on performance was found only for male zebra finches. For all three groups, discrimination performance stabilized for syllables more than 100 msec in duration. Thus, for syllables that were longer than 100 msec, reversals were discriminated nearly 100% of the time. However, for syllables shorter than 100 msec in duration, performance was positively correlated with syllable duration for male zebra finches (Pearson correlation; $r(21) = 0.725, p = 0.003$), but not for females ($r(21) = 0.258, p = 0.373$) or budgerigars ($r(21) = -0.215, p = 0.461$). Average discrimination performance vs. syllable duration is plotted for male zebra finches in Figure 3.11. Each point represents a syllable within one of the four songs. In Figure 3.11(a) syllables from individual birds are marked by color and shape. Figure 3.11(b) demonstrates that performance follows two different patterns, one for syllables under 100 msec, and one for syllables over 100 msec in duration.

A position effect on performance was found reliable only for budgerigars (Kruskal-Wallis one-way ANOVA, H=16.680, df=4, p=0.002). To test for differences in performance based on syllable position, syllables were grouped according to the 5 positions: first, second, middle, penultimate, and last. Post-hoc comparisons using Dunn's method showed that for budgerigars, performance on syllables in the penultimate position was significantly lower than syllables in the middle and last positions ($p < 0.05$). No position effect was found for male zebra finches (H=4.079, df=4, p=0.395). For females, a position effect was found (Kruskal-Wallis one-way ANOVA, H=10.440, df=4, p=0.034). However, this effect was not robust, as a post-



(a) Separated by song



(b) Separated by trend

Figure 3.11: Discrimination performance of male zebra finches on single syllable reversals as a function of syllable duration.

hoc Dunn's test did not find any significant differences between individual groups ($p > 0.05$). In addition, removal of the 2 sub-syllables from the analysis resulted in no position effect (H=8.968, df=4, p=0.062).

A significant effect of syllable type on performance was found for male zebra finches and budgerigars, which is consistent with the duration and position effects found in these groups, respectively. This will be explained below. To test the effect of syllable type on discrimination performance, syllables were categorized into 5 types: stacks, sweeps, noisy, high, and combo. Stacks are syllables that have flat harmonics and a tonal quality. Sweeps are syllables in which the harmonics sweep downward across time. Noisy syllables are short, and do not have any defining acoustic features. They appear noisy in a spectrogram. High syllables have peak power concentrated in the 4-8 kHz range. Combo syllables are syllables that contain two or more types, with less than 5 msec of silence separating them. All syllables over 100 msec were combo syllables, and all syllables less than 100 msec fell into the remaining 4 categories.

Male zebra finches showed a significant effect of syllable type on performance (Kruskal-Wallis one-way ANOVA, H=34.412, df=4, $p < 0.001$). Post-hoc tests using the Dunn's method showed that performance on combo syllables was significantly higher compared with the four remaining groups ($p < 0.05$). There were no other differences between groups. This explains the saturation in discrimination performance for syllables greater than 100 msec in duration. All of these syllables contain at least two distinct types of sub-syllables that become flipped in order when they are reversed in time, making reversals easy to discriminate. This is analogous to words that contain multiple syllables. For instance, the word "BIRDSONG" becomes "ƆИOƧᗡЯIᙠ". Reversing these words in time not only reverses the fine structure of each individual syllable (bird and song), but also reverses the order in which these syllables occur. This is an additional cue that can be used in reversal discrimination. Syllables that were shorter than 100 msec were equally distributed among the stack, sweep, noisy, and high groups, indicating that the duration effect seen for these short syllables is independent of syllable type. Budgerigars also showed a significant effect of syllable type on discrimination performance (Kruskal-Wallis one-way ANOVA, H=18.411, df=4, p=0.001). Post-hoc tests using the Dunn's method showed that performance was lowest for stack syllables, and was significantly lower compared with the combo syllable group ($p < 0.05$). Interestingly, three out of the four syllables in the penultimate position were stack syllables. Therefore, it is unclear whether the position effect seen in budgerigars is explained by syllable type, or vice versa. Results from experiments in Chapter 5, using Schroeder waveform harmonics suggest that it is syllable type, and not position that explain differences in syllable reversal discrimination.

Response latencies suggest a window of attention smaller than the motif

Given that all three groups had similarly high hit rates for most single syllable reversals, response latencies were also analyzed, as they can give a more precise measurement of discrimination performance. Raw response latencies for all four motifs showed pattern in which response latencies increased as the position of the syllable reversal within the motif increased. Thus, response latencies were the shortest for reversals of the first syllable, and longest for reversals of the last syllable in the motif. This pattern was the same for all species, and means that all birds used the same strategy for listening and making these discriminations. Birds listened to the motif, and responded after hearing the change, rather than listening to the entire motif, and then responding. This suggests that birds are listening along an attentional window that is shorter than the motif, rather than listening to the motif

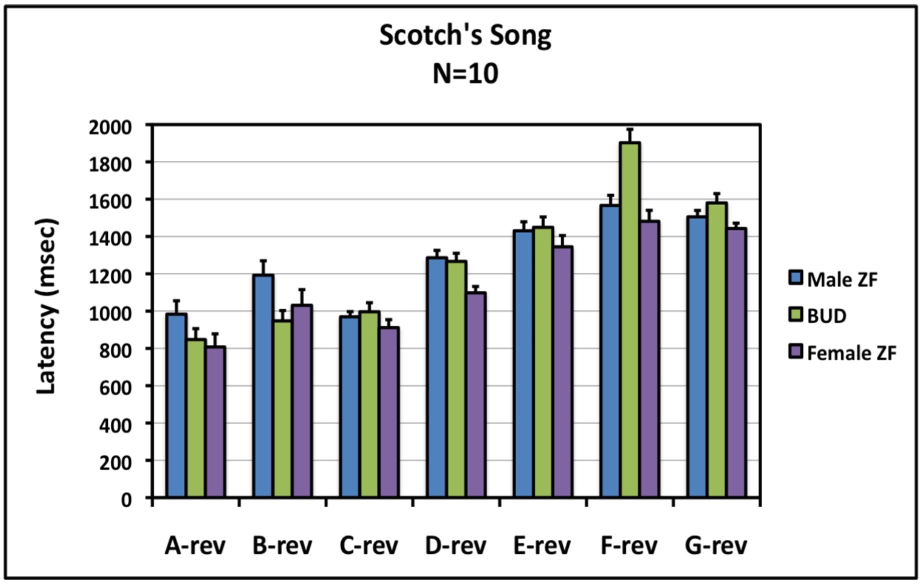


Figure 3.12: Average response latencies for syllable reversals in Scotch's motif. Male zebra finches, female zebra finches, and budgerigars are compared.

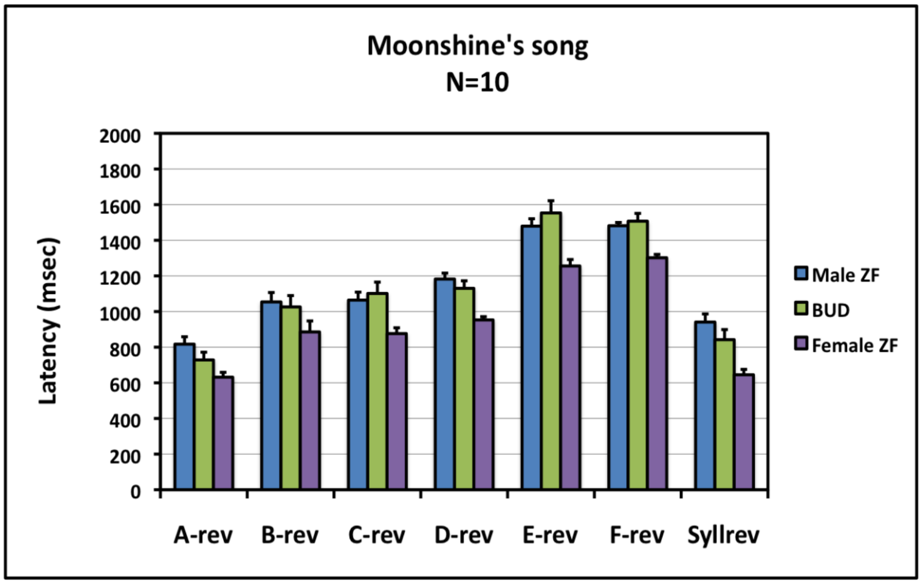


Figure 3.13: Average response latencies for syllable reversals in Moonshine's motif. Male zebra finches, female zebra finches, and budgerigars are compared.

as a whole. This also means that birds are able to make decisions about the motif without listening to it in its entirety. Raw latencies for the four songs are displayed in Figures 3.12, 3.13, 3.14, and 3.15

Response latencies for sub-syllables in Bear's and Julep's motifs further confirm

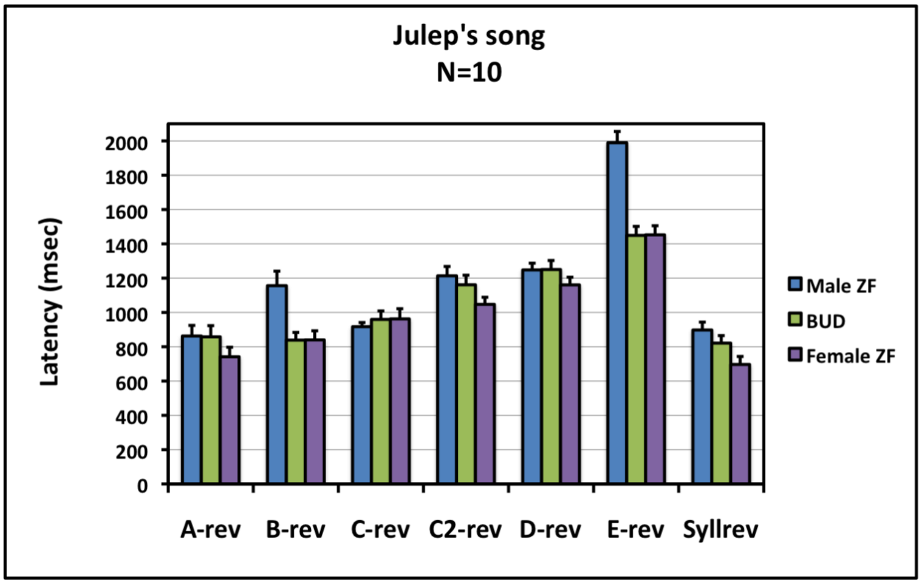


Figure 3.14: Average response latencies for syllable reversals in Julep's motif. Male zebra finches, female zebra finches, and budgerigars are compared.

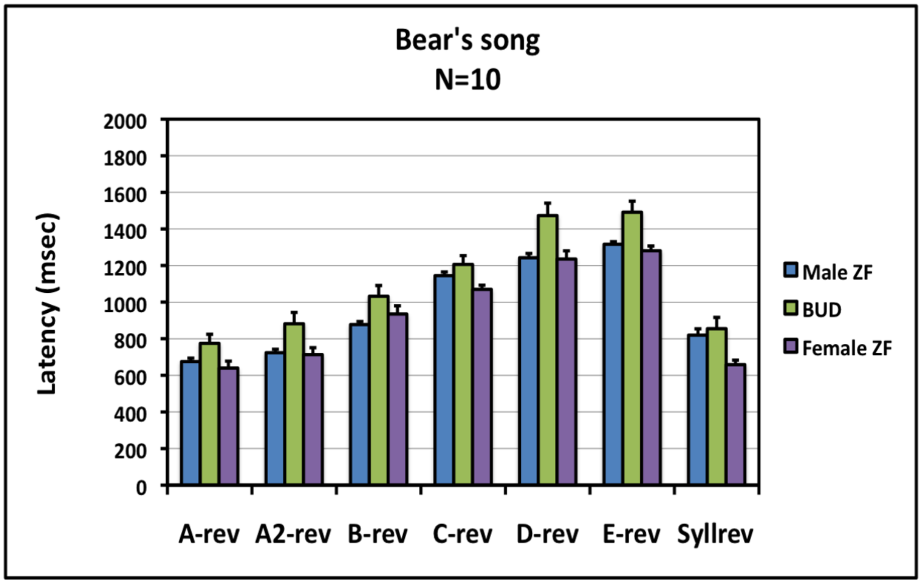


Figure 3.15: Average response latencies for syllable reversals in Bear's motif. Male zebra finches, female zebra finches, and budgerigars are compared.

that birds are listening to song motifs along a small window of attention. The average latencies for reversals of sub-syllables A2 in Bear's motif and C2 in Julep's motif are slightly higher than for reversals of the entire syllable (see Figures 3.14 and 3.15). These sub-syllables are the second portion of syllables A and C, and thus reversal of just these sub-syllables occurs later than if the entire syllable was reversed. We would expect reversals of just the first sub-syllable to have a similar response latency as the entire syllable. The fact that response latencies differ based on where in the syllable the reversal occurs is also an indicator that birds listen to song with the same temporal precision that they use to produce song.

Response latencies also revealed that male zebra finches may process song differently from females and budgerigars. In three of the song sets, one target consisted of a motif in which all syllables were reversed in time, but in the original order (Syllrev target). Male zebra finches had significantly longer response latencies for Syllrev targets than for A-rev targets (Mann-Whitney rank sum test, $p < 0.05$), even though both targets begin with the first syllable reversed. This was not the case for female zebra finches or budgerigars. This result is interesting, given that the target Syllrev begins the same way as the target A-rev. The only difference is that the target Syllrev contains B, C, D, etc. reversed as well, whereas the target A-rev (for any song), contains B, C, D, etc. in forward position. This suggests that even after males hear a change in song, they may still process any additional changes that occur within the same attentional window. Average latencies for these targets and specific Mann-Whitney values are shown in Table 3.3.

Lastly, significant group differences were seen in corrected response latencies, indicating differences in performance exist even though hit rates were similar between groups (Kruskall-Wallis, H=193.740, df=2, $p < 0.001$). Post-hoc tests using Dunn's method showed that all three groups were significantly different from one another, with males having the longest response latencies (median=712.00), females having the shortest (median=614.00), and budgerigars having intermediate response latencies (median=678.50). Response latencies are generally shorter for targets that are very different from the background, whereas they are longer when targets are more similar to the background. This means that females were able to discriminate targets from background with the greatest ease, whereas males had the most difficulty. This result does not seem to be a function of general response time, as this pattern is not seen for simple tone discrimination tasks. In this case, females had significantly slower response latencies (median=753 msec) compared with males and budgerigars (medians= 653 and 598 msec, respectively) (Kruskal-Wallis, H=52.305, df=2, $p < 0.001$). Instead, this result may be specific to song and song-like stimuli. Female zebra finches listen to directed song during mating displays, and their auditory experience with song may result in enhanced perception.

3.3.4 Discussion

All birds, regardless of gender or species, were able to easily discriminate most single syllable reversals within a set of four songs. Male zebra finch performance showed that all birds had the shortest response latencies for Bear's song. This suggests that birds are attending to the acoustic structure of song, and not the song identity.

For all three groups, performance saturated for syllables that were 100 msec in duration or longer. For syllables shorter than 100 msec, performance was positively correlated with syllable duration only for male zebra finches. Thus, males appear to be quite sensitive to the duration of syllables within song, whereas this is not the case with females and budgerigars. Additionally, Bear's song had the longest average syllable duration (3/5 syllables had durations greater than 100 msec), and this may be why male zebra finches had the shortest response latencies for this song.

Table 3.3: Results of Mann-Whitney Rank Sum Test comparing latencies for targets A-rev and Syllrev in Experiment 2

Group	**Song**	**Median Latency A-rev**	**Median Latency Syllrev**	**U**	**Sample Size**	**p**
Males	MoonBOS	689.00	808.50	4047.500	$n_1 = n_2 = 80$	0.004
	JulepBOS	669.00	779.50	4039.000	$n_1 = n_2 = 80$	0.004
	BearBOS	644.00	707.00	4127.500	$n_1 = n_2 = 80$	0.002
Females	MoonBOS	539.5	527.5	1844.500	$n_1 = n_2 = 60$	0.817 (n.s.)
	JulepBOS	624.00	616.50	1651.500	$n_1 = n_2 = 60$	0.437 (n.s.)
	BearBOS	551.00	581.00	2005.000	$n_1 = n_2 = 60$	0.283 (n.s.)
Bud	MoonBOS	632.00	706.50	2077.500	$n_1 = 59; n_2 = 60$	0.103 (n.s.)
	JulepBOS	668.00	731.50	1896.000	$n_1 = n_2 = 60$	0.616 (n.s.)
	BearBOS	727.00	861.00	1902.000	$n_1 = n_2 = 60$	0.594 (n.s.)

In two stimulus sets, Julep's and Bear's songs, reversals of a sub-syllable were tested as well as entire syllables (see Figures 3.8, 3.9, and 3.10). Males, females, and budgerigars discriminated these sub-syllable reversals on average 88.75 – 100.00% of the time, similar to the range of average hit rates for full syllables. Thus, even smaller changes to individual portions of syllables can be discriminated easily. It is possible that birds are listening on a finer time scale than the individual units of song, possibly along a sliding window. Response latency data further confirmed this, by showing that birds respond to changes to song almost immediately, instead of listening to the entire motif and then responding. Furthermore, for male zebra finches, targets with all syllables reversed had longer response latencies than targets with just the first syllable reversed. It is possible that after birds hear a change in song, they continue to process changes that occur in the temporal window, and thus take longer to respond.

3.4 Experiment 3: Discriminability of changes to inter-syllable interval duration within the song motif

3.4.1 Introduction

In Experiment 1, both interval duration increases and single syllable reversals were targets in the same stimulus set, testing which type of change to the song motif is more salient to the birds. In this experiment, only temporal envelope changes were presented. Since fine structure changes are so salient to birds, it is possible that birds ignore temporal envelope changes when fine structure changes are also present in the test set. Isolating temporal envelope changes allows us to ask if birds are capable of discriminating changes to inter-syllable intervals at all. This also allows for specific questions to be asked about perception of interval changes, such as whether original interval duration or position within of the interval within the motif affects duration discriminability.

3.4.2 Methods

Subjects

Four male zebra finches and one female zebra finch were used in this experiment.

Stimuli

Targets consisted of single interval doublings (shown in Figure 3.2) at each location within the song motif. Since all motifs contain less than 7 interval locations, the remaining targets consist of single syllable removals from the middle of the motif. An example of this target is shown in Figure 3.16. Removal of a syllable from the motif is a similar change to an interval increase in that a large gap is produced, which affects the temporal envelope of the song motif. Fine structure of the remaining syllables is not affected by this type of change to the song motif.

These additional targets were added to the target set so that all sessions contain 7 target trials and 3 sham trials per 10 trial block, as in Experiment 1. All further experiments also contain 7 target trials and 3 sham trials per block, so that all experiments are equally comparable in terms of target to sham ratio. This 7:3 ratio was chosen based on previous work in psychoacoustics discrimination experiments.

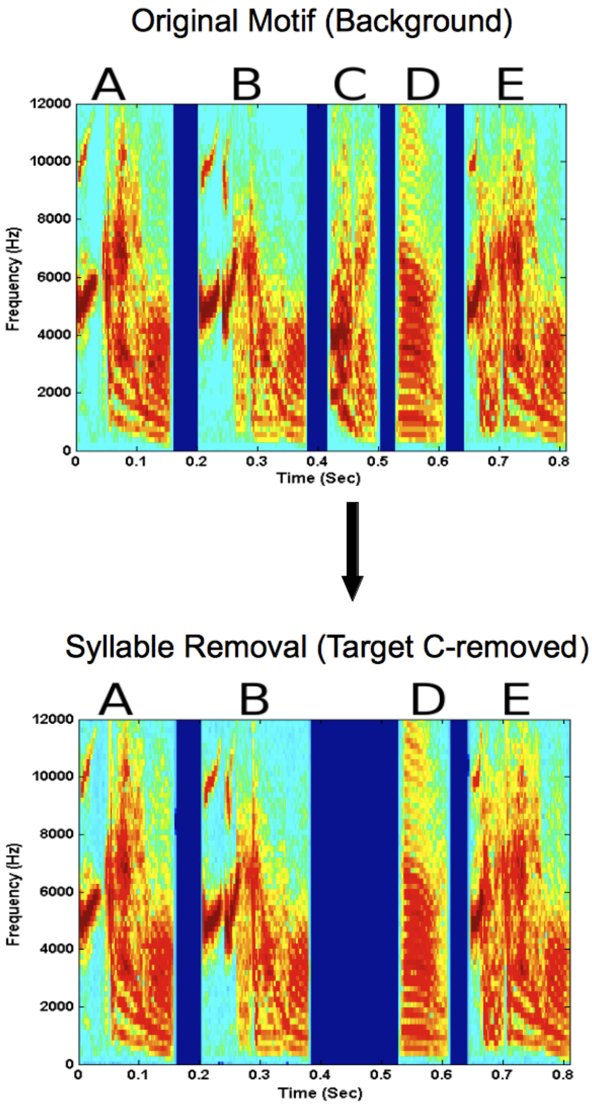


Figure 3.16: Original song motif (top) that serves as the repeating background, and the same motif with syllable C removed (bottom) that serves as the target. Dark blue in the spectrogram indicates pure silence between the syllables.

Table 3.4: Specific targets for Experiment 3: Natural song

Bear's song (A,B,C,D,E)	**Julep's song (A,B,C,D,E)**
1. Interval 1	1. Interval 1
2. Interval 2	2. Interval 2
3. Interval 3	3. Interval 3
4. Interval 4	4. Interval 4
5. B removed	5. B removed
6. C removed	6. C removed
7. D removed	7. D removed

Experimental Design

Due to the difficulty of this task, birds ran on a limited stimulus set of two songs. The set of 7 targets for both songs is listed in Table 3.4.

Training

Since birds had already been acclimated to song previously, there are no training procedures for this experiment.

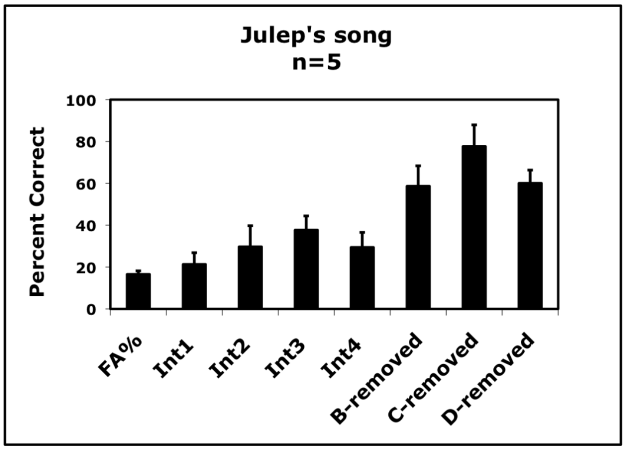
Testing Procedures

The background motifs were presented at 70 dB SPL, at a rate of once per 1500 msec, as in Experiments 1 and 2. After a variable amount of time (2-6 msec) a target was presented. After target presentation, birds had to peck the report key within 2.5 sec for food reward. Failure to peck within 2.5 sec resulted in a "miss". For each 10-trial block, 7 targets (see Table 3.4) and 3 sham trials were presented in random order.

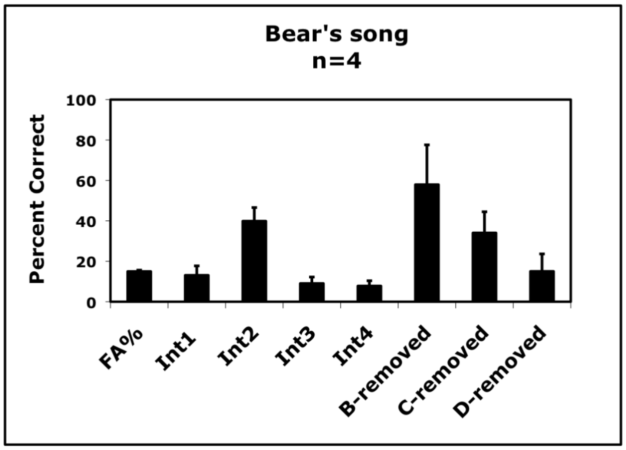
Due to the difficulty of this task, most sessions that birds ran had a false alarm rate above 20%. However, each bird started the session with a relatively low false alarm rate, and the rate increased as the session continued. Thus, only complete blocks of trials in which the false alarm rate did not exceed 20% were included in the analysis. Because of this, not every bird ran 200 valid trials. Birds ran between 50 and 200 valid trials for each song with the exception of one bird that did not run any valid trials on Bear's song. Because performance is presented as percent correct, any differences in the number of valid trials run by birds is normalized. There was no strong indication that performance improved with the number of valid trials ran. Percent correct hit rates for all birds were averaged together per song, since the hit rates for the female zebra finch did not differ significantly from the males for either song tested (Rank sum test, $p > 0.05$).

3.4.3 Results

Birds discriminated interval doublings on average less than 40% of the time. Figure 3.17 shows performance on both Bear and Julep's song. Analysis was done across both songs. These results indicate that poor performance on interval doublings in Experiment 1 is not a result of context, as birds are still unable to do the task even when only temporal envelope changes are tested. There was no effect of position on interval doubling discrimination performance (Kruskal-Wallis one-way ANOVA, H=4.364, df=3, p=0.225). Birds performed better on single syllable removals compared with interval doublings, and this reached significance (Rank sum test, U=232.000, $n_1 = 27; n_2 = 40$, $p < 0.001$). This is likely because this type of change results in a much larger gap between syllables, as well as changes the sequence of song. However, performance on single syllable removals was not significantly correlated with syllable duration (Pearson correlation, r(4)= 0.586, p=0.221).



(a) Discrimination performance: Julep's motif



(b) Discrimination performance: Bear's motif

Figure 3.17: Average discrimination performance of zebra finches on single interval doublings and single syllable removals in natural song, presented in the same testing session. Error bars show standard error of the mean.

3.4.4 Discussion

The results of this experiment show that when only temporal envelope (i.e. interval) changes are presented, birds do show a slight improvement in interval duration discrimination. However, performance was still relatively poor compared to performance on syllable reversals in the previous experiment. Interval increases were discriminated less than 40% of the time. In addition, performance on interval increases was similar to the false alarm rate. From the results of Experiments 1 and 3 it is clear that changes to interval durations in natural song are not salient to birds. Single syllable removals were more easily discriminated. However, even the highest average hit rate for a single syllable removal was lower than the average hit rate for almost all single syllable reversals in Experiment 1. These results provide further confirmation that changes to intervals, or the overall envelope of the song motif are not as easily discriminated by zebra finches as changes that occur within syllables.

There are two additional hypotheses for why temporal envelope changes are not as salient as fine structure changes. One is that the spectral structure of syllables may act as a distractor when birds listen to song. Given the diversity and complexity of spectral cues in syllables, perhaps birds are only attending to syllables and their structure, rather than the rate and rhythm with which syllables are sung. This hypothesis will be tested in Experiment 8, in which the spectral structure of syllables is replaced with random noise.

The second hypothesis is that changes to intervals are not easily discriminated, due to temporal auditory masking, which may occur as a result of fatigue in the auditory system. It has been shown in both humans and parakeets that the auditory threshold for detection of a pure tone is higher when that tone occurs after a noise burst, than when the tone occurs in isolation (Dooling and Searcy 1980). This is termed *forward masking*, and can occur up to 100 msec after the presentation of the succeeding burst. Since intervals between syllables are fairly short in duration (~ 40 msec), it is possible that forward masking is occurring during the intervals, resulting in an inability to discriminate interval changes. Although Dooling and Searcy (1980) found humans and parakeets to be similar in the the time course for recovery from forward masking, this test was done using tones and noise bursts. It is possible that humans and birds show different amounts of temporal masking when listening to song, and this may be why human subjects were able to discriminate interval changes to song in Experiment 1. The following experiment tests the hypothesis that poor interval discrimination in birds is due to forward auditory masking, by measuring discrimination performance on syllable reversals within a motif that does not contain intervals.

3.5 Experiment 4: Discriminability of syllable reversals in a song motif with limited global temporal information

3.5.1 Introduction

This experiment is a modification of Experiment 2 that is designed to ask the question of whether global temporal structure in song affects perception of fine structure within syllables. More specifically, are zebra finches able to easily discriminate single syllable reversals in a song motif in which there are no inter-syllable intervals? Inter-syllable intervals may aid in the auditory processing of syllables by allowing the allowing the auditory system to recover from fatigue. If this is the case, we would expect to see forward auditory masking, in which auditory thresholds occurring after syllables are elevated for a short amount of time. This could explain the insensitivity of zebra finches to interval changes, if temporal masking is occurring during the intervals. If temporal masking is occurring during intervals, then removal of intervals in the motif should result in decreased discriminability of fine structure changes in some syllables, since forward masking would then occur during the presentation of syllables.

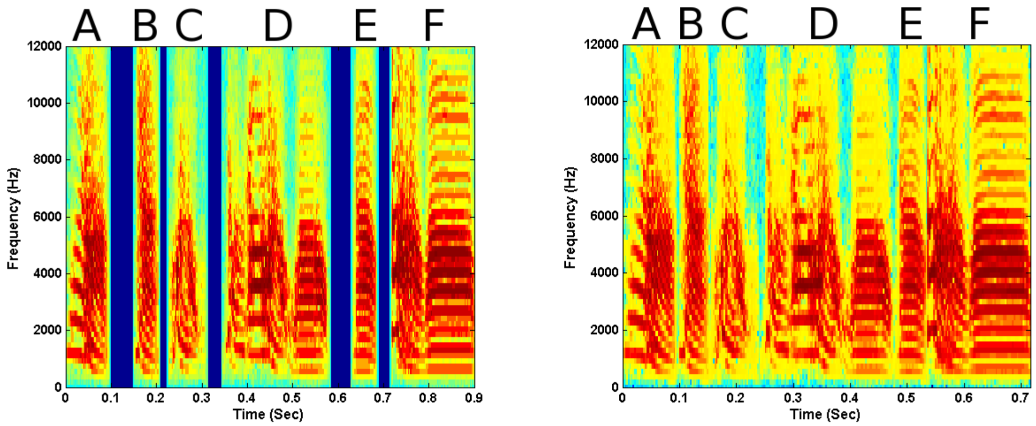
3.5.2 Methods

Subjects

Four male zebra finches were used in this experiment.

Stimuli

Stimuli consisted of the same songs used in Experiment 2, except that all inter-syllable intervals have been deleted from song. Song syllables were untouched, and contained the natural song envelope. Thus, the natural off ramp of one syllable, and the on ramp of the next syllable served as a boundary between the two syllables. However, the rate of delivery of syllables was immediate, as all intervals were abolished. An example comparing natural song, and the same song with no intervals is illustrated in Figure 3.18. Targets consisted of single syllable reversals, as in Experiment 2, however only full syllables were tested. In Experiment 2, sub-syllables were tested in Julep's and Bear's songs. For this experiment, those sub-syllable targets were replaced by targets in which all syllables are reversed in time, but remained in the correct order (target Syllrev).



(a) Natural song

(b) Song containing no intervals

Figure 3.18: Comparison of a natural song, and the same song with all inter-syllable intervals removed. The only difference between the two songs is the rate of delivery of the syllables. Dark blue in the spectrogram indicates pure silence between the syllables

Experimental Design

Each of the four male zebra finches was tested on the BOS, and the other three conspecific songs.

Training

Since birds were acclimated to song, no further training was necessary.

Testing Procedures

The background motif was presented at 70 dB SPL, at a rate of once per 1500 msec, as in all previous experiments. After a variable amount of time (2-6 msec) a target was presented. After target presentation, birds had to peck the report key within 2.5 seconds for food reward. Failure to peck within 2.5 seconds resulted in a "miss". For each 10-trial block, 7 targets and 3 sham trials were presented in random order. Birds were run on 300 trials for each stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate did not exceed 20%.

3.5.3 Results

Overall, reversal discrimination performance was very similar, regardless of whether inter-syllable intervals were present or not. Average hit rates for single syllable reversals in motifs with no intervals ranged from $68.75 - 100\%$, which is very similar to the range seen in Experiment 2 ($77.50 - 100\%$). A Mann-Whitney rank sum test confirmed that there was no significant difference in average performance between natural motifs, and motifs that had inter-syllable intervals removed (U=4834.000, $n_1 = n_2 = 92$, p=0.060).

3.5.4 Discussion

Birds were still able to discriminate single syllable reversals in a song that contained no intervals between the syllables. Thus, removing global temporal information (specifically tempo) does not affect perception of syllable structure in song. This result suggests that perception of syllable fine structure is not dependent upon perception of temporal envelope in song, and that forward auditory masking is not likely the reason for poor interval duration discrimination in Experiments 1 and 3. If forward masking did affect interval perception, then we would expect removal of intervals from the motif to subsequently affect syllable perception. This was not the case, as there was no difference in performance for single syllable reversal discrimination when intervals were removed. This could mean that forward masking does not take place when listening to song, or it could mean that syllables are presented at a high enough sound pressure level that discrimination performance is not affected by forward masking. Either possibility suggests that forward masking cannot sufficiently explain poor interval discrimination performance.

The final experiment in this chapter examines the role familiarity plays in birds' ability to discriminate changes to the fine structure of song syllables. Familiarity with an auditory stimulus, and more specifically a sequential auditory stimulus, allows one to form expectations about the next element in the sequence. It is possible that because syllables are produced so precisely via the syrinx and surrounding muscles, that there are general principles that syllables follow. Reversals of the fine structure in syllables may result in syllables are physically impossible to produce, and thus do not occur in nature. Temporal reversals in an unfamiliar stimulus in which there are no general principles or expectations may be more difficult to discriminate. To test this, birds were tested on temporal reversals of syllables that occur in song motifs that are played backwards.

3.6 Experiment 5: Discrimination of syllable reversals within a time reversed song motif

3.6.1 Introduction

The goal of this experiment is to test whether birds can discriminate single syllable reversals in a stimulus that contains the same spectral content, but no longer resembles song. Thus, birds can no longer rely on expectations of what sounds "correct" in the song. Time reversed song can easily be discriminated from forward song (Braaten et al. 2006), and has been shown to elicit no neural responses in the avian forebrain, even when the reversed song is the BOS (Solis and Doupe 1997; Theunissen and Doupe 1998). In addition, time-reversed song contains upsweeps and other acoustic features that are not naturally produced by zebra finches. It is possible that in Experiment 2, birds are able to discriminate syllable reversals so easily because a song motif is a familiar acoustic environment. Thus, a reversed syllable within a motif is a foreign sound within this familiar environment, and also a sound that violates expectations of this environment. This experiment asks whether birds can discriminate temporal reversals equally well within a stimulus background that is less familiar.

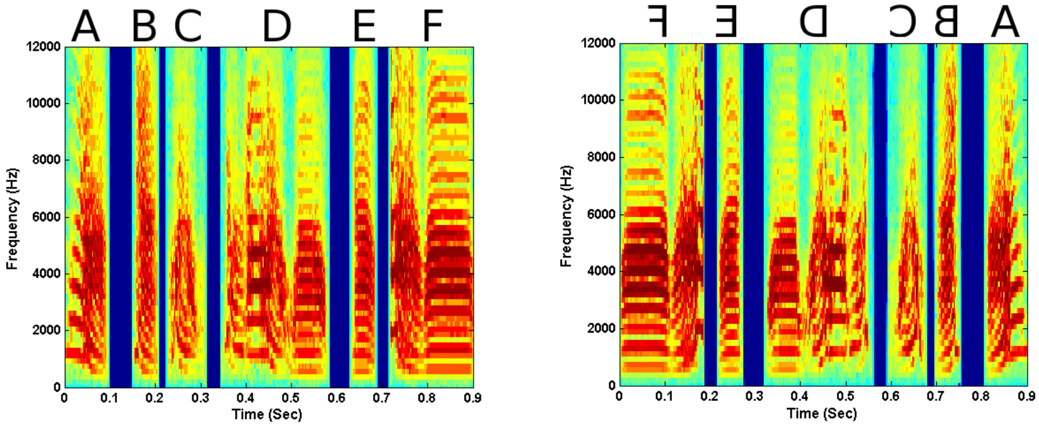
3.6.2 Methods

Subjects

Four male zebra finches were used in this experiment.

Stimuli

Stimuli consisted of the same four songs used in Experiment 2, except that songs were reversed in time such that both syllable order and individual syllables were reversed. This is equivalent to simply flipping the entire song in time. Targets consisted of single syllable reversals, and were the same targets as in Experiment 2 (see Table 3.2). Since the background motif is reversed, these targets consist of a single forward syllable within the reversed motif. An example of the background motif is illustrated in Figure 3.19(b), and a target is shown in Figure 3.20. The target sets for Moonshine's, Bear's, and Julep's songs each contained one target in which all syllables were reversed. Since the original background song is reversed in time, this target consists of all forward syllables played in reverse order (i.e. E,D,C,B,A).



(a) Forward song (b) Reversed song

Figure 3.19: Comparison of a natural song, and the same song reversed in time. Both global syllable order and individual syllable structure is reversed in time. Dark blue in the spectrogram indicates pure silence between the syllables.

Experimental Design

Each of the four male zebra finches was tested on the BOS, and the other three conspecific songs.

Training

Since birds were acclimated to song, no further training was necessary.

Testing Procedures

The background motifs was presented at 70 dB SPL, at a rate of once per 1500 msec, as in all previous experiments. After a variable amount of time (2-6 msec) a target was presented. After target presentation, birds had to peck the report key within 2.5 sec for food reward. Failure to peck within 2.5 sec resulted in a “miss”. For each 10-trial block, 7 targets and 3 sham trials were presented in random order. Birds were run on 300 trials for each stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate

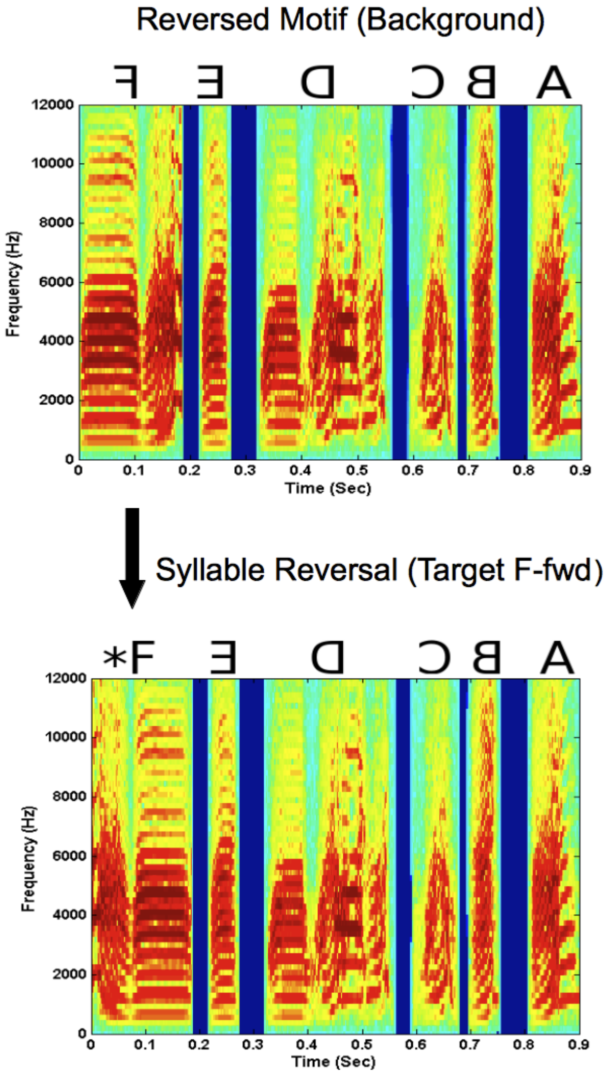


Figure 3.20: Reversed song motif (top) that serves as the repeating background, and the same motif with a single syllable reversed (bottom) that serves as a target. Note that the target contains a single forward syllable among a reversed song motif. Dark blue in the spectrogram indicates pure silence between the syllables.

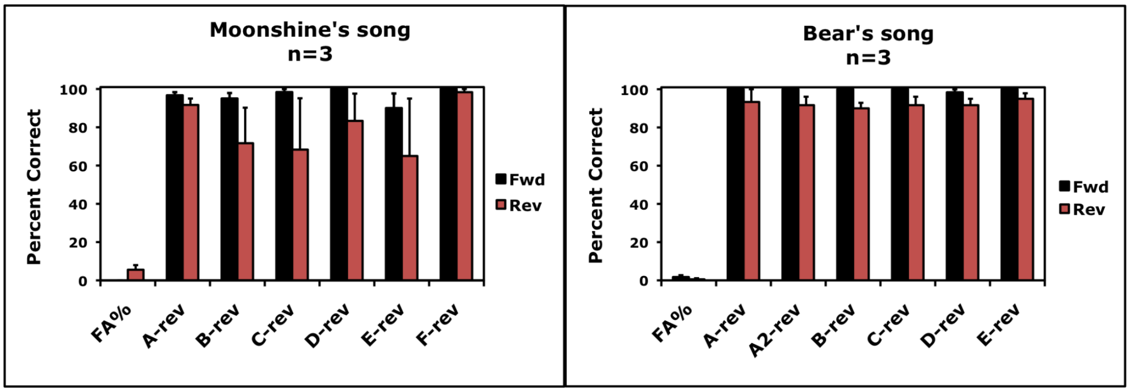
did not exceed 20%. One bird whose false alarm rate exceeded 20% for all songs was not included in the analysis.

3.6.3 Results

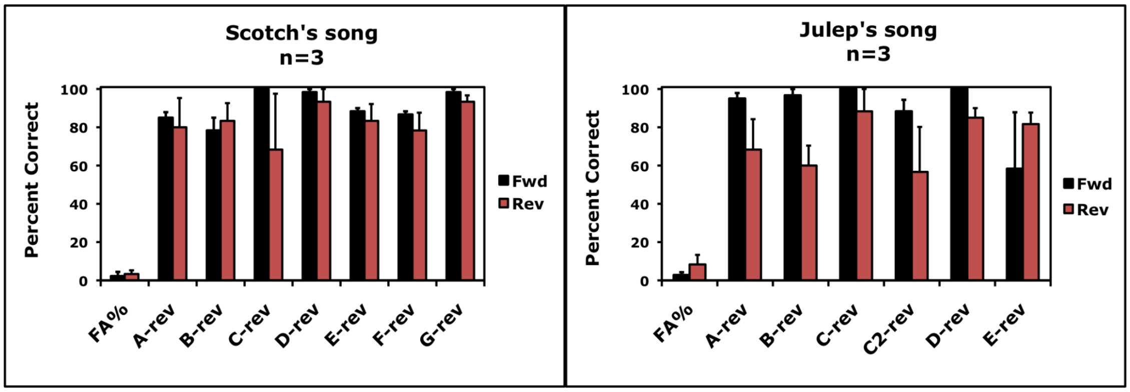
Overall, there was a significant difference in average discrimination performance between motifs played in the forward position ($94.07 \pm 13.22\%$), and in the reverse position (82.07±22.96%) (Mann-Whitney rank sum test, U=4041.000, $n_1 = n_2 = 75$, $p < 0.001$). These results are shown in Figure 3.21. Decline in average discrimination performance for individual syllables was not significantly correlated with syllable duration (Pearson correlation; r(23)=0.0465, p=0.825). There was also no effect of syllable type on decline in average performance (Kruskal-Wallis one-way ANOVA; H=1.906, df=4, p=0.753). Lastly, there was no effect of syllable position on decline in average performance (Kruskal-Wallis one-way ANOVA; H=8.161, df=4, p=0.086). In other words, decline in average discrimination performance for motifs reversed in time was not systematic, and could not be accounted for by syllable duration, syllable type, or syllable position within the motif.

3.6.4 Discussion

Birds were still able to discriminate single syllable reversals in a motif that is temporally reversed, but performance was significantly worse compared to natural song. There was no indication that certain syllable types, positions, or durations were more prone to worse performance when motifs were reversed in time. The only difference in the stimulus set from Experiment 2 was that the background and target motifs were temporally reversed. This resulted in a stimulus with the same spectrum as song, but was unfamiliar to birds. The discrimination task (i.e. temporal reversals) was the same as in Experiment 2, except that the reversals took place in



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 3.21: Average discrimination performance of male zebra finches on single syllable reversals in forward song (black) and reversed song (red). Error bars show standard error of the mean.

an unfamiliar context. Thus, it seems likely the proficiency with which birds can discriminate temporal reversals of single song syllables is in part because song is a familiar acoustic environment.

The structure and morphology of song syllables in oscine songbirds is under precise neuromuscular control. Thus, fine structure in syllables reflects the physical limitations of the production system. Song syllables are produced by the syrinx, usually during expirations of air that vibrate the medial tympaniform membranes. Studies of brown thrashers show that the phonology of syllables, specifically fundamental frequency, frequency modulation, and amplitude modulation, have been shown to be tightly controlled through activity in the surrounding syringeal muscles, and air flow through the syrinx (Goller and Suthers 1996). Similar mechanisms are seen in zebra finches and other songbirds as well. More specifically, fundamental frequency is largely determined by electrical activity in the ventral syringeal muscle. Increases in activity result in an increase in the fundamental frequency, whereas decreases in activity in this muscle result in a decrease in fundamental frequency. Oscillations in activity of the ventral syringeal muscle correspond with frequency modulation produced in song syllables. Amplitude modulation is regulated in several ways, through the use of airflow through the syrinx. These complicated, and well coordinated control mechanisms for song syllable production suggests that the song syllables of birds are produced quite precisely. When syllables are temporally reversed, the resulting fine structure corresponds to a syllable that is physically impossible to produce, and thus does not occur in nature. Perhaps birds are able to discriminate these reversals so easily, because reversals violate the general principles of syllable production. Since song is a familiar acoustic environment to both zebra finches and budgerigars (since they are housed with the finches), reversals may stand out as unnatural elements in this environment. If this is the case, birds may not need to compare background and target motifs to determine a difference in fine structure. They may simply need to compare adjacent syllables to determine if a violation of syllable structure has been made.

Familiarity also plays a similar role in auditory discrimination in humans, as seen in a study by Jacobsen et al. (2005). In a passive listening task, two familiar, nonlinguistic sounds (breaking dishes and the Microsoft Windows chime) were presented as familiar sounds, and the time reversed versions of these sounds were presented as unfamiliar sounds. One sound served as the standard, or background, and the other served as the deviant, or target. The mismatch negativity (MMN), which is an event-related brain potential, was used as a measure of deviance detection. Passive listeners showed a larger amplitude MMN for deviant sounds that were inserted into a sequence of familiar (forward) sounds, compared to when deviant sounds were inserted into a sequence of unfamiliar (time reversed) sounds. In other words, it was easier to detect a deviant sound when the background was made of a familiar context. Jacobsen et al. also found that MMN amplitudes were greater for familiar deviants than for unfamiliar deviants. In the current experiment, the task was to discriminate a single forward syllable within a backwards motif, which is a familiar sound to birds. Nonetheless, this task was more difficult for the birds than discriminating reversals in a forward motif. One reason for this could be that birds are less able to hold the previous motif in memory while listening to the next motif to determine if a change is present.

This effect has also been seen in active listening tasks. Paquette and Peretz (1997) showed that subjects are more accurate and much faster at discriminating between musical instruments when the sounds are played in the forward direction, compared to when they are played in reverse. Subjects were presented a pair of instruments dichotically, either played forward or reversed, and were asked to determine whether the pair contained a violin. Thus, the task required discriminating a specific target from the rest of the sounds presented. This was much easier for sounds played forward than in reverse, even though subjects had been familiarized with the reversed sounds. Familiarity with the sounds provided a large advantage to the listener.

The experiments in this chapter have shown that zebra finches do not attend to interval changes in song, are very attentive to syllable changes, and that male zebra finches are quite sensitive to the duration of individual syllables in making these discriminations. This suggests that changes to fine structure within syllables are much more salient than changes to the temporal envelope of the motif. However, this still leaves the question of which features are being attended to in the fine structure, as spectral fine structure, temporal fine structure, and individual syllable envelopes are all features that are modulated over the duration of the entire syllable. The next two chapters address question, through the use of synthetic song motifs that isolate these acoustic features, but maintain the overall timing and rhythm of song. Chapter 4 focuses on the role of syllable envelope and to a lesser degree fine structure in reversal discriminability of song syllables. Chapter 5, focuses on perception of temporal fine structure alone, in the absence of spectral or envelope cues.

Chapter 4

The role of spectral and amplitude envelope cues in the perception of syllable fine structure

4.1 Introduction

In Chapter 3, results from Experiments 1 and 3 showed that zebra finches are not very sensitive to changes in the temporal envelope of song, as they could not discriminate interval duration doublings. The reason for this insensitivity to interval changes remains unclear. However, they are much more sensitive to changes that occur to the fine structure within individual syllables. Fine structure is an all-inclusive term referring to modulations that occur over time within the syllable. These modulations includes spectral fine structure, temporal fine structure, and amplitude envelope cues. Whether birds are using all of these cues equally in making these fine structure discriminations has yet to be determined. The following experiments ask what role spectral structure plays in fine structure discrimination. This is done by replacing spectral structure in song syllables with Gaussian noise and asking how this affects how birds discriminate fine structure changes to the song motif.

The stimuli in these experiments consist of a synthetic song in which the amplitude envelope of each natural song syllable was filled with random Gaussian noise generated either from the same seed (Experiments 6 and 9), or random seeds (Experiments 7 and 8). These noise syllables are separated in time by the same duration as the natural song intervals, resulting in a noise song with the same overall rhythm as natural song. Thus, this noise song has the same syllable envelopes, and temporal envelope cues as natural song, but lacks the spectral structure that natural song contains. Gaussian noise is broad band, and the spectrum remains fairly stable over time, unlike natural song syllables which contain spectral variation over time, and across different syllables. The fine structure in noise is presumed to contain mostly temporal fine structure for this reason. However, since the temporal fine structure is more randomized than in song, we assume that the main cue remaining in noise syllables is amplitude envelope. Experiment 9 tests perception of changes to noise fine structure in the absence of amplitude envelope cues.

4.2 Experiment 6: Discriminability of single burst reversals at all locations within a synthetic same-seed noise song

4.2.1 Introduction

This extension of Experiment 2 examines the ability of birds to discriminate syllable reversals in the absence of song-like spectral cues. Song normally contains complex, time-varying spectral structure, which is unique for each individual song syllable in the motif. By replacing this spectral structure with the same piece of Gaussian noise, we are able to determine the role syllable envelope cues (and to a lesser degree fine structure cues) play in syllable reversal discriminability.

4.2.2 Methods

Subjects

Four male zebra finches, three female zebra finches, and three female budgerigars were used in this experiment.

Stimuli

All noise songs were created in Matlab using the amplitude envelopes of the four natural songs used in Chapter 3. For each bird's song, the syllables were isolated using Adobe Audition, and each syllable envelope was extracted using a Hilbert transform. For each song, a unique seed was used to generate noise bursts so that all bursts had the same beginning noise, and differed at the ends based on the duration of the burst (which matched the duration of the individual song syllables). This was generated using the randn function in Matlab, and specifying a random number for the seed. Each song (MoonBOS, BearBOS, ScotchBOS, JulepBOS) used a different seed. Then, each burst was multiplied by it's corresponding syllable envelope to generate a noise syllable similar to the natural syllable. These noise syllables were then concatenated with silence that was the same duration as the natural song intervals, resulting in a noise song that has the same syllable envelopes, syllable durations, and interval durations as natural song. In addition, each song syllable was filled with noise such that the beginning of each syllable is the same, while the ends of each syllable contain different noise based on the syllable's duration. The longest syllable in each song had a unique piece of noise not present in any

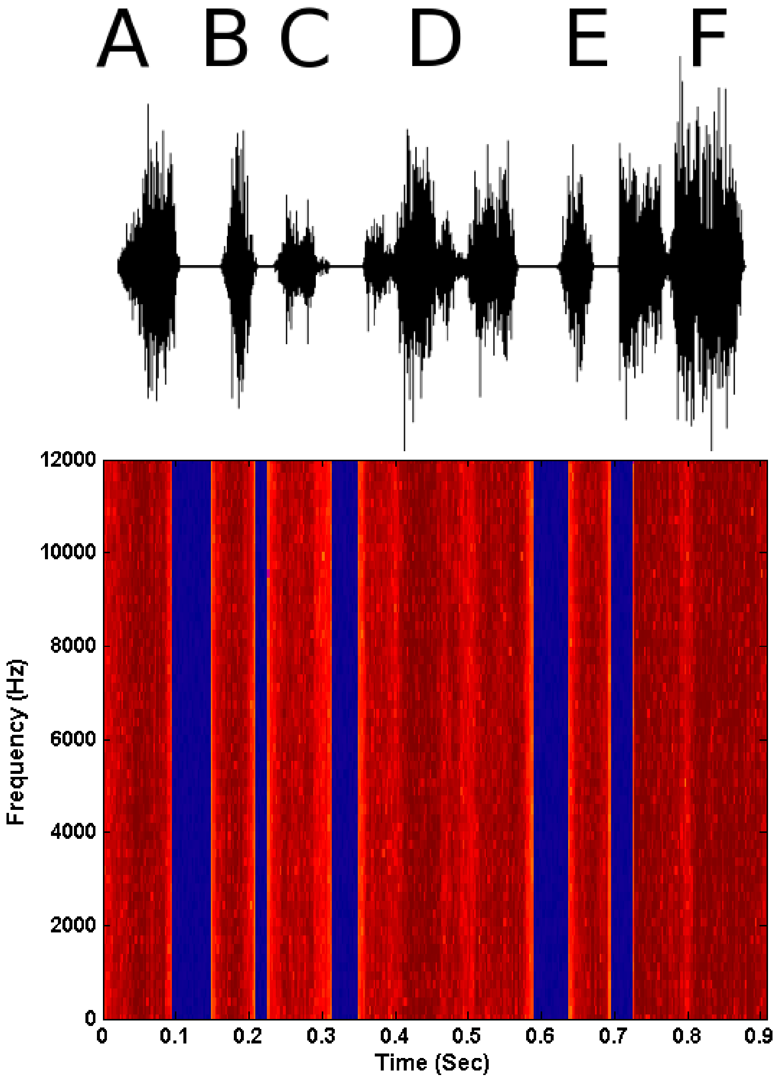


Figure 4.1: A synthetic noise song modeled after Moonshine's song motif. The top shows the amplitude envelope, and the bottom shows the spectrum. Unlike natural song, each syllable in noise song has a similar, flat spectrum that remains stable over time. Dark blue in the spectrogram indicates pure silence between the noise syllables.

other syllable, whereas the shortest syllable had a piece of noise that was present in every syllable. An example of Moonshine's noise song and its spectrum is shown in Figure 4.1.

For each stimulus set, each target contained a single noise syllable reversal, such

Table 4.1: Specific targets for Experiment 6: Same-seed noise song

Moonshine's song (A,B,C,D,E,F)	**Bear's song (A,B,C,D,E)**	**Scotch's song (A,B,C,D,E,F,G)**	**Julep's song (A,B,C,D,E)**
1. A reversed	1. A reversed	1. A reversed	1. A reversed
2. B reversed	2. B reversed	2. C reversed	2. B reversed
3. C reversed	3. C reversed	3. D reversed	3. C reversed
4. D reversed	4. D reversed	4. E reversed	4. D reversed
5. E reversed	5. E reversed	5. F reversed	5. E reversed
6. F reversed	6. All reversed	6. G reversed	6. All reversed
7. All reversed	7. All reversed	7. All reversed	7. All reversed

that all noise syllables at all locations were tested. For the three songs that contain less than 7 syllables, the remaining targets are motifs in which all noise syllables were reversed in time, but remained in the same sequential order (As in Experiment 2). For Scotch's noise song, which did contain 7 syllables, only 6 reversals were tested, and the 7th target was a motif in which all noise syllables were reversed (target "Burst rev"). Noise syllable B, which is a similar duration to noise syllable A (76 and 65 msec, respectively) was not tested.

Experimental Design

All birds were tested on the four song sets. Thus, each male was tested on the BOS, and the three other conspecific songs. The target set for each of the same-seed noise filled songs is listed in Table 4.1.

Training

To allow birds to be acclimated to noise songs, a random noise song was made from the training song used in Experiment 1. The noise song training set contained the same two types of targets that the natural training song had. One target consisted of the second interval quadrupled in duration, and the other target consisted of the entire noise song reversed in time. The interval target was presented on 30% of trials, and the reversal target was presented on 40% of trials. The remaining 30% of trials in the session were sham trials in which no target was alternated with the background. Birds ran on the training set for 2 sessions, or 200 trials before moving experimental stimulus sets.

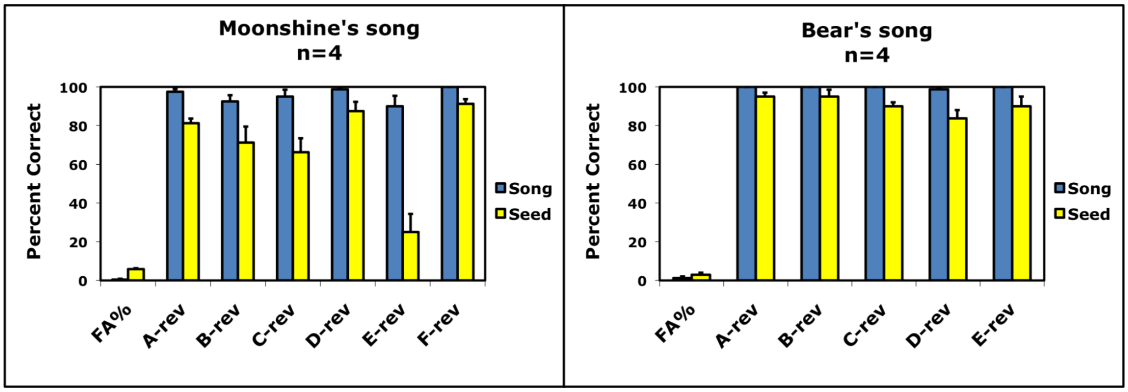
Testing Procedures

Same-seed noise songs were presented at 70 dB SPL, with a presentation rate of once per 1500 msec, and a response interval of 2500 msec. Thirty percent of trials were sham trials. For each 10-trial block, 7 targets and 3 sham trials were presented in random order. Birds were run until they were able to complete 300 trials in which the false alarm did not exceed 20%, and the last 200 trials were analyzed. For most birds, this required less than 600 trials.

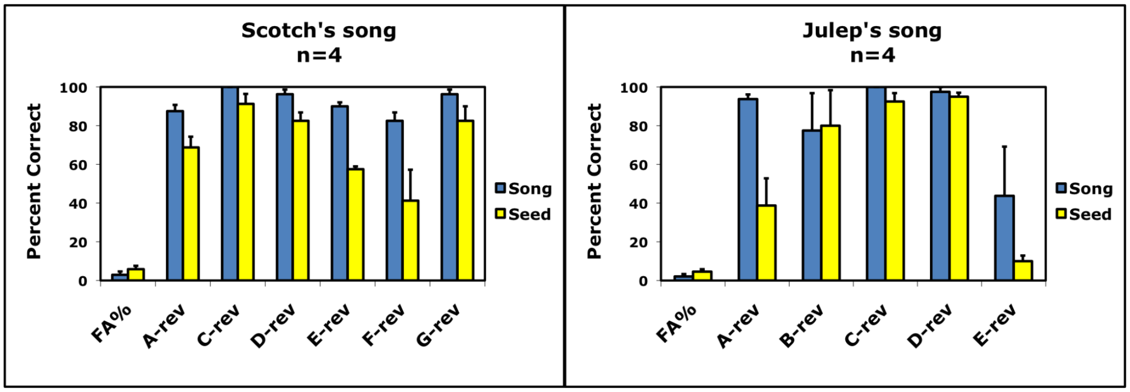
4.2.3 Results

A comparison of performance on single syllable reversals in natural song and same-seed noise song is shown in Figures 4.2, 4.3, and 4.4 for males, females, and budgerigars, respectively.

For all three groups, reversal discrimination performance was 70% or higher for the majority of the noise syllables (15/22 syllables for males, 17/22 for females



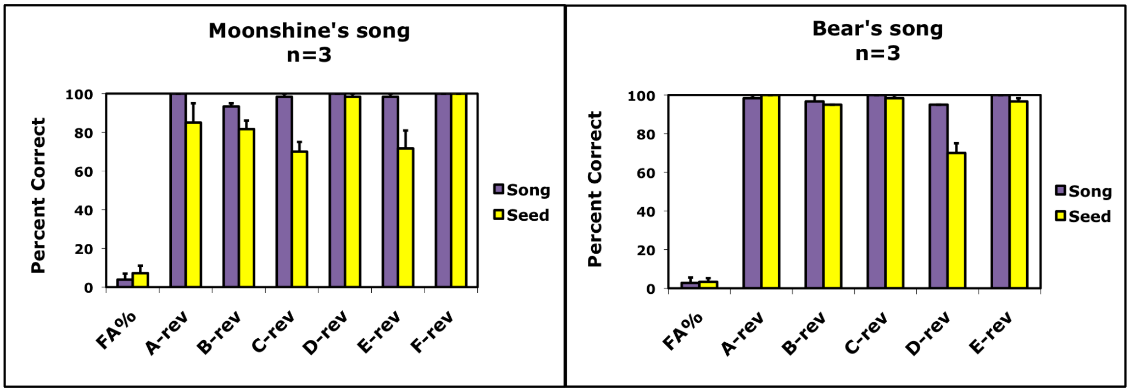
(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



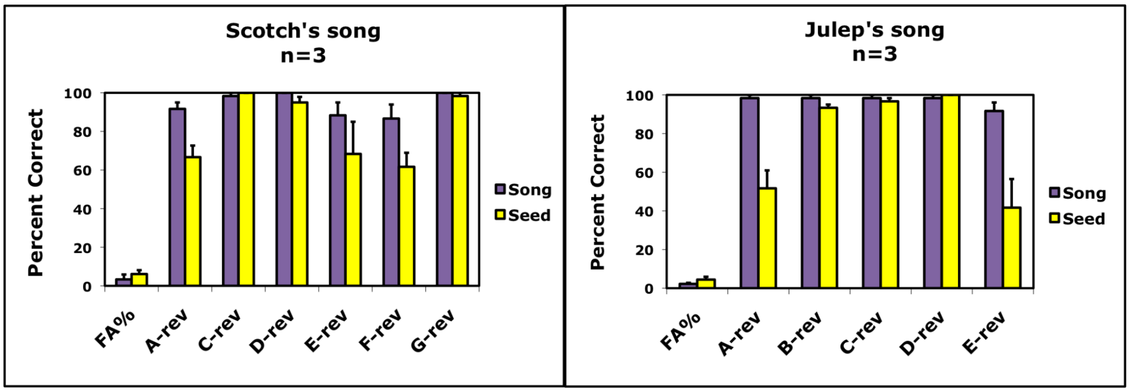
(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 4.2: Comparison of average discrimination performance of male zebra finches on single syllable reversals in natural song (blue), and same-seed noise song (yellow). Error bars show standard error of the mean.

and budgerigars), indicating that syllable envelope and fine structure do contribute to forward/reverse discriminability. If birds were using spectral structure of song syllables alone, we would expect much lower hit rates for same-seed noise songs. Compared with natural song, average performance across all syllables was significantly lower for same-seed noise songs (Rank sum test, $p < 0.05$, see specific values in Table 4.2). However, decline in average hit rate was not uniform across all syllables, and was in fact negatively correlated with syllable duration for all three groups (Pearson correlation, Males: r(20)=-0.534, p=0.010; Females: r(20)=-0.680, p=0.000; Budgerigars: r(20)=-0.593, p=0.004).



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

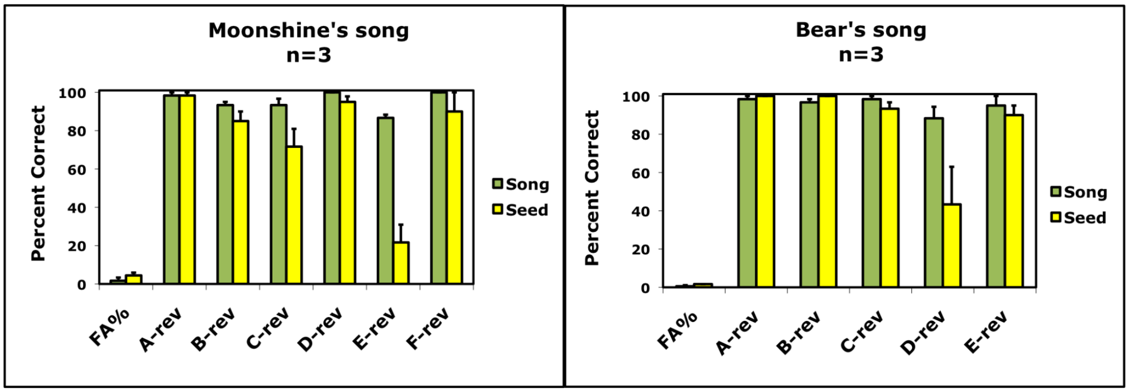


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

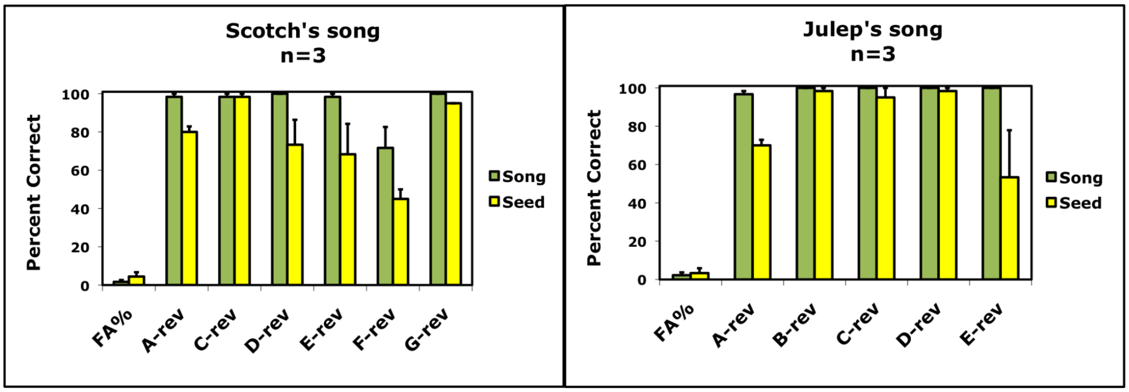
Figure 4.3: Comparison of average discrimination performance of female zebra finches on single syllable reversals in natural song (purple), and same-seed noise song (yellow). Error bars show standard error of the mean.

Table 4.2: Results of Mann-Whitney Rank Sum Test comparing performance in Experiments 2 and 6

Group	Median Natural	Median Seed	U	Sample Size	p
Males	100.00	85.00	1584.000	$n_1 = n_2 = 88$	< 0.001
Females	100.00	95.00	1245.000	$n_1 = n_2 = 66$	< 0.001
Bud	100.00	92.50	1212.000	$n_1 = n_2 = 66$	< 0.001



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 4.4: Comparison of average discrimination performance of budgerigars on single syllable reversals in natural song (green), and same-seed noise song (yellow). Error bars show standard error of the mean.

Similar to reversal discrimination of natural song syllables, only males showed a duration effect upon performance for same-seed noise syllables. For noise syllables less than 100 msec in duration, performance was positively correlated with duration for males (r(20)=0.628, p=0.022) but not for females (r(20)=0.418, p=0.155) or budgerigars (r(20)=0.233, p=0.443). For noise syllables greater than 100 msec in duration, all three groups showed a stabilization in performance at around 95%. A Kruskal-Wallis one-way ANOVA followed by post-hoc tests using Dunn's method showed that female zebra finches had significantly higher hit rates compared to male zebra finches for same-seed noise song (H=9.108, df=2, p=0.011, see Table 4.2 for median hit rates). There were no other significant differences between the groups.

4.2.4 Discussion

Most reversals of same-seed noise filled syllables had a hit rate of 70% or greater, which is well above the false alarm, or guessing rate. This suggests that while the spectral features of song syllables contribute to some of the forward/reverse discriminability, it does not account for all of it. If spectral patterning of individual syllables accounted for discriminability 100%, then we would expect a hit rate for forward/reverse discrimination of same-seed noise syllables to be much lower and closer to the false alarm rate. The fact that discrimination rates are this high means that for this task, birds are using the remaining features: syllable envelope and fine structure present in the noise. While the fine structure in noise differs from that of individual song syllables, it nonetheless remains a cue that birds may be using for this discrimination.

For all three groups, birds had significantly lower performance for same-seed noise song, compared with natural song. The decline in hit rate was negatively correlated with syllable duration, and this was also significant for all three groups. Thus, when spectral cues are removed from song, a decline in performance was seen for shorter syllables, whereas longer syllables were less affected. However, zebra finches were most sensitive to this, as they were the only group that showed a duration effect on performance for same-seed noise syllables shorter than 100 msec in duration. This is similar to results in Experiment 2, in which only male zebra finches showed a duration effect for reversals of natural song syllables shorter than 100 msec. Together, these results demonstrate that male zebra finches rely on temporal integration when listening to song and song-like stimuli more than females and budgerigars.

In this experiment, the amplitude envelopes of syllables were filled with noise generated from the same seed. Since syllables differ in duration, the longest syllable had unique noise, whereas the shortest syllable had the same piece of noise as all other syllables. Thus, the noise structure was shared among the syllables, and this pattern in the noise fine structure may provide an additional cue to birds. In addition to comparing the background and target motifs to determine if a change is present, another strategy birds could use is to compare the noise structure of the adjacent syllables in the motif. To test whether birds are using this fine structure cue, or whether they are relying on envelope cues, the next experiment replaces the same-seed noise with random noise. In this case, each syllable envelope is filled with a unique piece of random Gaussian noise. While the envelope remains the same, the fine structure of noise is now very different for each syllable. Performance on this set of syllables will provide insight into whether birds are attending more to envelope, or fine structure when listening to noise song. If birds are listening more to fine structure cues, then we expect performance to be worse for the random noise case. This is because with random noise there is no repeating patterns among the syllables. However, with same-seed noise the noise structure is the same for the beginning of each syllable. Thus, the noise structure should be more difficult to follow in the random noise case. If performance remains the same for both types of noise songs, then it is likely that birds are listening more to the envelope, which is the same in both cases.

4.3 Experiment 7: Discriminability of single burst reversals at all locations within a synthetic random noise song

4.3.1 Introduction

This experiment examines whether filling the envelope of song with random noise affects forward/reverse discriminability of individual syllables. The results of this experiment will be compared with performance in Experiment 6, in which the song envelope was filled with the same-seed noise for each syllable. In both cases, the song envelope remains the same and it is only the fine structure of the noise that differs. In the random seed case, the noise is unique for each syllable. In the same-seed case, the noise is shared between syllables. Whether or not performance differs from the same-seed noise case will provide insight into the degree to which birds are listening to the fine structure of the noise, versus the amplitude envelope of the noise.

4.3.2 Methods

Subjects

Four male zebra finches, three female zebra finches, and three female budgerigars were used in this experiment.

Stimuli

All random noise songs are created in Matlab using the amplitude envelopes of the four natural songs used in Chapter 3. For each song, the syllables were isolated using Adobe Audition, and each syllable envelope was extracted using a Hilbert transform. Then each syllable envelope was multiplied point by point by random Gaussian noise (generated from the randn function in Matlab without specifying the seed) that was the same duration as the envelope. This resulted in a random noise burst that had the same amplitude envelope as an individual song syllable. These noise syllables were then concatenated with silence that was the same duration as the natural song intervals, resulting in a random noise song that had the same syllable envelopes, syllable durations, and interval durations as natural song. While each syllable was filled with a different piece of random noise, the same noise-filled song serves as the background and target set throughout the entire experiment. Thus, the random noise was frozen and unchanging.

Experimental Design

The same experimental design was used as in Experiment 6, with the same targets presented for each stimulus set. The only difference is that the syllable amplitude envelopes for each song were each filled with a different piece of frozen random noise, rather than the same noise for each individual syllable. Similar to the previous experiment, for each song (with the except of Scotch's song; see Experiment 6), single noise syllable reversals were tested at all locations within the song.

Table 4.3: Results of Mann-Whitney Rank Sum Test comparing performance on Experiments 6 and 7

Group	**Median Seed**	**Median Rand**	**U**	**Sample Size**	**p**
Males	85.00	85.00	3966.500	$n_1 = n_2 = 88$	0.780 (n.s.)
Females	95.00	85.00	2462.000	$n_1 = n_2 = 66$	0.190 (n.s.)
Bud	92.50	90.00	2456.500	$n_1 = n_2 = 66$	0.200 (n.s.)

Training

Since birds were already acclamated to noise songs from Experiment 6, no further training was necessary.

Testing Procedures

Random noise songs were presented at 70 dB SPL, with a presentation rate of once per 1500 msec, and a response interval of 2500 msec. Thirty percent of trials were sham trials. For each 10-trial block, 7 targets (see Table 4.1) and 3 sham trials were presented in random order. Birds were run until they were able to complete 300 trials in which the false alarm did not exceed 20%, and the last 200 trials were analyzed. For most birds, this required less than 600 trials.

4.3.3 Results

There were no significant differences in overall performance between random noise and same-seed noise songs for males, females, or budgerigars (Rank Sum Test, $p < 0.05$, for specific values, see Table 4.3.

This suggests that changing the noise within the song envelope had little effect upon performance, and that syllable envelope is the main cue birds are attending to when listening to noise songs. To test this hypothesis, the rise and fall rates of syllable envelopes were calculated to determine if performance was correlated with forward/reverse symmetry of syllable envelopes. The rise/fall rates of the first and last 5 and 10 msec were calculated and the absolute value of the difference was taken as a measure of envelope asymmetry. A larger difference between the rise and fall rate indicates more forward/reverse asymmetry. If birds are listening to syllable envelope cues, then we expect syllables with larger asymmetry should have higher hit rates, whereas syllables with less asymmetry should have lower hit rates.

For all three groups, rise/fall difference was positively correlated with performance. These results are summarized in Table 4.4. Male and female zebra finch performance was significantly correlated with rise/fall asymmetry of the first and last 10 msec of the syllable for random noise song, but not for natural or same-seed noise song. A similar pattern was seen with budgerigars, but for the rise/fall asymmetry of the first and last 5 msec of the syllable. This demonstrates that when the task becomes more complicated by the addition of unique noise for each syllable, birds rely even more on syllable envelope cues to make forward/reverse discriminations.

There were no significant differences in performance on random noise song between males, females, or budgerigars (Kruskal-Wallis, H=2.739, df=2, p=0.254). This differs from same-seed song, in which females performed better than males. Performance on random noise song showed several similarities to same-seed noise song. First, performance on random seed noise song was significantly lower than natural song (Rank Sum Test, $p < 0.05$, specific values in Table A.3 in Appendix A). Second, the decline in average hit rate for random noise syllable reversals was negatively correlated with syllable duration (Males: $r(20) = -0.619, p = 0.00212$;

Table 4.4: Relationship between envelope rise/fall asymmetry and performance on Experiments 2, 6, and 7

Group	**Song Type**	**Pearson**	**df**	**p**
Males	Natural	0.189	20	0.400 (n.s.)
	Seed Noise	0.388	20	0.074 (n.s.)
	Random Noise	0.432	20	0.044
Females	Natural	0.279	20	0.209 (n.s.)
	Seed Noise	0.385	20	0.077 (n.s.)
	Random Noise	0.472	20	0.027
Bud	Natural	0.066	20	0.772 (n.s.)
	Seed Noise	0.349	20	0.111 (n.s.)
	Random Noise	0.436	20	0.043

Females: $r(20) = -0.583, p = 0.00443$; Budgerigars: $r(20) = -0.507, p = 0.0161$). Lastly, male zebra finch performance on random noise syllables showed a duration effect for syllables less than 100 msec (Pearson correlation; $r(20)) = 0.739, p = 0.00389$), whereas female zebra finch ($r(20) = 0.43, p = 0.139$) and budgerigar ($r(20) = 0.194, p = 0.524$) performance did not.

4.3.4 Discussion

The results of this experiment suggest that when spectral cues are removed from song, birds rely mainly on amplitude envelope cues in syllables for forward/reverse discriminability of syllables. The transition from same-seed noise in Experiment 6 to random noise in Experiment 7 did not affect overall performance. If birds were relying on mostly fine structure cues present in the noise, we would expect to see a decline in discrimination performance. Instead, performance was similar in both experiments, but was significantly correlated with syllable envelope rise/fall asymmetry in the random noise song case. Performance was not correlated with envelope asymmetry for natural song or same-seed noise song. Perhaps with these song types, other acoustic cues were available in the syllables such that amplitude envelope served as a supplementary cue. However, once the fine structure noise song became more complex, as in random noise song, amplitude envelope became an important cue.

The noise notifs used in Experiments 6 and 7 were made in such a way that noise syllable reversals contained both a reversal of the noise burst itself (i.e. the fine structure) and a reversal of the amplitude envelope. Even though, both envelope and fine structure were reversed, we inferred from the results that birds were mainly using envelope cues since performance did not change when we removed the repeating patterns in the noise syllables and replaced it with random noise. However, two additional tests as an extension of Experiment 7 may provide further evidence of this by separating out envelope and fine structure cues in the syllable reversal. Envelope cues can be isolated in noise motifs by creating the reversed noise syllables such that forward noise bursts are multiplied by a time reversed syllable envelope. This reversed syllable will then only change in the envelope, whereas the fine structure remains the same as in the forward syllable. Fine structure cues can be isolated by creating reversed noise syllables such that the time reversed noise bursts are multiplied by forward syllable envelopes. In this case, the reversed syllable will only change in the fine structure of the noise, and the envelope will be the same as in the forward syllable.

Performance on both of these types of syllable reversals can be compared with the original experiment in which both envelope and fine structure are reversed. If the primary use of envelope cues holds, we predict that birds will perform much better on syllable reversals in which the only the envelope is reversed than syllable reversals in which only the fine structure is reversed. While envelope changes may be the main cue birds use for this discrimination, performance on syllable reversals in which only fine structure is reversed will tell us the degree to which birds use fine structure as a cue. If performance does not exceed the false alarm rate, then we can conclude that birds only use envelope cues in noise syllable reversal discriminations. However, if birds have low hit rates that exceed the false alarm rate, we can conclude that while envelope is the main cue, birds do use fine structure changes to some extent.

The use of envelope cues in syllable discrimination in birds is paralleled by the use of envelope cues in speech recognition in humans. A study by Drullman (1994) tested speech intelligibility in humans after manipulating temporal envelope and fine structure cues in speech. One experimental manipulation involved filling the speech envelope of a sentence with random noise and testing subjects' ability to repeat the sentence they heard. On average, subjects were able to repeat back 98.30% of sentences, which is near perfect. When listening to noise speech, subjects were able to use envelope cues alone to understand sentences. In other words, when the envelope of speech remained intact, removal of fine structure cues had minimal effect on speech intelligibility. Conversely, when Drullman et al. kept the fine structure of speech intact, but replaced the speech envelope with a random envelope, the resulting speech was much less intelligible, and subjects could only repeat about 17% of sentences back. From these results, and the results of our experiments, amplitude envelope cues have been shown to be important in perception of both song and speech, and specifically play a role in understanding speech.

Fine structure (both spectral and temporal) is also important in speech perception but provides different information from envelope cues, specifically pitch perception and understanding speech when background noise is fluctuating. Envelope cues alone are not sufficient for these aspects of speech perception. Pitch perception is especially important in tonal languages where pitch shifts change the meaning of words. A study by Kong and Zeng (2006) found that in quiet, subjects could discriminate the tone of various syllables perfectly using only fine structure cues, but could only discriminate the tone of $70 - 80\%$ of syllables using only envelope cues. In noise, subjects had even more difficulty discriminating tone when only envelope cues were present whereas discrimination of tone using only fine structure cues was less affected. It has been suggested by Moore (2008) that fluctuations of temporal fine structure in both speech and nonspeech stimuli allow for listeners to "listen within the dips" of a fluctuating background in order to detect signals.

Perception of changes to noise structure will be further examined in Experiment 9, which tests birds' ability for fine structure discrimination when syllable amplitude envelope cues are absent. Multiple populations of noise are used to test the robustness with which zebra finches can make these discriminations. First, one final experiment using the synthetic random noise song will be presented. The following experiment utilizes the synthetic random noise song in order to test perception of interval changes when spectral structure is removed from song. Perhaps birds are unable to attend to the global timing of song, because they are attending to it's spectral structure within individual syllables. Since song has multiple cues, it is possible that birds are listening to the spectral content, but not overall rhythm of song. As in the previous experiment, spectral structure is removed and replaced with random noise in an attempt to draw attention to the overall timing and rhythm of song. If this results in an improvement in interval discrimination performance, this suggests that poor performance in Experiment 1 is a result of attentional constraints.

4.4 Experiment 8: Relative salience of envelope and fine structure cues in a synthetic random noise song

4.4.1 Introduction

The results of Experiment 1, showed that birds are much more sensitive changes in syllables (i.e. single syllable reversals) than they are to changes in the inter-syllable intervals (i.e. interval duration doublings). When presented both types of stimuli in the same experiment, they discriminate syllable reversals nearly 100% of the time, whereas they are unable to discriminate large changes to single intervals in the song motif. This leaves the question as to why syllables appear to be so much more salient than intervals, especially when intervals provide global timing and rhythm to song. One possibility is birds are uncertain of "where" to listen in the song (syllables vs. intervals). Because there are multiple cues in song, birds may direct their attention to the most salient features, ignoring all others. Perhaps the spectral structure in syllables is so salient to birds that it interferes with perception of temporal envelope cues in song. Previous auditory discrmination studies have shown that directing the listener's attention to the portion of the stimulus that contains the change improves discrimination performance (Leek et al. 1991; Leek and Watson 1984).

This experiment tests whether interference from the spectral structure of syllables contributes to the difficulty birds have in discriminating changes to intervals. The assumption is that spectral structure in individual syllables competes with temporal envelope cues in song for the bird's attention. Spectral structure wins out, and birds attend to syllables and ignore the intervals separating them. To test this, spectral structure in individual syllables is replaced by random noise, drawing attention to temporal envelope cues instead. If birds perform better on single interval doublings in a random noise song compared to natural song, this suggests that in natural song, there is interference present from the multiple acoustic cues such that birds are unable to attend to the global timing cues of song when the more salient spectral cues are also present. However, if there is no improvement in discrimination performance with random noise song, then this suggests a more general mechanism in which intervals between any type of sound stimulus may not be salient to birds.

4.4.2 Methods

Subjects

Four male zebra finches and one female zebra finch were used in this experiment.

Stimuli

Random noise songs were the same as those used in Experiment 7.

Experimental Design

This experiment was an extension of Experiment 1, and served as a control experiment. For this reason, male zebra finches (as well as 1 female zebra finch) were tested, and a modified stimulus set (3 out of 4 songs) was used. Julep's song was chosen to be eliminated from the stimulus set, since it contained the shortest syllable, and was the most difficult target set. The target set for this experiment was the same as that of Experiment 1 in Chapter 3. For each random noise song, the same intervals were doubled, and the same single syllables were reversed. The only difference was that the changes took place in random noise song, rather than the natural song. These targets are illustrated in Table 4.5.

Table 4.5: Specific targets for Experiment 8: Random noise song

:

Moonshine's song (A,B,C,D,E,F)	**Bear's song (A,B,C,D,E)**	**Scotch's song (A,B,C,D,E,F,G)**
1. Interval 1	1. Interval 1	1. Interval 1
2. Interval 3	2. Interval 2	2. Interval 4
3. Interval 5	3. Interval 4	3. Interval 6
4. A reversed	4. A reversed	4. A reversed
5. C reversed	5. B reversed	5. C reversed
6. E reversed	6. D reversed	6. E reversed
7. F reversed	7. E reversed	7. G reversed

Training

Since birds were acclimated to noise songs, no further training was necessary.

Testing Procedures

Random noise songs were presented at 70 dB SPL, with a presentation rate of once per 1500 msec, and a response interval of 2500 msec. Thirty percent of trials were sham trials. For each 10-trial block, 7 targets (see Table 4.5) and 3 sham trials were presented in random order. Birds were run on 300 trials, and the last 200 trials were used for analysis.

False alarm rates for this experiment were considerably higher and more variable than for Experiment 1. In order to better compare performance for these two experiments, hit rates and false alarm rates were converted to d-prime scores. D-prime measures an observer's sensitivity, or the observer's ability to discriminate between two stimuli (in this case the background and target motifs). D-prime not only takes into account the hit rate, but also the false alarm rate. The d-prime transformation is as follows:

$$d' = \text{z(hit rate)- z(false alarm rate)}, \tag{4.1}$$

where z is the conversion of the hit rate or false alarm rate into a z-score, or a unit of standard deviation. A hit rate or false alarm rate greater than 50% (chance) is converted into a positive z-score, whereas a hit rate or false alarm rate less than 50% is converted into a negative z-score. Maximum d' occurs when an observer is able to discriminate between two stimuli with a hit rate of 100% and a false alarm rate of 0%. To avoid the problem of infinite values, any scores of 100% were converted to $1/(2N)$, and any scores of 0.00% were converted to $1 - 1/(2N)$, where N is the number of trials used to calculate that score.

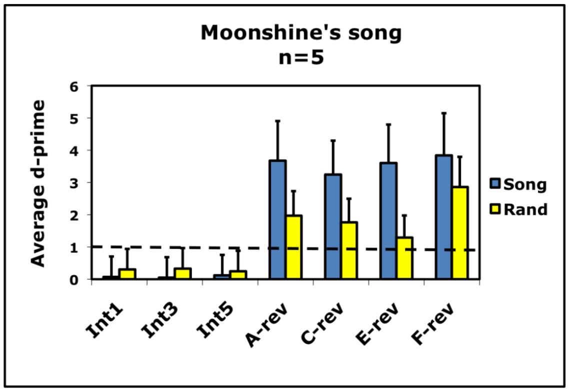
By using the d-prime measure, we were able to examine the discriminability of each target from the background, taking into account the higher false alarm rates. For each song tested, each bird's hit rate and false alarm rate were converted to d-prime scores, and the d-prime scores were averaged. Absolute values are reported, since any negative values were very small (> -1) and are likely due to chance variability (Macmillan and Creelman 2005). Since original hit rates for the female zebra finch did not differ significantly from the males for any of the songs tested (Rank sum test, $p > 0.05$), all five birds were analyzed together.

4.4.3 Results

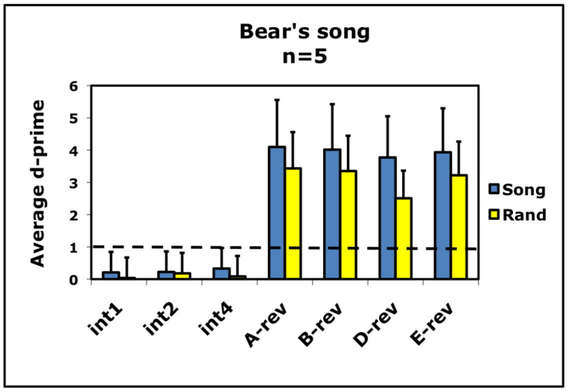
Overall, the birds' performance on random noise songs showed a similar pattern as the results of natural songs in Experiment 1. In other words, birds were not able to discriminate changes made to noise song intervals, but were able to easily discriminate changes made to most noise syllables. Since random noise song contains less acoustic cues than natural song, this task was more difficult. This resulted in different false alarm rates for the two experiments, with a higher rate for random noise song. To directly compare the performance in these two experiments, hit rates and false alarm rates for natural song (Experiment 1) and random noise song (Experiment 8) were converted to d-prime scores (see explanation of d-prime in the testing procedures). This comparison is shown in Figure 4.5.

By convention, a d-prime value of 1.00 is considered the threshold for discriminability, or the point at which the Gaussian distributions for two stimuli begin to no longer overlap. Using this convention, interval changes to both natural song and random noise song did not reach the threshold for discriminability. However, reversals for all syllables in natural song and random noise song were above this threshold, and were easily discriminated from the background. In addition, d-prime scores for interval doublings and syllable reversals in song did not have overlapping 95% confidence intervals, indicating that performance on syllable reversals is significantly higher than for interval doublings. This same result was maintained for noise song, with the exception of two syllables, Moonshine's syllable E and Scotch's syllable A. The d-prime scores for these two syllables did have overlapping 95% confidence intervals with the d-prime scores for interval doublings. However, this is due to a decrease in discriminability of syllable reversals in noise song, and not an increase in discriminability in interval doublings. D-prime scores and standard error values are shown in Table A.4 in Appendix A.

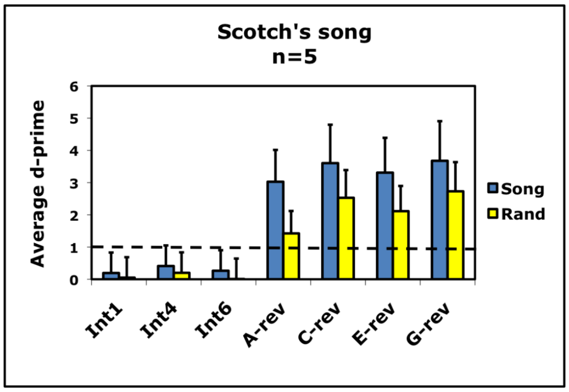
If poor performance on changes to intervals in natural song was due to attentional constraints, we would expect the sensitivity, or d-prime, for interval increases in random noise song to cross the threshold and have values greater than 1.00. However, this was not the case. Instead, d-prime scores for interval increases in noise song remained below threshold and were not discriminable from background. This provides evidence that poor performance on changes to intervals is not likely due to interference from the spectral structure in individual syllables.



(a) Discrimination performance: Moon's motif



(b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif

Figure 4.5: Average discrimination performance of male zebra finches on interval doublings and syllable reversals in natural song (blue) and random noise song (yellow). Performance is shown as average d-prime scores. The dashed line is at d-prime=1.00. Error bars show 95% confidence intervals.

4.4.4 Discussion

The goal of this experiment was to determine whether the unique and strong spectral patterning of individual syllables competes with temporal envelope cues in song for attentional resources. In other words, birds are unable to attend to the overall rhythm of song because attention is drawn to the spectral structure of the syllables. This could explain poor performance on discriminating interval changes, even when only interval changes are presented. Synthetic songs were made in which each syllable's amplitude envelope was filled with a different piece of frozen randomly generated noise. Thus, each syllable was always represented by the same piece of random noise, and each syllable had a unique piece of random noise. This resulted in a song-like stimulus that was consistent from rendition to rendition (like natural song is), but that also lacked any systematic spectral structure that varied over time.

Direct comparison of performance on interval changes in natural song and random noise song showed no change in discriminability. There was no improvement in the random noise song case. Thus, competition from spectral structure of syllables is not likely the cause of poor performance on interval changes in natural song. If there had been an improvement in interval discrimination in the random noise song case, this would indicate that birds are capable of attending to intervals between sound stimuli, but do not normally do so when listening to natural song. As this did not happen, another possibility is that birds do not attend to intervals between any type of sound stimuli. Experiments 1 and 8 test *gap duration discrimination*, in which the only change that occurs in the stimulus is a change to the gap between two sounds. Further gap discrimination experiments with simple stimuli such as pure tones should be done to measure the ΔT necessary to discriminate a change in the gap between two sounds. Thus far, previous research involving gaps has only tested zebra finches and budgerigars on *gap detection*, which measures the amount of silence needed between two sounds in order to perceive them as two separate sounds instead of one sound (Okanoya and Dooling 1990). With gap detection tasks, it is possible to complete the task while only attending to the sound stimuli to determine if one or two sounds are heard. Listening to the gaps is not necessary.

From the results of Experiment 1, we see that human subjects attend to intervals between song syllables more than zebra finches and budgerigars do. Human subjects were able to reliably discriminate increases in duration to the first and middle intervals in all songs tested. In addition, humans have been shown to attend to intervals separating non-song and non-speech stimuli. Abel (1972) tested human subjects on discrimination of temporal gaps between two Gaussian noise bursts, and found that ΔT not only depends on the original duration of the gap, but also the duration and amplitude of the marker sounds surrounding the gap. Thus, for both song, and non-song stimuli, humans are able to attend to gaps between sounds more effectively than birds. Perhaps birds and humans listen to and perceive song and sounds in general in different ways. Both humans and birds have demonstrated auditory stream segregation, or perceptual grouping, of tones based on frequency differences (Bregman 1990; MacDougall-Shackleton et al. 1998). Auditory segregation studies in humans have shown that gap discrimination within a sequence of tones is more accurate when the bordering tones are similar frequencies, compared to when the tones differ greatly in frequency (Kinney 1961). This held true for several patterns of tones that were tested. This is because tones of different frequencies are heard as separate entities, whereas tones of similar frequencies are grouped together. This result predicts that birds should have an easier time of gap discrimination in noise songs, because the spectrum of each syllable is broadband

and relatively unchanging. Thus, noise syllables should be perceptually grouped together. This is not the case for birds, as replacing syllable structure with noise did not increase discriminability of interval changes. This suggests that there may be fundamental differences in the way humans and birds perceive gaps and intervals between sounds.

4.5 Experiment 9: Forward/reverse discrimination using several different noise populations

4.5.1 Introduction

Zebra finches are able to discriminate reversals of single syllable envelopes filled with random or same-seed generated noise, as shown from Experiments 6 and 7. In this model of song, spectral cues are largely reduced since the noise is fairly uniform across all frequencies. However, in these experiments, both syllable envelope and fine structure cues were present. The purpose of this next experiment is to ask whether birds are still able to make these discriminations in the absence of syllable envelope as a cue, and test whether this discrimination can be made regardless of the noise that is used. By testing different populations (i.e., seeds) of noise, we are testing the robustness with which birds can follow the temporal fine structure of random noise.

Because there are fewer cues present, this task should be more difficult than previous ones. In taking this into consideration, a simpler model of song was used for this experiment. The stimuli used were a triplet of noise bursts that were 110 msec in duration, separated by 40 msec of silence. These values were chosen based on the parameter space of natural zebra finch song. The shortest songs contain three syllables (Sossinka and Böhner 1980), the average duration of a syllable is ~ 110 msec (Glaze and Troyer 2006), and the average duration of an inter-syllable interval is ~ 40 msec (Goller and Daley 2001; Wild et al. 1998). In addition, these values are further confirmed by the four songs used in the experiments in Chapter 3. Of the 23 syllables and 19 intervals in the data set, the average syllable duration is 108.63 ± 55.30 msec, and the average interval duration is 36.08 ± 10.53 msec. Thus, the stimuli still contained an overall timing that was behaviorally relevant, even if number and individual duration of syllables were simplified from the songs used in the previous experiments.

4.5.2 Methods

Subjects

Four male zebra finches and one female zebra finch were used in this experiment.

Stimuli

All noise bursts were generated in Matlab using the randn function, and specifying the seed. The following seeds were randomly chosen: 0, 14, 18, 27, 71, 223, 500, 850. None of the seeds used in this experiment overlapped with the seeds used in Experiment 6. Thus, birds were naive to the temporal structure of the bursts. Noise bursts were 110 msec in duration. Reversed versions of the bursts were created in Matlab by flipping the values of the forward burst vector. Background and target stimuli were made by concatenating three bursts of the same seed, with 40 msec of silence between bursts. An example of this type of stimulus is seen in Figure 4.6. Birds are tested on discrimination of single burst reversals within the triplet, similar to the testing of single syllable reversals in previous experiments.

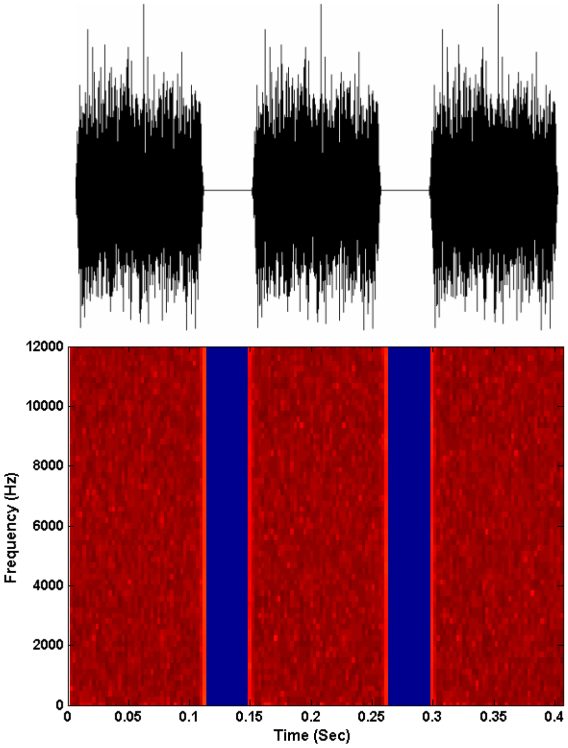


Figure 4.6: A triplet of identical noise bursts, all in the forward direction. The top shows the amplitude envelope, and the bottom shows the spectrum. Each burst in the triplet is 110 msec in duration, separated by 40 msec of silence (shown in the spectrogram in dark blue). In this example, each burst was generated using seed 850.

Experimental Design

Each seed was tested as a separate experiment, and the same order was presented to all birds. Within a given experiment, discrimination of single burst reversals was tested at all 3 locations within the triplet. In addition, an easier target in which all three bursts were reversed was also tested. This easier target ensured that the false alarm rate remained low, as this experiment is more difficult than previous tasks. The target set for each seed is shown in Table 4.6.

Table 4.6: Specific targets for Experiment 9: Triple burst seed test

Seed 0, 71, 850, 14, 18, 27, 500, 223
1. Burst 1 reversed
2. Burst 1 reversed
3. Burst 2 reversed
4. Burst 2 reversed
5. Burst 3 reversed
6. Burst 3 reversed
7. All bursts reversed

Training

As birds have already been acclimated to noise stimuli, there was no prior training for this experiment.

Testing Procedures

Triple burst stimuli were 450 msec in duration. They were presented at 70 dB SPL, at a rate of once per 1000 msec, with a response interval of 2000 msec. As in all previous experiments, 30% of trials were sham trials. Birds were run until performance had stabilized and they were able to run 200 continuous valid trials in which the false alarm rate did not exceed 20%. In most cases, birds ran between 200 and 300 trials in order to reach this criterion. In a few cases, birds ran more than this, never exceeding 600 trials. Since the female's average hit rates did not differ significantly from the males for any of the seeds tested, (Rank sum test, $p > 0.05$) all five birds were averaged together.

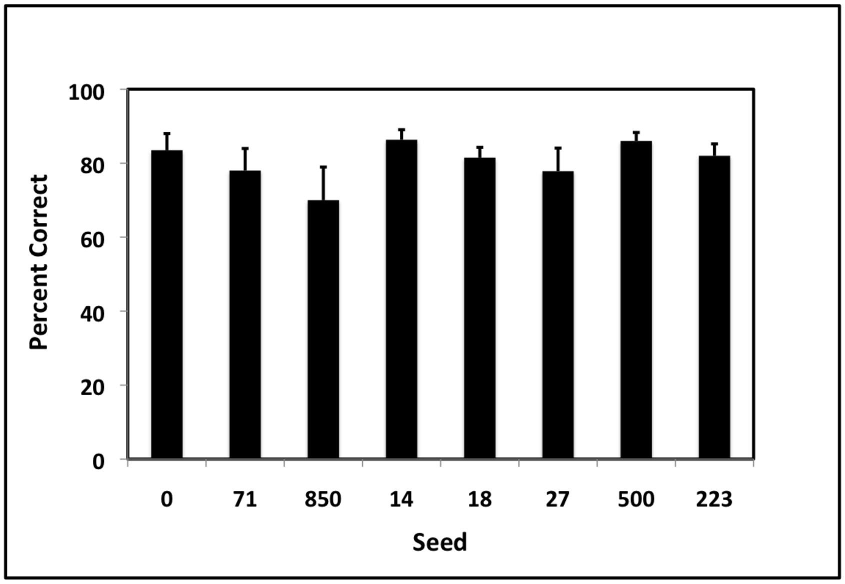


Figure 4.7: Average discrimination performance for single burst reversals for all eight seeds tested. Results are averaged across the three burst locations, and across the five birds tested. Error bars show standard error of the mean.

4.5.3 Results

On average, birds discriminated reversals of single bursts 70% of the time or greater, for all seeds tested (Figure 4.7). Across all seeds, there was no effect of position on performance (Kruskal-Wallis ANOVA, H=2.924, df=2, p=0.232). There was also no significant difference in average performance for any of the seeds tested (Kruskal-Wallis ANOVA, H= 10.599, df=7, p=0.157).

While average performance did not differ significantly for different populations of noise, there was some variability in individual performance (Figure 4.8). Some seeds had similar performance for all birds (18, 500), whereas others showed significant differences between birds (0, 71, 850, 14, 27, 223). These results are summarized in Table 4.7. Since individual performance within a seed passed normality and equal variance tests, a parametric one-way ANOVA was used. Individual performance for each seed was not significantly correlated with the number of trials ran (Pear-

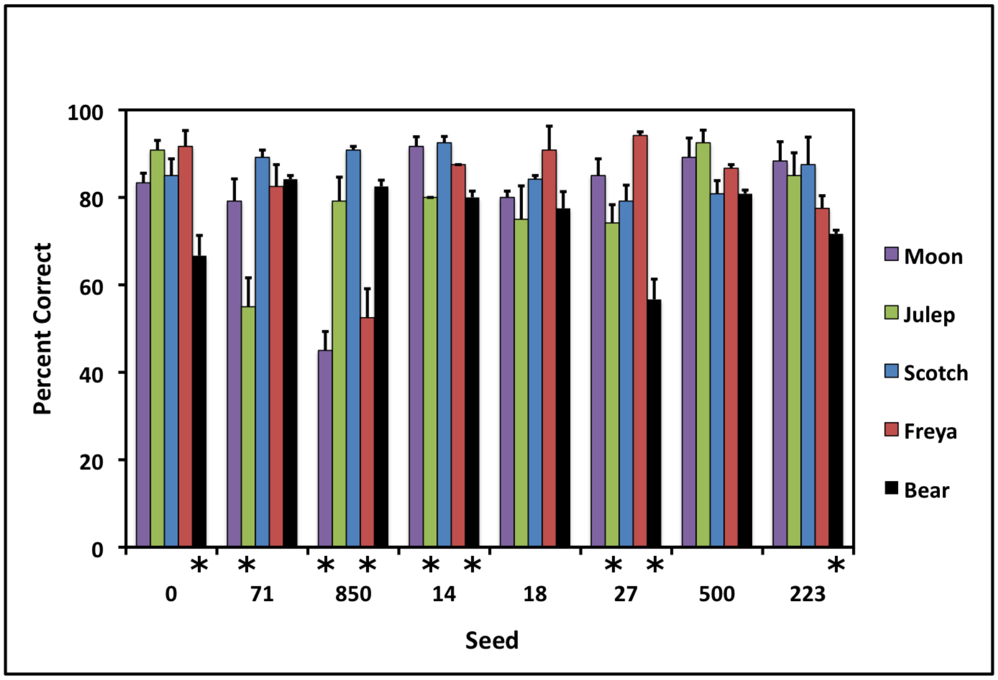


Figure 4.8: Individual discrimination performance for single burst reversals for all eight seeds tested. Results are averaged across the three burst locations. Error bars show standard error of the mean. Asterisks across the x-axis indicate significant differences from at least one other individual (detailed analysis provided in Table 4.7).

son correlation, r(38)=0.287, p=0.073), or the order in which seeds were presented (Pearson correlation, r(38)=0.116, p=0.476). These results suggest that individual differences in performance are due to the stimulus itself, rather than the testing conditions. More likely, the variation in performance is due to birds using different acoustic cues to complete the task. While the main cue present is temporal fine structure of the noise, there are small amplitude envelope changes, as well as small spectral changes that occur over time. It is possible that some birds are using these cues, and thus their performance on different seeds is not the same as the birds mainly using temporal fine structure cues.

Table 4.7: Results of one-way ANOVA for individual performance on each seed.

Results for Experiment 9						
Seed	**Bird**	**Mean**	**F**	**df**	**p**	**Tukey** $p < 0.05$
Seed 0	Moon	83.333 ± 3.819	8.600	4,10	0.003	Moon vs. Bear
	Julep	90.830 ± 3.819				Julep vs. Bear
	Scotch	85.000 ± 6.614				Scotch vs. Bear
	Freya	91.667 ± 6.292				Freya vs. Bear
	Bear	66.667 ± 8.036				
Seed 71	Moon	79.167 ± 8.780	9.106	4,10	0.002	Moon vs. Julep
	Julep	55.000 ± 11.456				Scotch vs. Julep
	Scotch	89.167 ± 2.887				Freya vs. Julep
	Freya	82.500 ± 8.660				Bear vs. Julep
	Bear	84.167 ± 1.443				
Seed 850	Moon	45.000 ± 7.500	21.095	4,10	< 0.001	Julep vs. Moon
	Julep	79.167 ± 9.465				Scotch vs. Moon
	Scotch	90.833 ± 1.443				Bear vs. Moon
	Freya	52.500 ± 11.456				Julep vs. Freya
	Bear	82.500 ± 2.500				Scotch vs. Freya
						Bear vs. Freya

Continued on next page

Results for Experiment 9 (continued)

Seed	**Bird**	**Mean**	**F**	**df**	**p**	**Tukey** $p < 0.05$
Seed 14	Moon	91.667 ± 3.819	20.500	4,10	< 0.001	Moon vs. Julep
	Julep	80.000 ± 0.000				Scotch vs. Julep
	Scotch	92.500 ± 2.500				Freya vs. Julep
	Freya	87.500 ± 0.000				Moon vs. Bear
	Bear	80.000 ± 2.500				Scotch vs. Bear
						Freya vs. Bear
Seed 18	Moon	80.000 ± 2.500	1.832	4,10	0.199	N/A
	Julep	75.000 ± 13.229				
	Scotch	77.500 ± 6.614				
	Freya	84.167 ± 1.443				
	Bear	90.833 ± 9.465				
Seed 27	Moon	85.000 ± 6.614	14.500	4,10	< 0.001	Moon vs. Bear
	Julep	74.167 ± 7.217				Julep vs. Bear
	Scotch	79.167 ± 6.292				Scotch vs. Bear
	Freya	94.167 ± 1.443				Freya vs. Bear
	Bear	56.667 ± 8.036				Freya vs. Julep
Seed 500	Moon	89.167 ± 7.638	3.473	4,10	0.050	N/A
	Julep	92.500 ± 5.000				
	Scotch	80.833 ± 5.204				
	Freya	86.667 ± 1.443				
	Bear	80.833 ± 1.443				

Continued on next page

Results for Experiment 9 (continued)						
Seed	**Bird**	**Mean**	**F**	**df**	**p**	**Tukey $p < 0.05$**
Seed 223	Moon	86.667 ± 1.443	5.021	4,10	0.018	Moon vs. Bear
	Julep	86.667 ± 1.443				Julep vs. Bear
	Scotch	87.500 ± 10.897				Scotch vs. Bear
	Freya	77.500 ± 5.000				
	Bear	71.667 ± 1.443				

4.5.4 Discussion

These results show that regardless of the population of noise used, birds do very well on fine structure discriminations when no syllable amplitude envelope cues are present. Repeating this with eight different populations of noise demonstrates the robustness with which zebra finches can follow fine structure in several different random patterns of noise. Each singular burst in a stimulus set contained 110 msec of a piece of random noise, with no repeating pattern within the burst. Thus, birds are relying on hearing small changes over short time scales when making these discriminations.

This also demonstrates that providing birds with roughly 100 msec of fine structure information is sufficient for them to make these discriminations. It would be interesting to test different durations of bursts (keeping the noise structure the same) to determine a duration vs. performance function to see whether this discrimination task is duration dependent as with natural song (Experiment 2). If it is duration dependent, then another question would be whether saturation of performance also occurs around 100 msec as with natural song, or whether there is an improvement in performance for bursts longer than 100 msec. A comparison of duration vs. performance for song syllables and noise bursts would further elucidate whether the temporal window findings in Experiment 2 extend to non-song sounds as well.

Even though noise bursts contain less amplitude envelope and spectral fine structure cues compared with natural song, small changes occur to both over time. Variation in individual performance suggests that birds may be relying on multiple cues to perform this task, even if those cues are relatively small compared with temporal fine structure cues. For a true test of birds' perception of temporal fine structure, the experiments in Chapter 5 make use of the Schroeder harmonic waveform, a stimulus in which the amplitude envelope and spectral profile remain constant across the entire duration. The only change that occurs is that the phase of each harmonic is either monotonically increasing (positive phase Schroeder) or decreasing (negative phase Schroeder). Thus, time reversals of Schroeder waveforms result in a change in phase, whereas amplitude envelope and spectra are unchanged.

Chapter 5

Perception of temporal fine structure in the context of global timing of song

5.1 Introduction

Results from the experiments presented in Chapter 4 showed that birds were able to discriminate between forward and reversed noise bursts in the presence (Experiments 6 and 7) and absence (Experiment 9) of syllable envelope cues. Much like speech, birds were able to rely on syllable envelope cues when spectral content had been replaced by random noise. In this case, fine structure cues were not necessary. However, Experiment 9 showed that birds are able to use fine structure cues when syllable envelope cues are removed. Individual variation in discrimination performance suggests that in addition to temporal fine structure cues, birds are also able to use small spectral and envelope cues present in the noise, and that different birds use different cues. Further experiments need to be done in order to truly isolate temporal fine structure cues from other cues that also occur over small time scales.

The stimuli used in these experiments were the Schroeder waveforms that have previously been used by Dooling et al. (2002) and Lauer et al. (2006). Schroeder waveforms are harmonic complexes in which the long-term amplitude envelope and spectrum remain constant over time, but the phase of the harmonics monotonically increases (positive-phase Schroeder), or monotonically decreases (negative-phase Schroeder) across frequencies. The result is a stimulus that only changes in phase information when reversed in time. The resulting frequency sweeps occur once per period of the waveform. Thus, temporal integration for discrimination of frequency sweeps can be tested by using a range of Schroeder waveform durations. Longer durations should be easier to discriminate, because more periods of the waveform and thus more frequency sweeps occur over the duration. Previous work by Lohr et al. (2006) has shown that zebra finches can discriminate reversals of single periods in a harmonic stimulus, and require an integration time of of 31.71 msec for 50% correct discrimination. Since experiments in this chapter test discrimination of positive and negative phase Schroeders that exist in a synthetic songs, the range of syllable durations will allow us to confirm whether temporal integration for discrimination of phase changes in Schroeders is similar to that found previously for harmonic stimuli.

5.2 Experiment 10: Discriminability of single Schroeder harmonic reversals within a song-like environment

5.2.1 Introduction

Temporal fine structure may play an important role in zebra finch song perception. Previous research has shown that zebra finches are quite proficient in detecting temporal fine structure changes within synthetic stimuli, specifically changes to phase and harmonic structure (Dooling et al. 2002; Lohr and Dooling 1998; Lohr et al. 2006). While these experiments test the perceptual limits of fine temporal processing in zebra finches, they do not ask whether birds use the same abilities when listening in a more natural setting to behaviorally relevant stimuli. All previous experiments testing zebra finches' ability to discriminate between phase changes have been in the context of short, single sounds (about the duration of a single syllable). The goal of this experiment is to test whether zebra finches are also able to discriminate changes to only temporal fine structure, when the stimuli had the same overall timing cues as song. This will be done by utilizing the Schroeder harmonic waveform, in a song-like context. Thus, synthetic Schroeder songs consist of Schroeder waveforms the same duration as individual song syllables, separated by the same duration as song intervals. This will both isolate temporal fine structure, and test discrimination of changes to temporal fine structure in a stimulus that has similar temporal envelope characteristics as song.

5.2.2 Methods

Subjects

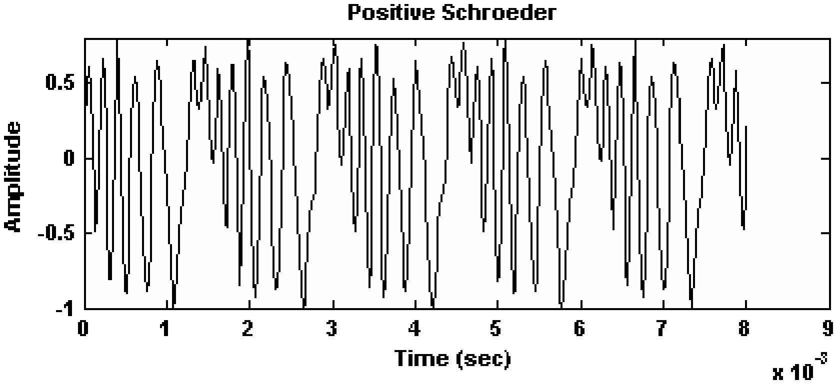
Four male zebra finches, three female zebra finches, and three female budgerigars were used in this experiment.

Stimuli

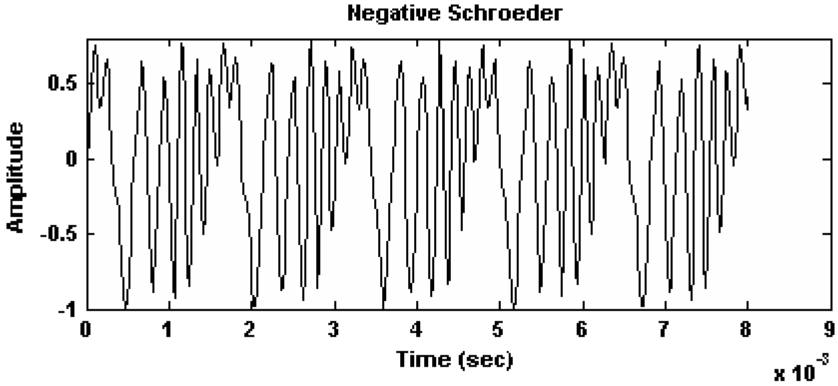
Schroeder waveforms were generated from a Matlab script written by Marjorie Leek. The waveforms had a fundamental frequency of 640 Hz, which is within the normal range of fundamental frequencies for both zebra finch contact calls (Simpson and Vicario 1990), and zebra finch song syllables (Williams 2001; Williams et al. 1989). The waveforms consisted of 10 components, had a frequency range of 640-6400 Hz, and had a 10 msec ramp at each end. The starting phases for each harmonic were determined by a modified version of the algorithm developed by Schroeder (1970):

$$\theta_n = C\pi n(n-1)/N, \tag{5.1}$$

where C is a scalar, n is the nth harmonic component, and N is the total number of harmonics in the waveform. This results in complexes in which the phase is either monotonically increasing or monotonically decreasing across frequency, and results in either upward or downward frequency sweeps within each period of the complex. The scalar term determines the speed and direction of frequency sweeps, and in this experiment we used +1 and -1. This scalar represents the slowest speed for frequency sweeps, resulting in a flat temporal envelope. A scalar of +1 means that the phase is monotonically increasing (positive phase Schroeder), and a scalar of -1 means that the phase is monotonically decreasing (negative phase Schroeder).



(a)

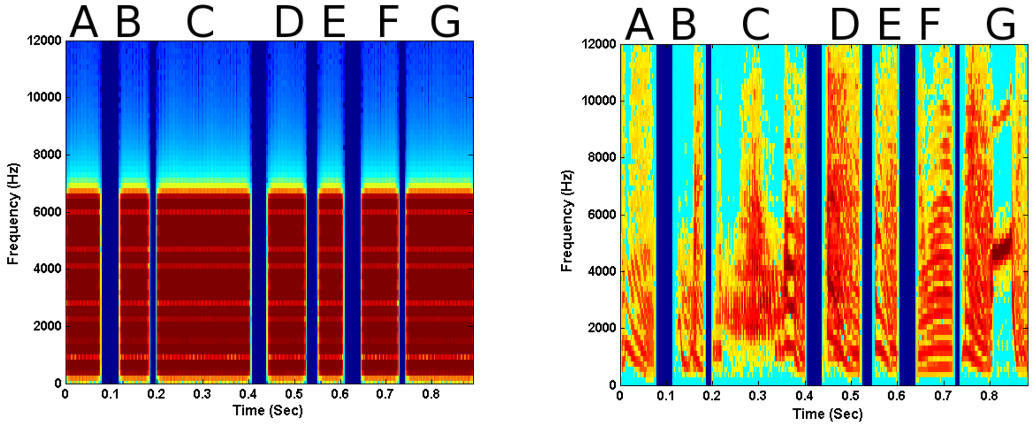


(b)

Figure 5.1: Examples of positive (a) and negative (b) phase Schroeder waveforms. Each waveform is 8 msec in duration. The fundamental frequency is 640 Hz, and extends to 6400 Hz. Positive and negative Schroeders are time reversed versions of one another, in which the phase is either monotonically increasing or monotonically decreasing across frequency.

Positive and negative phase Schroeders are shown in Figure 5.1.

For each song, Schroeders were generated to be the same duration as natural song syllables. Schroeders were concatenated with silence that was the same duration as natural song intervals. The resulting stimulus was a string of Schroeders with a fundamental frequency of 640 Hz, that had the same rhythm as that of natural song. For the background Schroeder song, only positive waveforms were concatenated. For targets, a single Schroeder was reversed in time so that there was one negative waveform, and the rest were positive waveforms. The task was to discriminate single



(a) Schroeder song

(b) Natural song

Figure 5.2: Comparison of a synthetic Schroeder song and the natural song from which it was modeled. Dark blue in spectrogram indicates pure silence between syllables. Both songs have the same syllable and interval durations and thus overall timing. However, Schroeder song lacks the variations in amplitude envelope and spectral features that natural song contains.

Schroeder reversals within a Schroeder song. A spectrogram of a Schroeder song is shown in Figure 5.2, along with a spectrogram of the original song from which the Schroeder song was modeled.

Experimental Design

The same design was used as in the previous experiments, in which all birds were tested on each Schroeder song, which were modeled after the songs of Moonshine, Bear, Scotch, and Julep. Thus, for male zebra finches, each was tested on Schroeder songs which were modeled after the BOS and three conspecific songs. The single Schroeder reversals corresponded with the same targets as in Experiments 2, 6, and 7, so that performance on single syllables could be compared in the different conditions. Table 5.1 shows the target set for each Schroeder song.

Table 5.1: Specific targets for Experiment 10: Schroeder songs

Moonshine's song (A,B,C,D,E,F)	**Bear's song (A,B,C,D,E)**	**Scotch's song (A,B,C,D,E,F,G)**	**Julep's song (A,B,C,D,E)**
1. A reversed	1. A reversed	1. A reversed	1. A reversed
2. B reversed	2. B reversed	2. C reversed	2. B reversed
3. C reversed	3. C reversed	3. D reversed	3. C reversed
4. D reversed	4. D reversed	4. E reversed	4. D reversed
5. E reversed	5. E reversed	5. F reversed	5. E reversed
6. F reversed	6. All reversed	6. G reversed	6. All reversed
7. All reversed	7. All reversed	7. All reversed	7. All reversed

Training

To acclimate birds to the Schroeder stimuli, birds ran 1-2 sessions in which all 7 targets had all Schroeders reversed (termed target "All reversed"). After birds ran on this with a false alarm below 20%, they began experimental stimulus sets in which targets had a single Schroeder reversal.

Testing Procedures

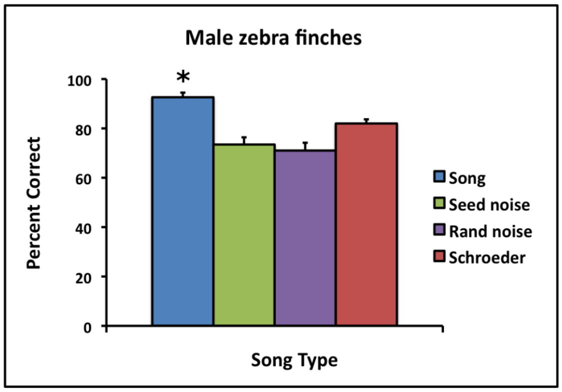
Schroeder songs were presented at 60 dB SPL, with a presentation rate of once per 1500 msec, and a response interval of 2500 msec. Because Schroeder songs lacked a song envelope, a presentation level of 60 dB was chosen for a more comfortable listening level. Thirty percent of trials were sham trials. For each block of 7 targets and 3 sham trials, the order in which targets and shams were presented was randomized. Birds ran on 300 trials, and the last 200 continuous valid trials were analyzed. Valid trials were trials in which the false alarm rate did not exceed 20%.

Budgerigars were unable to complete trials at 60 dB with a false alarm rate that did not exceed 20%, and were run on Schroeder songs that were presented at 80 dB SPL instead. To compare performance at the two different presentation levels, two zebra finches were also run on Schroeder songs presented at 80 dB SPL, and two budgerigars were run on Schroeder songs presented at 60 dB SPL. Since false alarm rates differed greatly between experiments, hit rates and false alarm rates were converted to d-prime scores, and differences were tested using 95% confidence intervals. D-prime scores that had overlapping confidence intervals were not significantly different. D-prime scores that had confidence intervals that did not overlap were significantly different.

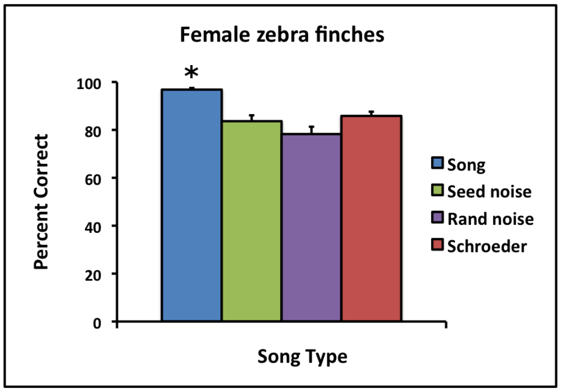
5.2.3 Results

Overall, birds were able to discriminate single Schroeder reversals at a high level of performance, even though the only acoustic cue present in Schroeder reversals was phase information. Overall performance on all four song types (natural, same-seed noise, random noise, and Schroeder) is shown in Figure 5.3 for each group. Performance was averaged across all syllables in all songs, and across all birds in the group. While performance on Schroeder song in general was lower compared to natural song, it was the same as or higher than random and same-seed noise songs (Kruskal-Wallis ANOVA, post-hoc Tukey test, see Table 5.2 for specific values).

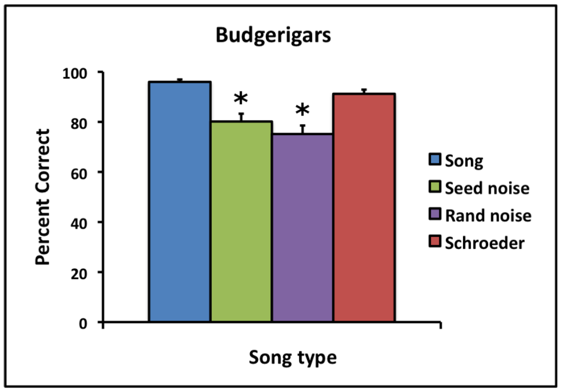
Performance on individual Schroeder reversals for the four songs is shown for for males, females, and budgerigars in Figures 5.4, 5.5, and 5.6, respectively. Average hit rates for natural song syllables are also shown in the plot for comparison purposes. Performance on most Schroeder reversals was closely matched to performance on the natural syllables. The decline in average performance compared to natural song was negatively correlated with syllable duration for zebra finches but not budgerigars (Males: $r(20) = -0.464, p = 0.0294$, Females: $r(20) = -0.620, p = 0.00206$, Budgerigars: $r(20) = -0.338, p = 0.123$). Budgerigar performance on Schroeder reversals was significantly higher than for male and female zebra finches,



(a) Discrimination performance: Male ZF



(b) Discrimination performance: Female ZF



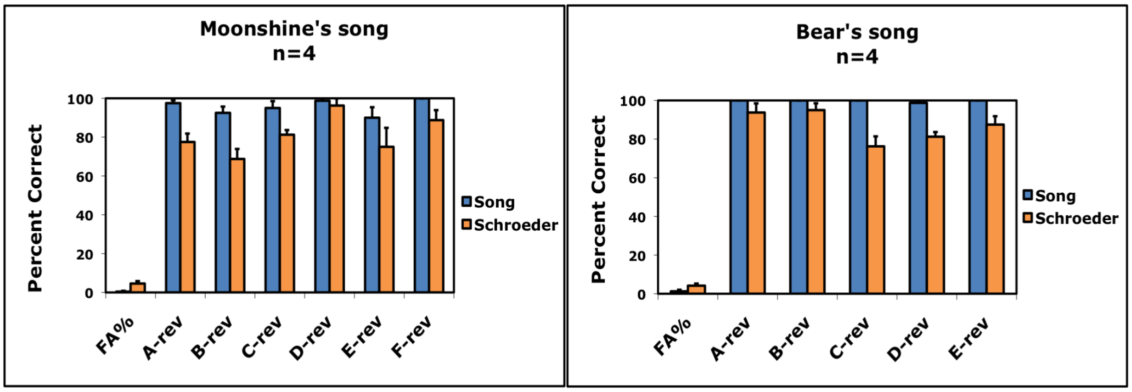
(c) Discrimination performance: Budgerigars

Figure 5.3: Overall performance on the four song types (natural, same-seed, random, and Schroeder) for males, females, and budgerigars. Performance on each song type was averaged across the 22 syllables tested. Error bars show standard error of the mean. Asterisks mark song types in which overall performance is significantly different from Schroeder song.

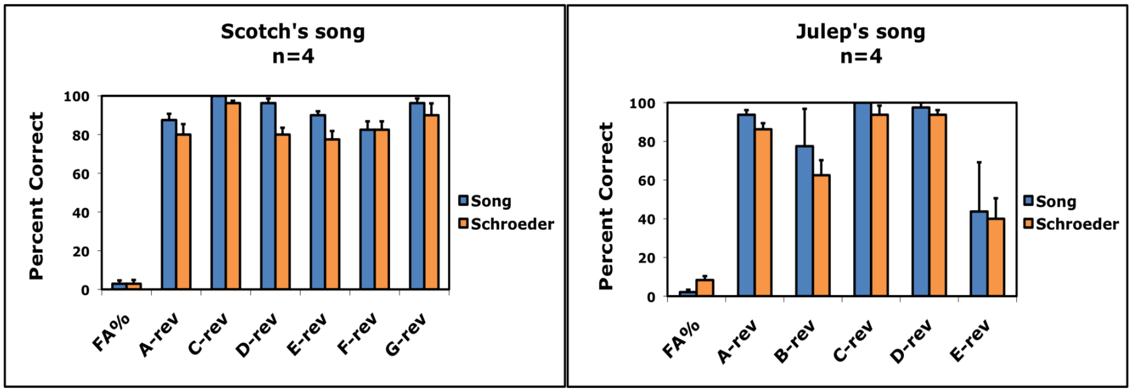
Table 5.2: Results of Kruskal-Wallis ANOVA comparing performance on the four song types.

Group	**Song Type**	**Mean**	**Median**	**H**	**df**	**p**	**Tukey** $p < 0.05$
Males	Natural	92.61 ± 17.55	100.00	67.916	3	< 0.001	**Natural vs. Schroeder**
	Seed noise	73.47 ± 27.33	85.00				Natural vs. Seed
	Rand noise	71.02 ± 29.65	85.00				Natural vs. Rand
	Schroeder	**81.99 ± 15.77**	**85.00**				
Females	Natural	96.82 ± 5.52	100.00	44.145	3	< 0.001	**Natural vs. Schroeder**
	Seed noise	83.64 ± 19.95	95.00				Natural vs. Seed
	Rand noise	78.26 ± 24.92	85.00				Natural vs. Rand
	Schroeder	**85.83 ± 14.40**	**90.00**				
Bud	Natural	95.98 ± 7.90	100.00	43.049	3	< 0.001	Natural vs. Seed
	Seed Noise	80.15 ± 25.34	92.50				Natural vs. Rand
	Rand Noise	75.15 ± 27.69	90.00				**Seed vs. Schroeder**
	Schroeder	**91.21 ± 13.65**	**97.50**				**Rand vs. Schroeder**

Boldface results in last column correspond to the asterisks in Figure 5.3



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

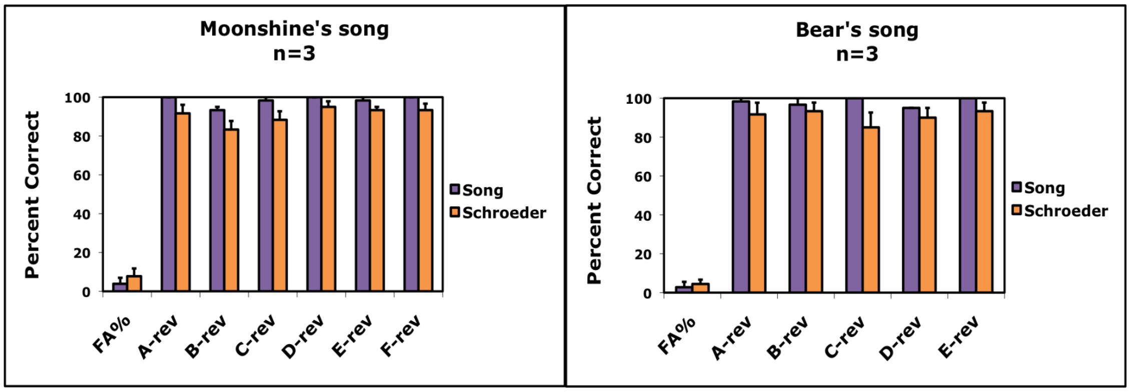


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

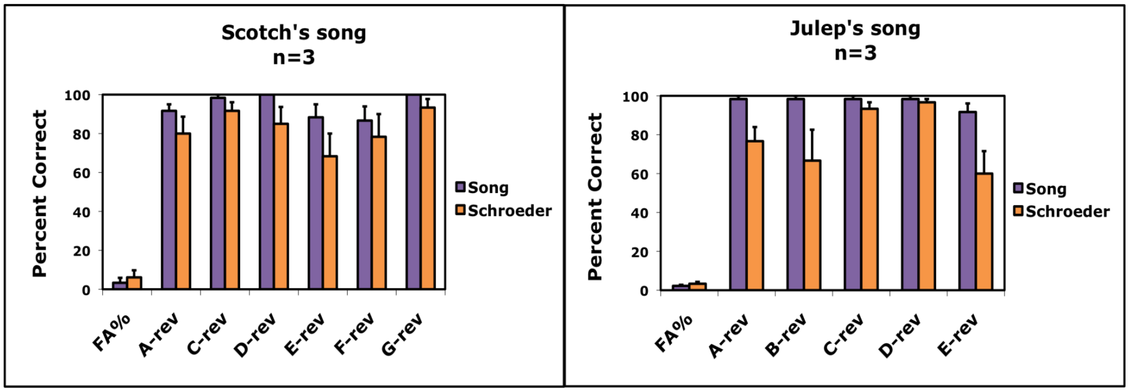
Figure 5.4: Comparison of average discrimination performance of male zebra finches on single syllable reversals in natural song (blue) and Schroeder song (orange). Error bars show standard error of the mean.

and this is likely due to the level difference, which will be discussed below (Kruskal-Wallis ANOVA, post-hoc Dunn's test, H=23.279, df=2, $p < 0.001$).

Interestingly, all three groups showed a duration effect for single Schroeder reversals (Figure 5.7). Performance stabilized for Schroeders longer than ~ 130 msec in duration (around 93% correct for zebra finches and 97% for budgerigars). For Schroeder reversals less than 130 msec, performance was positively correlated with Schroeder duration for all three groups (Males: $r(20) = 0.768, p = 0.00216$; Females: $r(20) = 0.685, p = 0.00983$; Budgerigars: $r(20) = 0.564, p = 0.04480$). Male zebra finch performance showed this duration effect with natural, random, same-seed, and



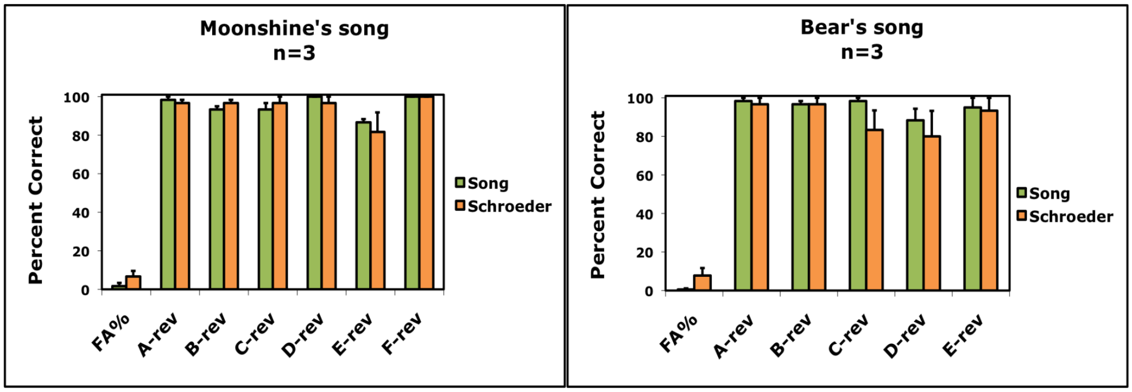
(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



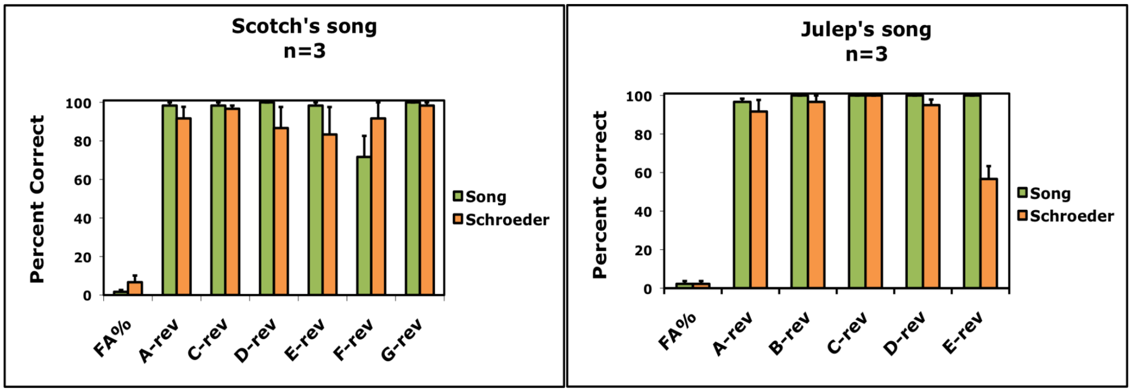
(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 5.5: Comparison of average discrimination performance of female zebra finches on single syllable reversals in natural song (purple) and Schroeder song (orange). Error bars show standard error of the mean.

Schroeder songs. Female zebra finches and budgerigars only showed this duration effect with Schroeder song. For the shortest duration Schroeder (30 msec), average percent correct discrimination was 40.00% for males, 60.00% for females, and 56.67% for budgerigars. These values are similar to the 31.71 msec seen by Lohr et al. (2006) for period reversal discrimination.

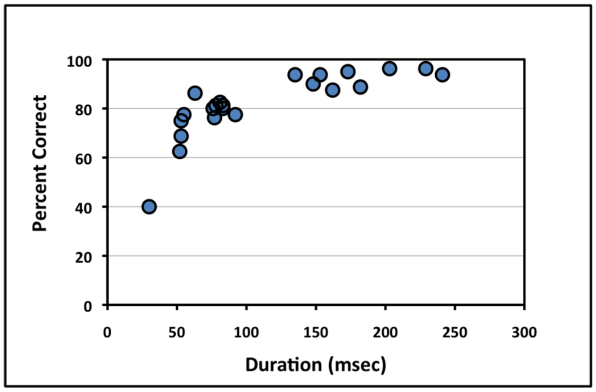


(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

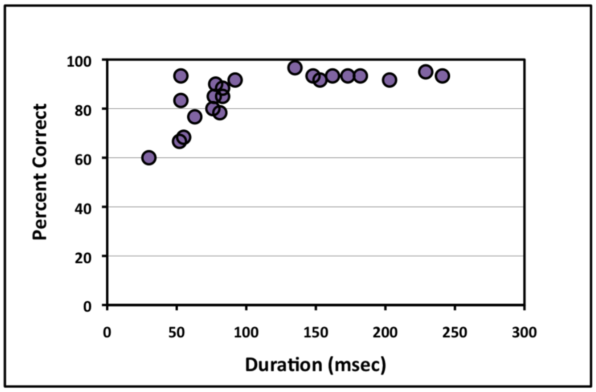


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

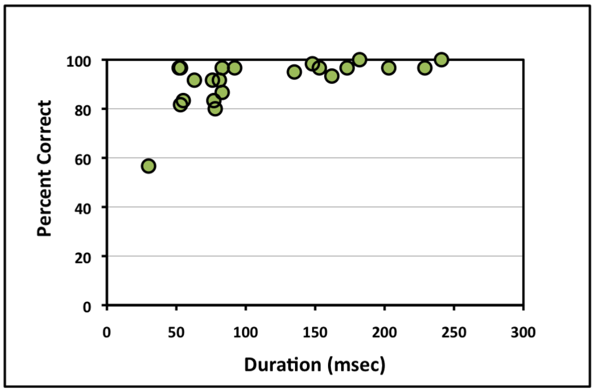
Figure 5.6: Comparison of average discrimination performance of budgerigars on single syllable reversals in natural song (green) and Schroeder song (orange). Error bars show standard error of the mean.



(a) Duration effect for male zebra finches



(b) Duration effect for female zebra finches



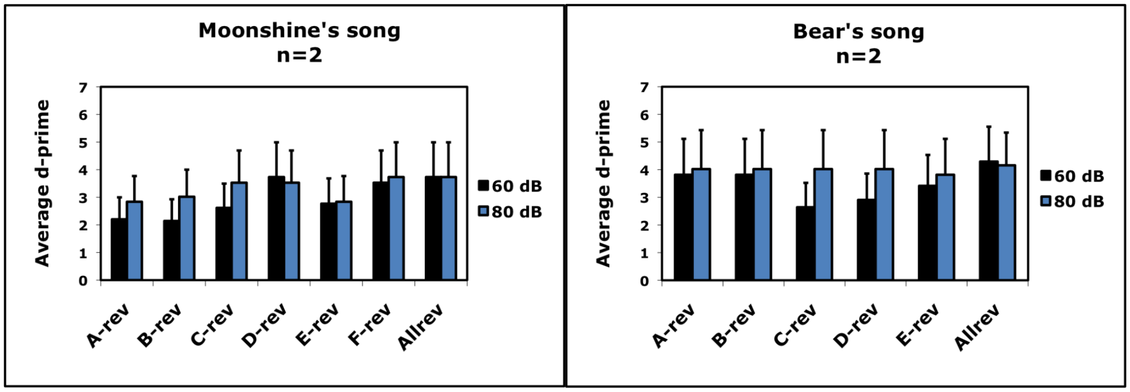
(c) Duration effect for budgerigars

Figure 5.7: Discrimination performance of males (blue), females (purple), and budgerigars (green) on single Schroeder reversals as a function of Schroeder waveform duration.

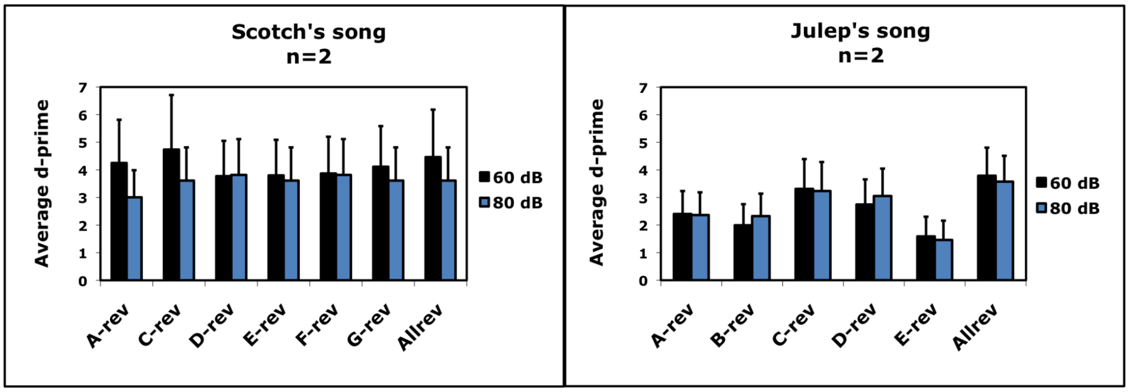
Level differences

Zebra finches discriminated single Schroeder reversals presented at 60 and 80 dB with equal sensitivity. This is illustrated in Figure 5.8. Specific d-prime scores and standard error values are shown in Table A.5 in Appendix A. For all single Schroeders tested, d-prime scores for 60 and 80 dB presentation levels had overlapping confidence levels, and thus any differences are likely due to chance. This indicates that zebra finch performance is saturated at 60 dB SPL, and an increase in the presentation level does not improve discrimination performance. Thus, fine temporal discrimination in zebra finches may not depend on the loudness of the stimuli. Further experimentation would be needed in order to confirm this.

Budgerigars in contrast, showed differences in sensitivity for single Schroeder reversals at different presentation levels. D-prime scores were significantly lower for roughly half of the Schroeder reversals presented at 60 dB, compared with the same Schroeder reversals presented at 80 dB. These results are shown in Figure 5.9. Specific d-prime and standard error values are shown in Table A.6 in Appendix A. Of the 13 Schroeders that showed increased discriminability when presented at 80 dB, 11 were less than 100 msec in duration. Likewise, of the 9 Schroeders that showed no change in discriminability, 7 were over 100 msec in duration. Together, these results indicate that budgerigar performance on temporal fine discriminations is dependent upon sound pressure level, and that the Schroeders most affected by level differences are ones that are generally less than 100 msec in duration.

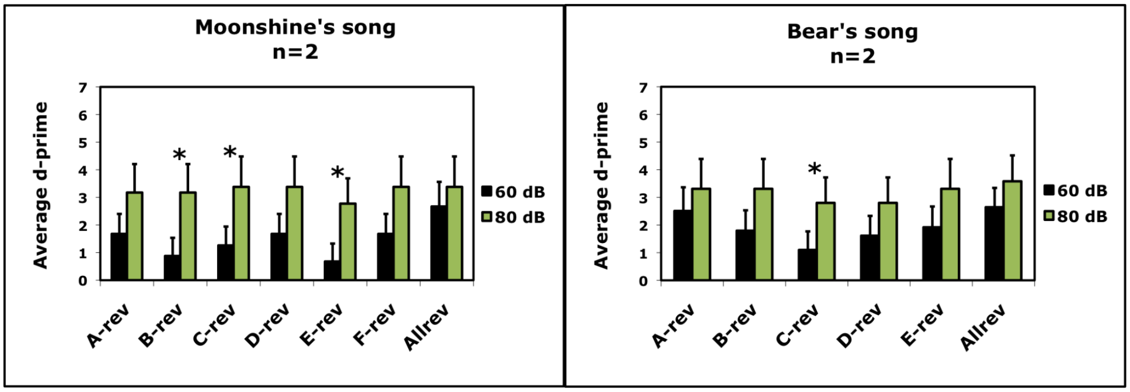


(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

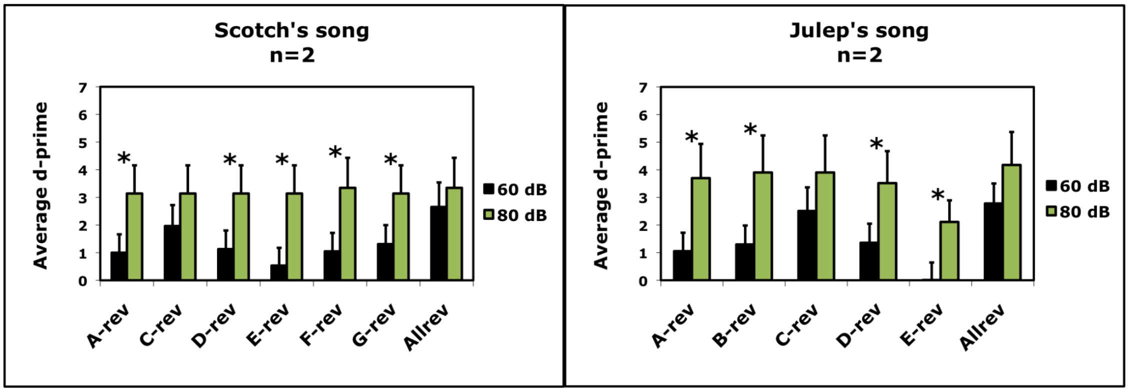


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 5.8: Comparison of discrimination performance of male zebra finches on single Schroeder reversals presented at 60 and 80 dB. Performance is shown as average d-prime scores. Error bars show 95% confidence intervals.



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure 5.9: Comparison of discrimination performance of budgerigars on single Schroeder reversals presented at 60 and 80 dB. Performance is shown as average d-prime scores. Error bars show 95% confidence intervals. Asterisks indicate significant differences between the two SPL levels.

5.2.4 Discussion

The high level of performance on single Schroeder reversals indicates that birds are capable of discriminating changes to temporal fine structure, when no other acoustic cues are present. All three groups showed a duration effect upon performance for Schroeders less than 100 msec in duration. This contrasts with reversal discrimination in natural and noise song (Experiments 2, 6, and 7), in which only male zebra finches showed a duration effect. This suggests that in natural song, male zebra finches may rely on different acoustic cues than female zebra finches and budgerigars. Perhaps male zebra finches focus on fine temporal cues even when other cues such as syllable envelope and spectral structure are present. Females and budgerigars on the other hand, may use a combination of the cues present in natural song. If this is the case, we expect females and budgerigars to perform similarly on all syllables, regardless of duration since multiple cues are available to them. However, in Schroeder synthetic song, females and budgerigars can only rely on temporal fine structure cues. In this case performance is dependent upon duration of the syllables. Since male zebra finches show a duration effect for natural, noise, and Schroeder songs, it stands to reason that in all three cases they are mainly using temporal fine structure cues to make these forward/reverse discriminations.

The results in this experiment confirm previous results regarding temporal fine structure discrimination Dooling et al. (2002); Lauer et al. (2006); Lohr et al. (2006) and show that birds are capable of temporal fine structure discrimination in a song-like acoustic environment. Schroeder harmonics in this experiment had a fundamental frequency of 640 Hz, similar to that of zebra finch vocalizations. Previous work has indicated that zebra finches may have greater temporal precision than other birds and humans, as they are able to discriminate between positive and negative phase Schroeders up to 1000 Hz. Performance in budgerigars and canaries drops for frequencies higher than 700 Hz, and performance in humans drops for frequencies higher than 300 Hz. Since frequency sweeps occur once per period, Schroeders with higher fundamental frequencies have shorter periods and the frequency glide occurs over a shorter time. Zebra finches are able to discriminate changes that occur over 1-2 msec, but other birds and humans need longer periods.

Zebra finches' proficiency at discriminating changes to temporal fine structure was further demonstrated in our experiments comparing performance at two different presentation levels. Zebra finches showed no difference in discrimination performance for Schroeders presented at 60 dB and 80 dB. In both cases, performance was roughly 80.00%, with a false alarm rate below 20%. Budgerigars however, were unable to complete the criterion of $< 20\%$ false alarm rate for Schroeder reversals presented at 60 dB. This high false alarm rate means that the task was very difficult for budgerigars. When d-prime scores were compared at the two presentation levels, budgerigars showed a vast improvement at 80 dB, whereas zebra finches showed no difference. It appears that budgerigars' ability to do fine temporal discriminations may be limited by the loudness of the stimulus, whereas zebra finches' ability is not. Recall in Experiment 2 that budgerigars had significantly worse performance on stack syllables, compared with other syllable types. Stack syllables are harmonic, and thus have fairly even amplitude envelopes and spectral components across time. Thus, reversal discrimination of stack syllables may be similar to Schroeders in which the main cue present is temporal fine structure.

An interesting follow up to this experiment would be a systematic test of Schroeder reversals within a Schroeder song, at various sound pressure levels. This could answer several questions, including the lowest presentation level that zebra finches could reliably discriminate single Schroeder reversals, and whether performance is correlated with duration for all sound pressure levels.

5.3 Experiment 11: Comparison of frequency discrimination thresholds within and outside of a song-like context

5.3.1 Introduction

One final question asked in this dissertation is whether auditory perception in birds is affected by context. Many previous studies measuring absolute thresholds for frequency discrimination, duration discrimination, and gap detection have used single or pairs of stimuli. In order to communicate with others, most of the sounds birds hear on a daily basis are embedded into a song context. Zebra finch song motifs contain at least three elements and as many as eight, all with different durations and spaced with different inter-syllable interval durations. To test whether thresholds for frequency discrimination differ depending on context, frequency discrimination thresholds were conducted in two cases and compared: One single element, and the same element embedded in a song context. Since zebra finch songs contain harmonic syllables, Schroeder waveforms and Schroeder songs were used. By using Schroeders, fundamental frequency could be manipulated, while all other cues remained the same.

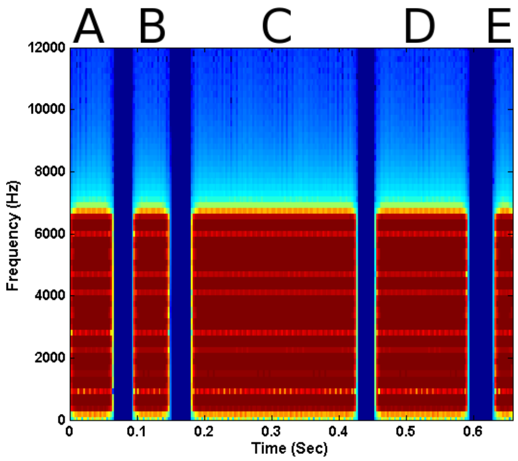


Figure 5.10: Julep's Schroeder song. The fundamental frequency of syllable D was shifted in 1 Hz steps for frequency discrimination threshold testing.

5.3.2 Methods

Subjects

Two male zebra finches were used in this experiment.

Stimuli

Stimuli consisted of either Julep's Schroeder song (see Figure 5.10), in which all syllables had a fundamental frequency of 640 Hz, or a single Schroeder waveform with a fundamental frequency of 640 Hz.

Experimental Design

For the song context case, targets consisted of the same Schroeder song as the background, but with syllable D (see Figure 5.10) shifted in fundamental frequency from 641 Hz (target 1) to 647 Hz (target 7), in 1 Hz steps. All other syllables (A,B,C, and E) had fundamental frequencies of 640 Hz. Julep's syllable D is 135 msec in duration.

For the single case, the same 135 msec duration Schroeder with a fundamental frequency of 640 Hz served as the background, and targets consisted of this single Schroeder with the fundamental frequency shifted from 641 Hz (target 1) to 647 Hz (target 7) in 1 Hz steps.

Training

Since birds were acclimated to Schroeder stimuli, no training was necessary.

Testing Procedures

For the song context case, Julep's Schroeder song was presented at 70 dB SPL, at a rate of once per 1500 msec, and a response interval of 2500 msec. For the singular case, Schroeders were presented at 70 dB SPL, at a rate of once per 500 msec, with a response interval of 2000 msec. Thirty percent of trials were sham trials. For each 10-trial block, targets and shams were presented in random order. Birds were tested until performance stabilized (at least 400 trials) such that the threshold for 50% correct frequency discrimination for the last 200 trials did not differ from the threshold for the last 100 trials by more than 1/3 Hz.

5.3.3 Results

For both birds tested, thresholds for 50% correct frequency discrimination did not differ substantially for Schroeder waveforms tested singly, and Schroeder waveforms within a Schroeder song (Figure 5.11). Differences between the two conditions were less than 1/2 Hz for both birds. There was a roughly 1 Hz difference in frequency discrimination thresholds between the two birds. However, this is likely accounted for by individual variation.

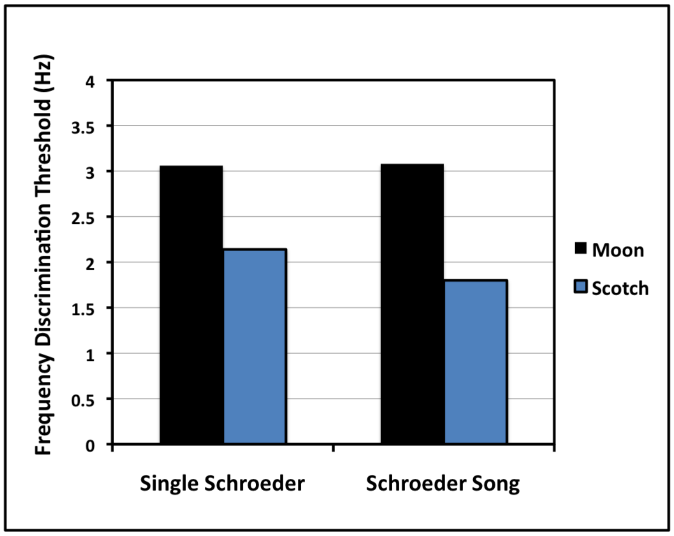


Figure 5.11: Comparison of thresholds for frequency discrimination in a single Schroeder and in a Schroeder syllable of a Schroeder song.

5.3.4 Discussion

In both cases, frequency shifts were made to a Schroeder waveform that had a fundamental frequency of 640 Hz, and a duration of 135 msec. Thresholds for 50% correct discrimination did not differ regardless of whether the Schroeder was presented singly, or as part of song-like stimulus. Birds could discriminate frequency changes equally well in the stimulus that contained multiple elements of differing durations, with a specific tempo. This suggests that birds are able to listen to multiple features at once and still discriminate changes in a single feature, in this case frequency.

One question brought up by the design of this experiment is whether birds are comparing syllable D in the target motif to syllable D of the background motif, or whether they are comparing syllable D in the target motif to the adjacent syllables in the target motif, C and E. The same question may be asked of Experiment 9, in which the stimulus contains the same noise burst repeated three times. There is no doubt that birds are discriminating these changes, but it is unclear what standard birds are using for comparison.

A suggested extension for Experiment 11 would be to create a Schroeder song in which the fundamental frequency of each of the syllables A,B,C,D, and E was different, so that adjacent syllables cannot serve as a standard for comparison. In order to directly compare with the above experiment, syllable D should have a fundamental frequency of 640 Hz, and target motifs should shift the fundamental of syllable D from 641-647 Hz in 1 Hz steps. If the threshold for frequency discrimination is significantly higher compared with the case in which all Schroeder syllables had the same fundamental frequency, then it is clear that in the first experiment, birds were using adjacent syllables for this task.

This type of Schroeder song would also more closely approximate natural song, since each syllable would be unique. Thus, a higher threshold would also indicate that frequency discrimination for an element within a song environment is more difficult than for that element presented alone, likely due to constraints in working memory and attention. Conversely, if thresholds are found to be similar to the original experiment, this indicates that even in a complex song environment, birds have the attentional mechanisms required to discriminate changes to just a single element.

The final chapter will discuss the implications for the results found in each of experiments presented, and future work that may provide further insight.

Chapter 6

General Discussion

The song of the zebra finch, though simple in its syllable sequencing, is quite complex in its acoustic structure. Song motifs contain between three and eight syllables, each of which has a unique pattern of spectral, temporal, and amplitude envelope cues. Motifs are sung in a repetitive manner, with little variation from rendition to rendition. Syllables and the intervals separating them have consistent durations that provide an overall rhythm to the song. While songs differ greatly between unrelated birds, related and neighboring birds share many of the same song elements, as song is learned through social interaction and tutoring.

Perception of song by both male and female zebra finches is essential to species survival. Males defend their territories and resources through singing to other males, and begin mating displays through singing to females. Thus, the role of song perception differs for males and females, but is equally important. Males must identify familiar birds based on the songs they sing to determine if they should move to another territory, whereas females must judge songs based on quality and complexity to determine if the singer is a good mate. In addition, the sexual dimorphism of song behavior may also play a role in perception, as females do not sing and lack the brain circuitry that males have for this purpose.

Through the use of auditory discrimination experiments, we have been able to determine the relative salience of time scales, and acoustic cues within zebra finch song motifs. The results presented in the previous chapters add to the way in which we think about zebra finch song, as a mode of animal communication, and as a model for human speech development.

6.1 The role of song perception in male and female zebra finches

Song serves two purposes for male zebra finches: to defend their territories from other males, and to show their genetic fitness to females as a potential mate. This suggests the possibility that males and females may listen for different acoustic cues in song, as each cue may provide a different piece of information. Male zebra finches need to identify and discriminate between singers, many of which sing a similar song. Females, on the other hand, need to use song to determine the health of the singer, and his ability to provide resources.

Results from our experiments suggest that male may attend more to temporal fine structure in song syllables than females. Because of this, males are more sensitive to the duration of song syllables than females. Temporal fine structure changes are perhaps the most subtle type of changes that can occur within song, and thus require a longer duration over which to discriminate. Temporal fine structure specifically refers to phase and harmonic structure in syllables, which contribute to the timbre and overall tone of song. These small changes may be important for male song perception specifically, because the songs of related zebra finches likely differ along this dimension, rather than in their amplitude envelope or spectral cues, which are transmitted with a high amount of fidelity during song learning. Two birds singing the exact copy of a song may still subtly differ in certain acoustic features such as timbre, simply due to individual differences in the vocal tract (Forstmeier et al. 2009).

Adult male zebra finches show a preference for the songs of their fathers (i.e. their tutor song) over unfamiliar conspecific songs. However, this preference does not extend to the songs of unfamiliar brothers that were tutored by the same bird, but in another clutch (Riebel and Smallegange 2003). The songs of unfamiliar brothers share considerably more song syllables with the father's song than conspecific songs. However, these songs are not treated the same as the tutor song. This preference for only the father's song means that male zebra finches do not generalize between songs that have similar syllabic structure and syllable sequencing as the tutor song. It is possible that because many neighboring zebra finch songs may share many common syllables, that male zebra finches must rely on temporal fine structure cues to differentiate between them.

Female zebra finches, on the other hand, show a preference for songs that are complex, as song complexity has been shown to be an indicator of mate quality. Specifically, song quality has been shown to correlate with developmental health (Spencer et al. 2003), neuroanatomical development (DeVoogd 2004), and the ability to learn new foraging tasks (Boogert et al. 2008). Song complexity includes the total number of syllables in the motif, the number of distinct syllables in the motif, and the spectral structure of the individual syllables. Spencer et al. (2003) found that that limiting access to food, or injecting cortisone (an indicator of stress) in male zebra finches during development resulted in songs with less syllables, and lower peak frequencies compared with normal controls. These findings suggest that the size and diversity of the syllable repertoire (via spectral structure) may be important factors for mate choice. This biological drive to attend to song quality over song identity may be one reason why females may attend more to spectral structure and amplitude envelope cues, but not temporal fine structure cues in song.

Vocal experience with song, and anatomical differences in the avian forebrain may also account for differences in song perception in male and female zebra finches. Since females do not sing, it is difficult to test this hypothesis experimentally. However, studies involving the zebra finch long call has provided insight into this possibility. Male and female zebra finches both produce long calls (LC), mainly in situations in which they can hear but not see other birds around them. It has been shown by Zann (1984) that male and female LC differ along three acoustic parameters. Compared with female LC, male LC are shorter in duration and are more stereotyped, have a higher fundamental frequency, and contain fast frequency modulations at either the beginning or the end of the call. Due to the combinatorial power of these three parameters, male LC are each unique to the individual, whereas female LC are more similar to one another. Male LC also differ from female LC in two other ways. First, the beginning portion containing fast frequency modulations is learned from a tutor, whereas female LC are innate (Zann 1985). Secondly, male LC are produced by the same vocal motor pathway that produces song (Simpson and Vicario 1990). In females, this pathway is very much reduced, as the nucleus RA is much smaller than in males. Lesions of HVC, RA, and NXIIts in the vocal motor pathway affect the morphology of male LC, but leave female LC intact.

The sexual dimorphism of LC is seen not only in production, but also in perception. In call back experiments, male zebra finches show a preference for female LC over male LC. This preference for female LC is also seen in females, but is much less pronounced. Moreover, male and female zebra finches were found to categorize LC using different sets of acoustic cues (Vicario et al. 2001b, 2002). Males reliably discriminate LC based on gender, using fundamental frequency, duration, and fast frequency modulation cues, all of which are present in male LC. Females on the other hand, only use duration cues when discriminating LC, and are insensitive to fundamental frequency and fast frequency modulation cues. Female LC lack fast frequency modulation, and vary much less in fundamental frequency than male LC. Thus, it is possible that vocal production acts to shape auditory perception.

Further work by Vicario et al. (2001a) suggests that this link may be through anatomical structures in the vocal motor pathway. Lesions of RA in male zebra finches change the morphology of their LC, and makes them similar to female LC. In other words, the duration becomes longer, the fundamental frequency lower, and fast frequency modulations are lost. In addition, males with lesions of RA lost their preference for female LC, and no longer discriminated LC based on fundamental frequency and fast frequency modulations. Once they lost these acoustic cues in their own LC, they also no longer used them to discriminate others' LC. Recent work by Lei and Mooney (2010) has shown that neurons in HVC receive auditory feedback from vocalizations through indirect connections with the thalamic nucleus ovoidalis. Perturbation of ovoidalis results in altered vocalizations. Perhaps perception of certain features in calls and song is dependent upon production of those features, and the auditory feedback that results from their vocal output.

It is likely that male and female zebra finch perception of song is shaped by vocal experience, anatomical constraints, and biological drives, in concert. Our results provide one indication of this, suggesting that males and females may focus on different acoustic cues when listening to song. Further work in this area could provide valuable insight into the biological basis of song perception. More specifically, one area for research would be how male and female zebra finches treat songs of related zebra finches. These songs could be nearly identical in the syllables and sequencing, but differ in other fine temporal characteristics due to individual differences in vocal tracts. If males and females perceive these songs differently, this would provide further evidence that males and attend to different cues when listening to the same set of songs, and that this is likely due to differences in how song is used by males and females. Based on our previous results, we would expect male zebra finches to be better than females at discriminating between two similar songs. Moreover, we expect that males would perform best when one of the songs tested is the tutor song.

6.2 Implications for human vocal development

Birdsong has served as a model for human vocal development for several decades. The basis for this comes from parallels with speech seen in both development and maintenance, such as learning during a critical period, similar stages of development, and the necessity of auditory feedback. However, there have been very few demonstrations that birdsong and speech share perceptual similarities. As the purpose of both song and speech is to be perceived by a receiver as a form of communication, this has been a missing piece in the model. Our results suggest some similarities in the way song and speech are perceived, as well as some differences. A better understanding of these similarities and differences will not only inform us of the limitations of this model, but also how to utilize it to the fullest.

Results from Experiments 6 and 7 in this dissertation suggest that the role of envelope cues in song and speech seem to be similar. Birds were able to reliably discriminate reversals of syllables in which the syllable envelope was filled with random noise, and performance was significantly correlated with the rise/fall asymmetry of the syllable amplitude envelope. Performance on natural song was not correlated with rise/fall asymmetry of the syllable envelope, even though the same song envelope is present. Thus, when spectral cues are absent from song, birds are able to rely on syllable envelope cues to complete the task. A similar use of envelope cues in speech was shown by Drullman (1994) in which speech envelopes were filled with random noise, and listeners were asked to repeat back sentences that they heard. Subjects were able to repeat back nearly all of the sentences, indicating that in quiet listening conditions envelope cues are sufficient for speech intelligibility.

In both cases, both birds and humans can reliably use envelope cues in the absence of spectral structure in order to understand song and speech. While discrimination experiments are not a direct test of understanding, discrimination between a natural song motif, and a motif in which a single syllable has been reversed requires recognizing the syllables in the motif and forming an expectation of what should come next in the sequence. The fact that removal of spectral structure does not affect performance in birds and humans indicates that much of the information present in song and speech can be transmitted largely through envelope cues alone.

Perception of fine structure cues in speech and song appear to differ, specifically when presented alone. In speech, perception in non-quiet conditions requires the presence of fine structure cues in addition to envelope cues. Since background noise is often present, either in the form of other speakers or environmental noise, both cues are used simultaneously in most cases. However, when envelope cues are removed from speech, the remaining fine structure cues are not enough to reliably understand speech Drullman (1994). In fact, speech becomes nearly unintelligible. Zebra finches on the other hand, are able to discriminate syllable reversals when only fine structure, specifically temporal fine structure, cues are present. Removing the spectrum and amplitude envelope cues from syllables had little effect on discrimination performance. While the temporal fine structure in Schroeder song is not the same as natural song syllables, zebra finches are nonetheless able to rely on such types of cues in the absence of spectrum and amplitude envelope.

Birds performed best when all acoustic cues were present in song (as humans do with speech), but they performed nearly as well when only temporal fine structure cues were present. When only envelope cues were present, birds were still able to do the task, but performed worse. These results suggest that fine structure cues, and specifically temporal fine structure, may be more important than envelope cues in song perception. However, to confirm this, reversal discrimination studies would need to be done with song in which the natural syllable fine structure cues remained, and syllable envelope cues are removed or replaced with a random envelope. If birds performed just as well as with Schroeder song, then birds may use fine structure in song differently from how humans use fine structure in speech. If these differences continued to hold after further testing, it could be either due to differences in avian and human anatomy of the ear, or simply due to environmental differences. Australian zebra finches live in grasslands and forests, usually close to water. Perhaps spectral and temporal fine structure is necessary for transmission of song through vegetation, even more so than human speech.

One way in which to test the influence that habitat has on zebra finch perception of spectral and temporal fine structure in song is to test perception in a subspecies of zebra finch, the Timor zebra finch (*Taeniopygia guttata guttata*). The Timor zebra finch lives on the Timor islands of Indonesia, and sings both a higher pitched and faster song than the Australian zebra finch (*Taeniopygia guttata castanotis*). Due to the dense vegetation of the rainforest habitat of the Timor finch, it is possible that Timor zebra finches rely on time structure even more so than the Australian zebra finch for the propagation of song. We would expect Timor zebra finches to be nearly perfect on reversal discriminations of syllables that contain only fine structure, but much worse than Australian finches on reversal discriminations of syllables that contain only syllable envelope cues. Perhaps other oscine birds that live in more open areas would rely on syllable envelope cues more than fine structure cues, as humans do in speech. A comprehensive study of several songbird species from different habitats on the perception syllables in their own songs would give insight into how environmental constraints shape the relative salience of temporal envelope and fine structure cues in song.

Another possibility for the differences in fine structure perception in song and speech could be in the vocalizations themselves. Perhaps envelope cues in speech contain more information than envelope cues in song. Conversely, fine structure in song may contain more information than fine structure in speech. This falls in line with our results. While Drullman (1994) found that subjects could repeat back $\sim$ 98% of sentences made of noise speech, zebra finches could discriminate reversals of noise syllables $\sim$ 70% of the time or greater. Birds showed a much larger decrease in performance compared to natural song than humans did when compared to natural speech. To test this hypothesis, zebra finches could be tested on reversals of song syllables with envelope cues removed, and speech syllables with envelopes removed. A baseline test is necessary to determine how birds perform on reversals of natural speech syllables when all acoustic cues are present. If performance is the same for natural speech and speech with envelope removed, this suggests that birds are better than humans at utilizing fine structure cues present in vocalizations. If birds perform worse on reversals of speech when envelope cues are removed, then this suggests that song and speech and fundamentally different in their fine structure cues. A recent study has shown that zebra finches are able to categorize human speech phonemes, independent of the speaker (Ohms et al. 2010). It would be interesting to determine exactly which cues birds are using for this task, and whether humans use the same cues as birds in this categorization.

6.3 How electrophysiology and psychoacoustics inform one another

A great deal has been learned about the neural mechanisms of song learning and production through electrophysiological recordings of the neurons in the anterior forebrain and motor production pathways, both in awake behaving birds and those asleep. However, little has been done to test predictions made from electrophyiology in actual behavior, specifically song perception. Given that the anterior forebrain pathway is a specialized circuit for song learning and is separate from the pathway for song production, neuronal responses in Area X and LMAN may reflect specializations necessary for song perception. If this is the case, then some of the most important findings from recording from Area X, LMAN, and HVC_X neurons can guide what behavioral and perceptual questions should be asked. Here, we focused on three aspects of song perception that have been previously tested using electrophysiological techniques: BOS sensitivity, local vs. global temporal sensitivity, and temporal windows for listening to song.

6.3.1 Bird's own song (BOS) sensitivity

Electrophysiological studies have shown that neurons in the anterior forebrain pathway of adult zebra finches exhibit preferences for the BOS over other stimuli. Neurons in HVC (Margoliash and Fortune 1992; Theunissen and Doupe 1998), Area X (Solis and Doupe 1997), and LMAN (Solis and Doupe 1997) show the largest firing responses when the BOS is presented. These neurons respond much less to presentation of conspecific songs, or even to the tutor song (Nick and Konishi 2005). These specializations for the BOS begin during song learning, and it was shown by Nick and Konishi (2005) that HVC of juvenile zebra finches actually respond best to the tutor song during the early sensorimotor phase of learning. However, by the late sensorimotor phase, preferences for the tutor song decrease, and preferences for the BOS increase and persist through adulthood. Thus, BOS sensitivity may be used for song learning, refinement, and maintenance throughout life. However, the question still remains whether this sensitivity for the BOS affects how song is perceived. Does BOS sensitivity translate into birds being more sensitive to changes in their own songs over conspecific songs?

Results found in Experiment 2 suggest that during syllable discrimination tasks, birds are attending to the acoustic structure of the songs, and not the actual identity of the songs. All four males showed a similar pattern in which the average response latencies were the shortest for Bear's song. Response latency is a more sensitive measure than hit rate, and can tell us not only if birds can discriminate changes to song, but exactly *when* they discriminate these changes. Thus, it appears that the BOS does not provide an advantage in perceiving changes to song. Instead, one song appears to be an easier target set for all of the birds. This song had the longest average syllable duration of the four songs tested. We showed in Experiment 2 that birds were listening to songs with an attentional window shorter than the motif, as they did not need to listen to the entire motif to make a decision. Instead, they responded immediately upon discriminating a change. Since reversal discrimination performance saturated for syllables longer than 100 msec, it is likely that the attentional window is approximately 100 msec in duration.

If birds are indeed listening along a temporal window and listening to song in chunks, then perhaps we should not expect to see a BOS effect upon performance in our experiments. Birds are not listening to the motif as a whole. However, if we change the task so that birds are forced to listen to the entire motif before making a decision, then we may see differences in performance based on the identity of the song. This will also test memory constraints in the zebra finch. If birds must wait for the entire motif before responding (as in a same-different task), we expect that birds will more accurately discriminate reversals of syllables later in the motif due to a recency effect. However, we also expect that birds may be able to remember more reversals in their own songs than in conspecific songs. Since the BOS is the most familiar song to the zebra finch, it is possible that it will be able to hold changes to the BOS in memory longer than changes to conspecific songs. A Go/No-go study by Cynx (1993) showed that discrimination between two songs was only affected by removal of syllables when the song tested was the BOS. Birds completed this task by perching on different branches for the Go and No-go stimuli. This type of task differs from key pecking in that responding takes more time, allowing birds to listen to entire motifs before making a decision.

6.3.2 Local vs. global temporal sensitivity

In addition to AFP neurons exhibiting sensitivity to the bird's own song, it has also been shown that these neurons are sensitive to the temporal properties of the BOS. Solis and Doupe (1997) showed that neurons in LMAN and Area X respond with the greatest firing rates for the BOS, but respond much less to presentation of the BOS played in reverse. When songs are played in reverse, not only is the global order of syllables changed, but the local, fine timing of individual syllables is also changed. However, when the BOS is played such that the global order of syllables is reversed, but the local timing of individual syllables is maintained (termed *reverse order song*), neurons show an intermediate response. Additionally, neurons in HVC have exhibited temporal combination sensitivities such that some neurons respond best to a single syllable, some respond best to pairs of syllables, and some respond best to several syllables in a specific order (Margoliash and Fortune 1992). This suggests neuronal sensitivities to certain acoustic features of song. These preferences disappear when the song or syllable is played in reverse. Interestingly, Margoliash and Fortune (1992) also found that for the neurons that respond best to several syllables, neuronal responses remained strong even when an interval between the syllables was increased by over 200 msec.

In Experiment 1, birds were tested on two types of changes to song: single interval increases, and single syllable reversals. Results showed that all birds (regardless of gender or species) were insensitive to changes to intervals, even though intervals were doubled in duration. However, birds were quite sensitive to single syllable reversals, even for syllables that were broadband, or contained fairly flat harmonics. These perceptual responses make sense given what was seen by Margoliash and Fortune (1992). Neurons in HVC are still sensitive to the BOS even when intervals between syllables are greatly increased. Even with this change to the BOS, the song is still considered the BOS, as shown by the unchanged firing rate. However, reversing single syllables within the set of syllables the neuron is sensitive to greatly reduces the neuron's firing rate. Thus, changing the fine structure of a small part of the song reduces neuronal preferences, almost as if that change results in another song that is no longer the BOS. Perhaps changes to intervals and the overall envelope of song are not necessarily a violation of song, whereas changes to actual syllables within the song do violate the nature of song and are thus discriminated more easily. Evidence for this is also seen in a perception study by Braaten et al. (2006) in which birds were tested in a Go/No-go procedure. Birds were trained to respond to forward song, and to withhold response to reversed song. Probe trials consisted of *reverse order* songs that reversed syllable order but maintained syllables in the forward position, and *syllable reversed* songs that reversed individual syllables but maintained the syllable order. Birds treated reverse order songs like forward songs and responded to them, and they treated syllable order songs like reversed songs and withheld response. Thus, birds consider changes to local temporal structure a violation of song, whereas they do not consider changes to global temporal structure a violation of song.

Perhaps this insensitivity to temporal envelope features of song, specifically duration intervals, is due to the fact that intervals between syllables do not convey any relevant information to birds. Intervals do not give any indication of the identity of the song, which is the main piece of information males listen for in song. Likewise, intervals do not add to the complexity of song, which is the main piece of information females listen for in song. While the resulting songs of zebra finches are quite stereotyped and rhythmic, this may be due to production mechanisms and may not have anything to do with perception. All of the relevant information in song seems to be held in the fine structure of the syllables.

6.3.3 Temporal windows for listening to song

While it has been previously suggested that neurons in the anterior forebrain encode song over long temporal windows ($\sim$ 300 msec), recent work by Kojima and Doupe (2008) has suggested a shorter temporal window on the order of $\sim$ 100 milliseconds. As shown previously, neurons in the AFP respond best to the bird's own song. The same neurons respond less to reverse order song (EDCBA), and not at all to the BOS played in reverse (ƎᗡƆᗺA). This suggests something about the way in which neurons are integrating information over the course of song. Responses to reverse order song should be similar to reversed song (i.e. almost no response) if neurons were integrating over the entire song. Likewise, if neurons were integrating over single syllables, responses to reverse order song should be similar to forward song.

Thus, neurons are integrating auditory information over a set window length that includes multiple syllables.

Kojima and Doupe (2008) found that mean syllable duration of the BOS determined how sensitive neurons were to reverse order BOS. When the BOS had long syllables, neurons responded similarly to the BOS and reverse order BOS. Thus, neurons were not very selective. However, when the BOS contained shorter syllables, neurons responded much less to reverse order BOS, and were thus very selective for temporal order. The selectivity index of the neurons decreased as mean syllable duration reached 150-200 msec. Kojima and Doupe (2008) reasoned that songs with shorter syllables would have more syllables contained in the temporal window, and a difference in the temporal order of the syllables would be detected more easily. In addition, they found that the selectivity index, and acoustic similarity of forward and reverse order BOS were best correlated (negatively) for correlation windows of 100 msec.

A similar temporal window has been found behaviorally in the male zebra finches tested in Experiments 2 and 10. For both natural and Schroeder modeled songs, syllable reversal discrimination performance was positively correlated with syllable duration. In the natural song case, zebra finches showed no further improvement in syllables greater than 100 msec in duration. In the Schroeder song case, they showed no further improvement in Schroeder waveforms greater than $\sim$ 130 msec in duration. This plateau in performance suggests that birds are listening along a 100-130 msec temporal window, after which more information does not improve performance. This is particularly striking in the Schroeder song case, since waveforms do not change over time as natural song syllables do. These similar results suggest that birds perceive songs with a similar time scale that neurons in the AFP use to process songs.

6.4 Conclusions

From the experiments conducted in this dissertation, we now know more about the way in which zebra finches perceive song. When zebra finches are listening to song, changes to syllables are much more salient than changes to inter-syllable intervals. Birds are extraordinarily sensitive to reversals of syllables in song, regardless of syllable type or location within the motif. Birds listen with a window of attention of approximately 100 milliseconds, the average length of a syllable. Discrimination performance is duration dependent for syllables shorter than 100 milliseconds. Birds are sensitive to envelope features of syllables, particularly when fine structure and spectral cues are not available. Temporal fine structure information alone is sufficient for birds to discriminate syllable-like stimuli in a song motif. In sum, the correct perception/identification of syllables in song can be maintained by different cues, but perception is most robust when all cues are present in song. This is reminiscent of human speech, in which multiple redundant cues are used for recognition of phonemes.

Appendix A

Supplementary Tables

Table A.1: Results of Mann-Whitney Rank Sum Test comparing performance on single interval doublings and single syllable reversals in Experiment 1

Song	**Median Int**	**Median Rev**	**U**	**Sample Size**	**p**
MoonBOS	0.00	100.00	1200.00	$n_1 = 30$; $n_2 = 40$	< 0.001
JulepBOS	5.00	95.00	936.50	$n_1 = 27$; $n_2 = 36$	< 0.001
BearBOS	0.00	100.00	1200.00	$n_1 = 30$; $n_2 = 40$	< 0.001
ScotchBOS	2.50	100.00	1200.00	$n_1 = 30$; $n_2 = 40$	< 0.001

Table A.2: Results of Kruskal-Wallis ANOVA for BOS latency analysis in Experiment 2

Bird	**Song**	**Mean Latency**	**Median Latency**	**H**	**p**	**Dunn's $p < 0.05$**
Moonshine	MoonBOS	722.04 ± 269.64	637.00	27.967	< 0.001	BearBOS vs.MoonBOS
	JulepBOS	844.03 ± 524.60	607.5			BearBOS vs. JulepBOS
	BearBOS	584.13 ± 103.64	568.00			
	ScotchBOS	763.81 ± 498.34	548.00			
Julep	MoonBOS	806.05 ± 385.55	687.00	19.487	< 0.001	BearBOS vs. MoonBOS
	JuleBOS	782.32 ± 414.27	647.00			
	BearBOS	657.39 ± 154.39	630.00			
	ScotchBOS	775.88 ± 449.96	635.00			
Bear	MoonBOS	1054.58 ± 425.87	888.50	44.058	< 0.001	BearBOS vs.MoonBOS
	JulepBOS	1399.89 ± 647.35	1125.00			BearBOS vs. JulepBOS
	BearBOS	857.58 ± 148.91	826.00			BearBOS vs. ScotchBOS
	ScotchBOS	1073.25 ± 437.22	884.00			
Scotch	MoonBOS	777.76 ± 234.90	713.00	106.324	< 0.001	BearBOS vs. MoonBOS
	JulepBOS	777.93 ± 349.40	688.50			BearBOS vs. JulepBOS
	BearBOS	674.53 ± 138.43	648.00			BearBOS vs. ScotchBOS
	ScotchBOS	970.14 ± 475.70	794.50			

Table A.3: Results of Mann-Whitney Rank Sum Test comparing performance in Experiments 2 and 7

Group	**Median Natural**	**Median Rand**	**U**	**Sample Size**	**p**
Males	100.00	85.00	1630.000	$n_1 = n_2 = 88$	< 0.001
Females	100.00	85.00	962.00	$n_1 = n_2 = 66$	< 0.001
Bud	100.00	90.00	942.50	$n_1 = n_2 = 66$	< 0.001

Table A.4: D-prime scores and standard error (se) for zebra finch performance on single interval doublings and single syllable reversals in natural and random noise songs. Standard error calculations were based on the number of trials for that individual target. All targets used 20 trials.

Zebra finches	**Natural**		**Rand**	
Target	**d′**	**se**	**d′**	**se**
Moon int1	0.07	0.32	0.30	0.32
Moon int3	0.05	0.32	0.33	0.33
Moon int5	0.12	0.32	0.25	0.32
Moon A	3.68	0.63	1.97	0.39
Moon C	3.24	0.54	1.76	0.37
Moon E	3.60	0.61	1.29	0.35
Moon F	3.84	0.67	2.86	0.48
Bear int1	0.21	0.32	0.04	0.32
Bear int2	0.22	0.32	0.18	0.32
Bear int4	0.33	0.33	0.09	0.32
Bear A	4.10	0.74	3.43	0.57
Bear B	4.02	0.72	3.35	0.56
Bear D	3.77	0.65	2.51	0.44
Bear E	3.93	0.70	3.22	0.53
Scotch int1	0.19	0.32	0.05	0.32
Scotch int4	0.41	0.33	0.20	0.32
Scotch int6	0.27	0.32	0.01	0.32
Scotch A	3.03	0.50	1.42	0.36

Continued on next page

Zebra finches	Natural		Rand	
Target	**d′**	**se**	**d′**	**se**
Scotch C	3.60	0.61	2.53	0.44
Scotch E	3.31	0.55	2.11	0.40
Scotch G	3.68	0.63	2.73	0.46

Table A.5: D-prime scores and standard error (se) for zebra finch performance on single Schroeder reversals in Schroeder songs presented at 60 and 80 dB SPL. Standard error calculations were based on the number of trials for that individual target. All targets used 20 trials, except Julep allrev and Bear allrev, which used 40 trials (as they accounted for 2 targets in the set).

Zebra finches	60 dB		80 dB	
Schroeder syllable	**d′**	**se**	**d′**	**se**
Moon A	2.20	0.41	2.84	0.48
Moon B	2.14	0.40	3.02	0.50
Moon C	2.62	0.45	3.53	0.60
Moon D	3.73	0.64	3.53	0.60
Moon E	2.77	0.47	2.84	0.48
Moon F	3.53	0.60	3.73	0.64
Moon allrev	3.73	0.64	3.73	0.64
Julep A	2.40	0.42	2.36	0.42
Julep B	1.99	0.39	2.32	0.42
Julep C	3.31	0.55	3.24	0.54
Julep D	2.75	0.46	3.05	0.51

Continued on next page

Zebra finches	**60 dB**		**80 dB**	
Schroeder syllable	**d′**	**se**	**d′**	**se**
Julep E	1.59	0.36	1.46	0.36
Julep allrev	3.79	0.52	3.58	0.48
Bear A	3.82	0.66	4.02	0.72
Bear B	3.82	0.66	4.02	0.72
Bear C	2.64	0.45	4.02	0.72
Bear D	2.91	0.49	4.02	0.72
Bear E	3.42	0.57	3.82	0.66
Bear allrev	4.29	0.64	4.16	0.60
Scotch A	4.25	0.80	3.01	0.50
Scotch C	4.74	1.01	3.61	0.61
Scotch D	3.77	0.65	3.82	0.66
Scotch E	3.80	0.66	3.61	0.61
Scotch F	3.87	0.68	3.82	0.66
Scotch G	4.11	0.75	3.61	0.61
Scotch allrev	4.46	0.88	3.61	0.61

Table A.6: D-prime scores and standard error (se) for budgerigar performance on single Schroeder reversals in Schroeder songs presented at 60 and 80 dB SPL. Standard error calculations were based on the number of trials for that individual target. All targets used 20 trials, except Julep allrev and Bear allrev, which used 40 trials (as they accounted for 2 targets in the set).

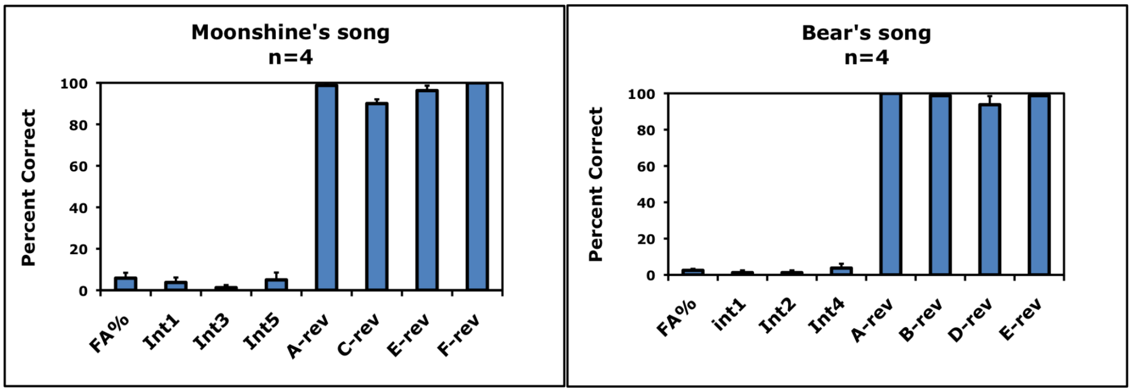
Budgerigars	**60 dB**		**80 dB**	
Schroeder syllable	**d′**	**se**	**d′**	**se**
Moon A	1.68	0.37	3.18	0.53
Moon B	0.88	0.34	3.18	0.53
Moon C	1.26	0.35	3.38	0.56
Moon D	1.68	0.37	3.38	0.56
Moon E	0.68	0.33	2.77	0.47
Moon F	1.68	0.37	3.38	0.56
Moon allrev	2.67	0.45	3.38	0.56
Julep A	1.05	0.34	3.70	0.63
Julep B	1.30	0.35	3.90	0.68
Julep C	2.51	0.44	3.90	0.68
Julep D	1.36	0.35	3.52	0.59
Julep E	0.01	0.32	2.11	0.40
Julep allrev	2.78	0.37	4.18	0.61
Bear A	2.51	0.44	3.31	0.55
Bear B	1.80	0.38	3.31	0.55
Bear C	1.10	0.34	2.80	0.47
Bear D	1.61	0.37	2.80	0.47

Continued on next page

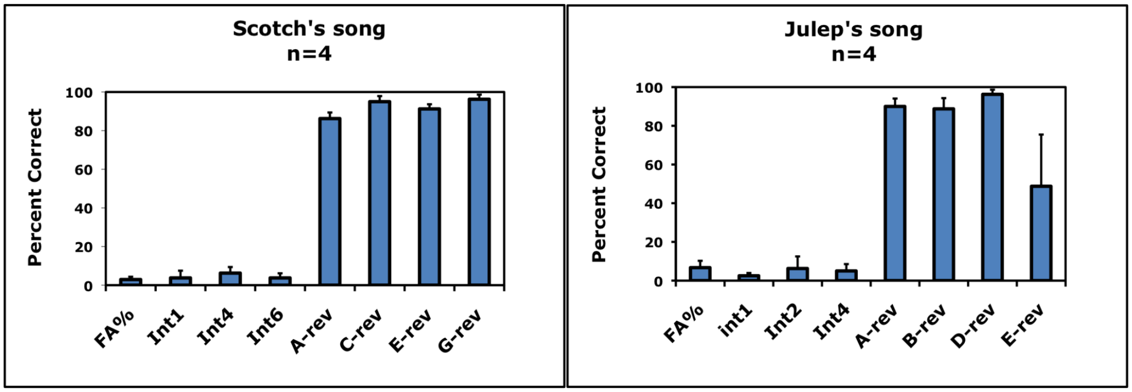
Budgerigars	**60 dB**		**80 dB**	
Schroeder syllable	**d′**	**se**	**d′**	**se**
Bear E	1.92	0.38	3.31	0.55
Bear allrev	2.64	0.36	3.58	0.48
Scotch A	1.00	0.34	3.14	0.52
Scotch C	1.96	0.39	3.14	0.52
Scotch D	1.13	0.34	3.14	0.52
Scotch E	0.53	0.33	3.14	0.52
Scotch F	1.05	0.34	3.34	0.55
Scotch G	1.31	0.35	3.14	0.52
Scotch allrev	2.65	0.45	3.34	.55

Appendix B

Supplementary Figures

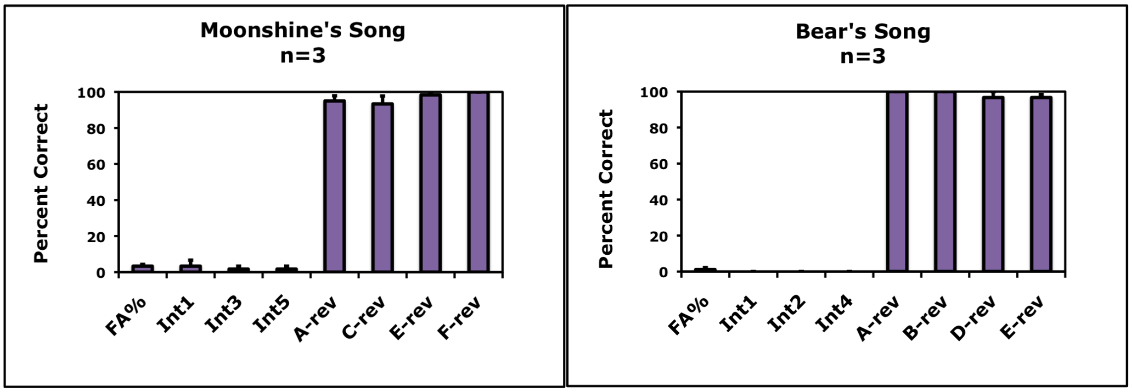


(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

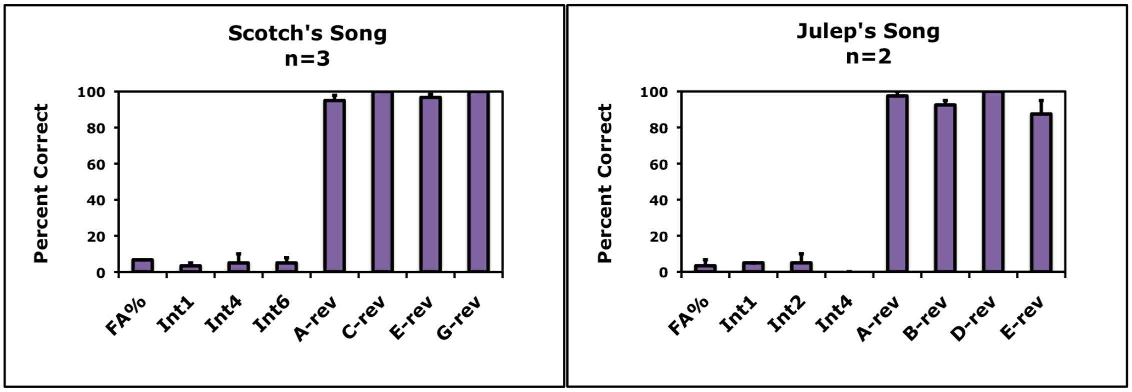


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure B.1: Average discrimination performance of male zebra finches on single interval doublings and single syllable reversals in natural song, presented in the same testing session. Error bars show standard error of the mean.

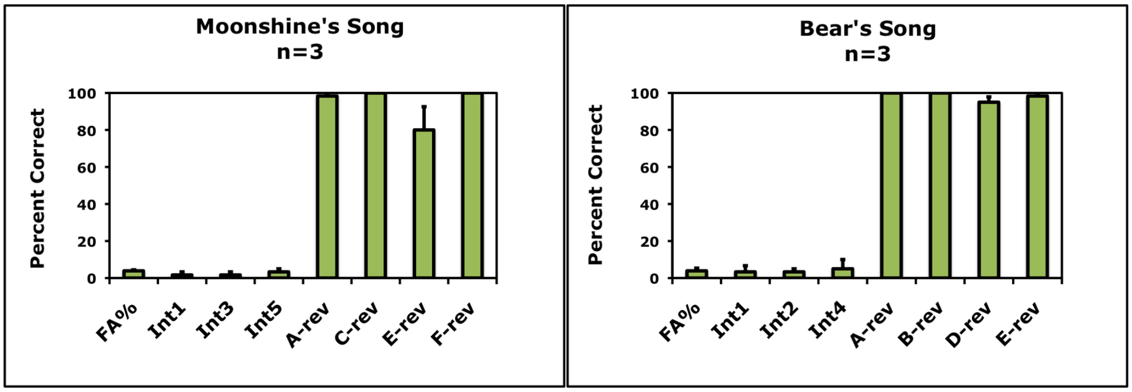


(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif

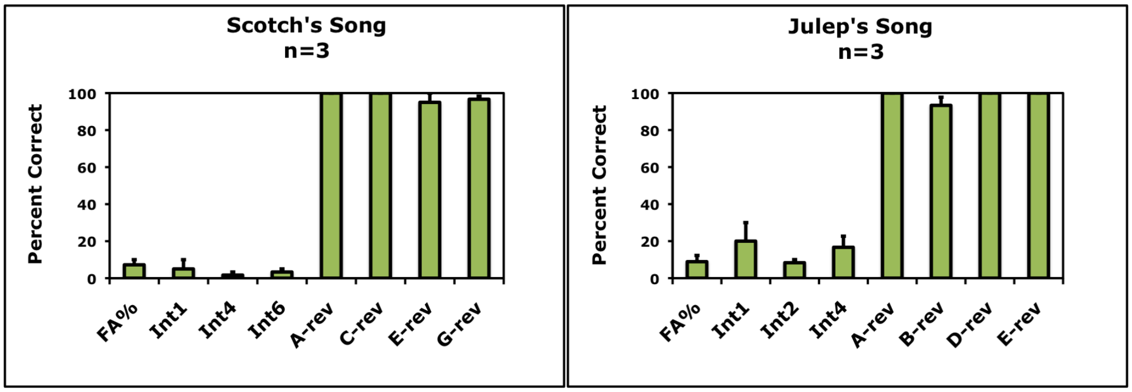


(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure B.2: Average discrimination performance of female zebra finches on single interval doublings and single syllable reversals in natural song, presented in the same testing session. Error bars show standard error of the mean.



(a) Discrimination performance: Moon's motif (b) Discrimination performance: Bear's motif



(c) Discrimination performance: Scotch's motif (d) Discrimination performance: Julep's motif

Figure B.3: Average discrimination performance of budgerigars on single interval doublings and single syllable reversals in natural song, presented in the same testing session. Error bars show standard error of the mean.

Appendix C

Abbreviations

AFP anterior forebrain pathway

BOS bird's own song

DLM medial portion of the dorsolateral nucleus of the anterior thalamus

EM energetic masking

dph days post hatch

HVC$_{\mathbf{RA}}$ RA-projecting HVC neurons

HVC$_{\mathbf{X}}$ area X-projecting HVC neurons

IM informational masking

LC long calls

LMAN lateral magnocellular nucleus of the anterior nidopallium

Nif nucleus interfacialis

NXIIts nerve XII, Tracheosyringeal nerve

RA	robust nucleus of the archopallium
Uva	uvaeformis

Bibliography

Abel, S. (1972). Discrimination of temporal gaps. *J. Acoust. Soc. Am.*, 52:519–524.

Amin, N., Doupe, A., and Theunissen, F. (2007). Development of selectivity for natural sounds in the songbird auditory forebrain. *J Neurophysiol*, 97:3517–3531.

Boogert, N., Giraldeau, L., and Lefebvre, L. (2008). Song complexity correlates with learning ability in male zebra finches. *Anim. Behav.*, 76:1735–1741.

Bottjer, S., Miesner, E., and Arnold, A. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, 224:901–903.

Braaten, R. F., Petzoldt, M., and Colbath, A. (2006). Song perception during the sensitive period of song learning in zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, 120:79–88.

Brainard, M. and Doupe, A. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature*, 404:762–766.

Brainard, M. and Doupe, A. (2001). Postlearning consolidation of birdsong: stabilizing effects of age and anterior forebrain lesions. *J. Neurosci.*, 21:2501–2517.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press.

Chi, Z. and Margoliash, D. (2001). Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron*, 32(5):889–910.

Cooper, B. and Goller, F. (2006). Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *J. Neurophysiol.*, 95(6):3798–3809.

Cynx, J. (1993). Conspecific song perception in zebra finches (*Taeniopygia guttata*). *J Comp Psychol.*, 107(4):395–402.

DeVoogd, T. (2004). Neural constraints on the complexity of avian song. *Brain, Behavior, and Evolution*, 63:221–232.

Dooling, R. and Haskell, R. (1978). Auditory duration discrimination in the parakeet (*Meopsittacus undulatus*). *J. Acoust. Soc. Am.*, 63(5):1640–1642.

Dooling, R., Leek, M., Gleich, O., and Dent, M. (2002). Auditory temporal resolution in birds: discrimination of harmonic complexes. *J. Acoust. Soc. Am.*, 112(2):748–759.

Dooling, R. and Searcy, M. (1980). Forward and backward auditory masking in the parakeet (*Melopsittacus undulatus*). *Hearing Research*, 3:279–284.

Doupe, A. (1997). Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.*, 17(3):1147–1167.

Doupe, A. and Solis, M. (1997). Song- and order-selective neurons develop in the songbird anterior forebrain during vocal learning. *J. Neurobiol.*, 33(5):694–709.

Drullman, R. (1994). Temporal envelope and fine structure cues for speech intelligibility. *J. Acoust. Soc. Am.*, 97(1):585–591.

Eales, L. (1985). Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Animal Behaviour*, 33:1293–1300.

Farries, M. and Perkel, D. (2002). A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. *J. Neurophysiol.*, 22(9):3776–3787.

Floody, O. and Arnold, A. (1997). Song lateralization in the zebra finch. *Hormones* *and Behavior*, 31:25–34.

Forstmeier, W., Burger, C., Temnow, K., and Deregnaucourt, S. (2009). The genetic basis of zebra finch vocalizations. *Evolution*, 63(8):2114–2130.

Glaze, C. and Troyer, T. (2006). Temporal structure in zebra finch song: implications for motor coding. *J. Neurosci.*, 26:991–10051.

Goller, F. and Daley, M. (2001). Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. *Proc. Biol. Sci.*, 268(1483):2301–2305.

Goller, F. and Suthers, R. (1996). Control of song phonology by syringeal muscles. *J. Neurophysiol.*, 76(1):287–300.

Hahnloser, R., Kozhevnikov, A., and Fee, M. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419:65–70.

Immelmann, K. (1969). Song development in the zebra finch and other estrilid finches. In Hinde, R., editor, *Bird vocalizations*, pages 61–74. Cambridge University Press.

Jacobsen, T., Schroeger, E., Winkler, I., and Horvath, J. (2005). Familiarity affects the processing of task-irrelevant auditory deviance. *J. Cog. Neurosci.*, 17(11):1704–1713.

Kao, M. and Brainard, M. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J. Neurophysiol*, 96:1441–1455.

Kidd, G., Mason, C., Richards, V., Gallun, F., and Durlach, N. (2007). Informational masking. In Yost, W., editor, *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources.*, volume 29, pages 143–189. Springer US, New York.

Kinney, J. (1961). Discrimination in auditory and visual patterns. *Am. J. Psychol.*, 74:529–541.

Kittelberger, J. and Mooney, R. (1999). Lesions of an avian forebrain nucleus that disrupt song development altar synaptic connectivity and transmission in the vocal premotor pathway. *J. Neurosci*, 19(21):9385–9398.

Kojima, S. and Doupe, A. (2008). Neural encoding of auditory temporal context in a songbird basal ganglia nucleus and its independence of birds' song experience. *European Journal of Neuroscience*, 27:1231–1244.

Kong, Y. and Zeng, F. (2006). Temporal and spectral cues in mandarin tone recognition. *J. Acoust. Soc. Am.*, 120(5):2830–2840.

Lauer, A., Dooling, R., Leek, M., and Lentz, J. (2006). Phase effects in masking by harmonic complexes in birds. *J. Acoust. Soc. Am.*, 119(2):1251–1259.

Leek, M., Brown, M., and Dorman, M. (1991). Informational masking and auditory attention. *Perception & Psycophysics*, 50(3):205–214.

Leek, M. and Watson, C. (1984). Learning to detect auditory pattern components. *J. Acoust. Soc. Am.*, 76(4):1037–1044.

Lei, H. and Mooney, R. (2010). Manipulation of a central auditory representation shapes learned vocal output. *Neuron*, 65(17):122–134.

Liu, W., Gardner, T., and Nottebohm, F. (2004). Juvenile zebra finches can use multiple strategies to learn the same song. *PNAS*, 101(52):18177–18182.

Lohr, B. and Dooling, R. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J. Comp. Psychol.*, 112(1):36–47.

Lohr, B., Dooling, R., and Bartone, S. (2006). The discrimination of temporal fine structure in call-like harmonic sounds by birds. *J. Comp. Psychol.*, 120(3):239–51.

Luo, M., Ding, L., and Perkel, D. (2004). An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. *J Neurosci*, 21:6836–6845.

MacDougall-Shackleton, S., Hulse, S., Gentner, T., and White, W. (1998). Auditory scene analysis by european starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *J. Acoust. Soc. Am.*, 103(6):3581–3587.

Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: A User's Guide.* Lawrence Erlbaum Associates, second edition.

Margoliash, D. and Fortune, E. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J. Neurosci.*, 12(11):4309–4326.

Moore, B. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *JARO*, 9:399–406.

Morrison, R. and Nottebohm, F. (1993). Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. *J. Neurobiol.*, 24:1045–1064.

Nelson, R. and Palmer, S. (2007). Familiar shapes attract attention in figure-ground displays. *Perception & Psychophysics*, 69(3):382–392.

Nespor, A. and Dooling, R. (1997). Discrimination among natural and altered motifs of the song of the zebra finch (*Taeniopygia guttata*): A comparative study. *Bird Behavior*, 12:15–28.

Nick, T. and Konishi, M. (2005). Neural song preference during vocal learning in the zebra finch depends on age and state. *J. Neurobiol*, 62(2):231–242.

Nordeen, K. and Nordeen, E. (1992). Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.*, 57:58–66.

Nordeen, K. and Nordeen, E. (1993). Long-term maintenance of song in adult zebra finches is not affected by lesions of a forebrain region involved in song learning. *Behav. Neural Biol.*, 59:79–82.

Ohms, V., Gill, A., Van Heijningen, C., Beckers, G., and ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proc.*

R. Soc. B., 277:1003–1009.

Okanoya, K. and Dooling, R. (1990). Temporal integration in zebra finches (*Poephila guttata*). *J. Acoust. Soc. Am.*, 87(6):2782–2784.

Okanoya, K. and Yamaguchi, A. (1997). Adult Bengalese finches (*Lonchura striata var. domestica*) require real-time auditory feedback to produce normal song syntax. *J. Neurobiol.*, 33(4):343–356.

Olveczky, B., Andalman, A., and Fee, M. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLOS Biol.*, 3(5):e153.

Paquette, C. and Peretz, I. (1997). Role of familiarity in auditory discrimination of musical instrument: a laterality study. *Cortex*, 33.

Prather, J., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451(7176):305–310.

Price, P. (1979). Developmental determinants of structure in zebra finch song. *J. Comp. Physiol. Psychol.*, 93:268–277.

Riebel, K. and Smallegange, I. (2003). Does zebra finch (*Taeniopygia guttata*) preference for the (familiar) father's song generalize to the songs of unfamiliar brothers? *J. Comp. Psychol.*, 117(1):61–66.

Scharff, C. and Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning. *J. Neurosci.*, 11:2896–2913.

Schroeder, M. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans. Information Theory*, 16(1):85–89.

Simpson, H. and Vicario, D. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.*, 10(5):1541–1556.

Solis, M. and Doupe, A. (1997). Anterior forebrain neurons develop selectivity by an intermediate stage of birdsong learning. *J. Neurosci.*, 17(16):6447–6462.

Sossinka, R. and Böhner, J. (1980). Song types in the zebra finch (*Poephila guttata castanottis*). *Z. Tierpsychol.*, 17:123–132.

Spencer, K., Buchanan, K., Goldsmith, A., and Catchpole, C. (2003). Song as an honest signal of developmental stress in the zebra finch (*Taeniopygia guttata*). *Hormones and Behavior*, 44:132–139.

Theunissen, F. and Doupe, A. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *J. Neurosci.*, 18(10):3786–3802.

Vicario, D., Naqvi, N., and Raskin, J. (2001a). Behavioral discrimination of sexually dimorphic calls by male zebra finches requires an intact vocal motor pathway. *J. Neurobiol.*, 47:109–120.

Vicario, D., Naqvi, N., and Raskin, J. (2001b). Sex differences in discrimination of vocal communication signals in a songbird. *Animal Behaviour*, 61:805–817.

Vicario, D., Raskin, J., Naqvi, N., Thande, N., and Simpson, H. (2002). The relationship between perception and production in songbird vocal imitation: what learned calls can teach us. *J. Comp. Physiol. A*, 188:897–908.

Watson, C. (2005). Some comments on informational masking. *Acta Acustica United with Acustica*, 91(3):502–512.

Watson, C., Kelly, W., and Wroton, H. (1976). Factors in the discrimination of tonal patterns. ii. selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America*, 60(5):1176–1186.

Watson, C., Wroton, H., Kelly, W., and Benbassat, C. (1975). Factors in the discrimination of tonal patterns. i. component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America*, 57(5):1175–1185.

Wild, J., Goller, F., and Suthers, R. (1998). Inspiratory muscle activity during bird song. *J. Neurobiol.*, 36(3):441–453.

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (taeniopygia guttata). *J Exp. Biol.*, 204(20):3497–3506.

Williams, H., Crane, L., Hale, T., Esposito, M., and Nottebohm, F. (1992). Right-side dominance for song control in the zebra finch. *J. Neurobiol.*, 23(8):1006–1020.

Williams, H., Cynx, J., and Nottebohm, F. (1989). Timbre control in zebra finch *Taeniopygia guttata* song syllables. *J. Comp. Psychol.*, 103(4):366–380.

Zann, R. (1984). Structural variation of the zebra finch distance call. *Z. Tierpsychol.*, 66:328–345.

Zann, R. (1985). Ontogeny of the zebra finch distance call. i. effects of cross-fostering to bengalese finches. *Z. Tierpsychol.*, 68:1–23.