Title
Behavioral and neural correlates of song tempo in the Bengalese finch

Author
Merfeld, Emily

Publication Date
2021

Peer reviewed|Thesis/dissertation

Behavioral and neural correlates of song tempo in the Bengalese finch

by
Emily Merfeld

THESIS
Submitted in partial satisfaction of the requirements for degree of
MASTER OF SCIENCE

in

Neuroscience

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Evan Feinberg
4542BD2C6E14414...

Evan Feinberg
Chair

DocuSigned by:
DEVANAND MANOLI
[illegible]

DEVANAND MANOLI

DocuSigned by:
Massimo Scanziani
[illegible]

Massimo Scanziani

DocuSigned by:
Kevin Yackle
863799D3911346A...

Kevin Yackle

Committee Members

Acknowledgments

I am deeply grateful for the opportunity to work alongside the amazingly talented people in the Brainard lab. Michael Brainard has provided invaluable support and feedback and has helped me become a better scientist. His infectious curiosity to understand and characterize the world, and his expectation of high rigor and good work, has been inspiring during my time in the lab, and will continue to inspire me throughout my career. I am so thankful for all the help that everyone in lab has provided, with everything from experimental design to troubleshooting hardware to analysis. Many thanks to each member of my committee – Evan Feinberg, Devanand Manoli, Massimo Scanziani, and Kevin Yackle - for providing valuable feedback on drafts of this thesis.

I also want to thank my friends and family for supporting me over these past years. In particular, life has been so much more joyful due to morning walks at Mori Point with Lily Ben-Avi and city bike rides with Vikram Chan-Herur. Finally, thank you to my little family here in San Francisco – Adam Bercu, Annie, and Luna – you three are truly lights of my life.

Behavioral and neural correlates of song tempo in the Bengalese finch

Emily Merfeld

Abstract

Song tempo in male Bengalese finches is highly stereotyped within-bird but variable across the population, and is an important courtship signal that guides mate choice. Here, the value of song tempo for communicating other individual-specific behavioral traits was examined. The calling behavior of male Bengalese finches was examined by playing a distance call of a male conspecific through a speaker and quantifying the bird's call response latency and the number and frequency of call responses. Song tempo was found to be anticorrelated with call response latency, and positively correlated with both the percent of calls that were responded to and the number of call responses per trial. In females, who do not sing, calling behaviors in the call playback paradigm were not associated with the sire's song tempo and no within-nest clustering of behavior was observed, suggesting that female calling behavior is not dependent on genetic background. However, there was strong within-bird clustering of call response features, such that females with low latency tended to produce many calls per trial and responded to most trials (and vice versa). In a separate set of experiments, electrophysiological data was recorded from the sensorimotor nucleus HVC while an adult male Bengalese finch listened to a variety of stimuli presented at different tempos. A tuning curve centered at a behaviorally-relevant tempo was identified. These findings suggest that song tempo may provide valuable information to potential mates about a suite of other behaviors, and the potential for individual-specific HVC dynamics to guide song learning and calling behaviors is discussed.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: Introduction

Bengalese finches are highly social organisms who produce two types of vocalizations: songs and calls. Songs are highly stereotyped and complex vocalizations produced by the male, while calls are shorter and simpler vocalizations produced by both sexes. In a variety of songbirds including Bengalese finches, vocal exchanges are particularly important for the formation and maintenance of socially monogamous pair bonds (Bolund et al., 2012; d'Amelio et al., 2017A; Riebel, 2009; Elie et al., 2010; Gill et al., 2015; Perez et al., 2015; Prior et al., 2020).

Songs are sufficient to guide mate choice (Riebel, 2009), and therefore are expected - like other sexual signals - to provide honest indicators of the singer's fitness as a mate (Dougherty, 2021). Signals that reflect the fitness of the producing organism are expected to be stereotyped within-bird but variable across the population; this allows the receiving organism to use inter-individual differences in the signal to select a mate. Both songs and calls fit this requirement, and each type of vocalization is sufficient for identification of the vocalizer (Geberzahn and Deregnaucourt, 2020; d'Amelio et al., 2017B; Elie and Theunissen, 2018).

What may vocalizations communicate about the fitness of a bird? Some prior literature has directly linked song features to fitness (Gil and Gahr, 2002). For instance, the preference of female birds for complex song over simple song (Nowicki and Searcy, 2004) has been linked to observations that developmental stress reduces song complexity (e.g., Nowicki et al., 2000; Zann and Cash, 2008). However, other studies with conflicting results (e.g., Gil et al., 2006; Brumm et al., 2009) make it difficult to determine whether features such as song complexity reliably reflect the fitness of the singer. An alternative theory is that vocalizations hold information about the compatibility of the courting bird with the courted bird. Behavioral synchrony - the phenomenon in which animals engage in the same task or coordinate their behavior to benefit the pair or group

- has been linked to reproductive success across several bird species (Griffith, 2019; Roth et al., 2021). For example, assortative pairing according to personality traits (e.g., aggression or exploratory drive) is linked to reproductive success in Eastern bluebirds (Harris and Siefferman, 2014), great tits (Dingemanse et al., 2004; Both et al., 2005), and zebra finches (Schuett et al., 2011). Intriguingly, the temporal coordination of magpie-lark duet increases during pair bonding, and more coordinated songs are more effective for territorial defense (Hall and Magrath, 2007). Although estrildid finches - such as zebra finches and Bengalese finches - do not duet, they do exchange precisely-timed antiphonal calls (Benichov et al., 2016; Benichov and Vallentin, 2020; Ma et al., 2020). The temporal synchrony of these exchanges increases with the duration of the pair bond in reproductively successful but not reproductively unsuccessful zebra finch pairs (Gill et al., 2015). Together, these findings suggest that selecting a compatible mate is linked to reproductive success, and that a bird may be motivated to select a mate with whom temporal vocal synchrony could be easily achieved. This leads to my first central hypothesis: song - the main courtship signal of the Bengalese finch - is an honest signal that sends information to the female about the call timing of the male, thereby allowing her to assess the vocal compatibility of the singer.

A correlation between song tempo and call timing may be underlaid by common neural circuitry at the level of the sensorimotor nucleus HVC. HVC auditory activity is precisely aligned to particular moments in song (Hahnloser et al., 2002) and artificial (Long and Fee, 2008) or natural (Aronov and Fee, 2012) changes in HVC activity are linked to changes in song tempo. Intriguingly, the same cells that are active during song production are also active during the production of precisely-timed calls (Benichov and Vallentin, 2020). Moreover, in males, HVC disinhibition reduces call response latency and increases call timing precision, whereas

increasing HVC inhibition has the opposite effect (Benichov and Vallentin, 2020). Whether altering HVC activity also leads to changes in call timing in females has not been examined. However, in both sexes, call timing precision is reduced drastically by the ablation of the robust nucleus of the arcopallium, which is directly downstream of HVC (Benichov et al., 2016). In males, HVC also produces activity prior to the onset of a female partner's call, and this "predictive" activity is necessary for the auditory-evoked HVC response (Ma et al., 2020). These findings suggest that rather than dictating song timing specifically, HVC exerts control over multiple types of precisely-timed vocalizations.

Song tempo arises from a combination of genetic and experience-dependent factors (Mets and Brainard, 2018; Mets and Brainard, 2019). Recent work has identified a putative mechanism by which innate HVC properties bias males to sing at a particular tempo. Intra-HVC expression of a zinc transporter protein - which, based on *in vitro* experiments, appears to increase neuronal excitability - correlates with song tempo. Moreover, artificially reducing intrasynaptic zinc levels - thereby mimicking the action of the zinc transporter - increases song tempo, while knocking down the zinc transporter reduces song tempo (Mets et al., 2021). Since activity in HVC is causally linked to song timing (Long and Fee, 2008) and call timing (Benichov and Vallentin, 2020), innate differences in HVC excitability may give rise to individual-specific song timing and call timing. Together, these findings lead to my second central hypothesis: HVC is innately tuned to produce and encode songs and calls with particular tempos.

Chapter 2 examines whether song tempo predicts calling behavior in male Bengalese finches. I hypothesized that males with higher song tempos would have shorter call response latencies and higher call rates. Indeed, in two cohorts of males, song tempo was anticorrelated with the latency to respond to a call played through a speaker and the probability of call response

per trial. Also in this chapter, I examined whether calling behavior in the female Bengalese finch can be predicted from the sire's song tempo. Although females do not sing, the sire's song tempo should reflect genetic and experience-dependent contributions to the female's behavior and underlying neural circuitry. I predicted that females with fast-singing fathers should respond more quickly and more often in the call playback paradigm. Although calling behavior was stereotyped within-bird and variable across the population, there was no relationship between sire song tempo and calling behavior, and no within-nest clustering of calling behavior was observed. Finally, preliminary results from untutored females are discussed.

Chapter 3 uses a larger sampling of nests and birds to further examine whether call features cluster within-nest in female Bengalese finches. I hypothesized that two timing-related features of spontaneous calls - inter-lobe interval and inter-call interval - would cluster within-nest, reflecting learned and genetic contributions to these behaviors. I found that although each female's inter-lobe interval and inter-call interval was stereotyped and individual-specific, these features did not cluster within-nest. Taken together, the results from females in Chapters 2 and 3 suggest that female calling behavior is highly idiosyncratic and largely independent of genetic background.

Chapter 4 examines whether HVC auditory activity is tuned to encode behaviorally-relevant tempos. In one adult male Bengalese finch, HVC was sensitive to tempo manipulations of the bird's own song. Moreover, playbacks of spectrally neutral white noise "syllables" played at various tempos revealed a tuning curve centered at 4-5 elements per second. This tempo is similar to the tempo of the introductory notes of the bird's song. These preliminary results lay the groundwork for future experiments that may characterize HVC tempo tuning curves in a wider sample of juvenile and adult males.

Taken together, these results suggest that a male's song tempo communicates important information about calling behavior, and that this tempo may be supported by individual-specific HVC dynamics.

CHAPTER 2: Call features are stable within-bird phenotypes that are well-predicted by song tempo in males

2.1 Results

2.1.A. Call response features are correlated with song tempo in naturally-reared male Bengalese finches.

The latency to respond to a call stimulus is stereotyped within-bird but variable across-bird (Benichov et al., 2016). Song tempo is similarly stereotyped within-bird but variable across-bird due to a combination of genetic and environmental factors (Mets and Brainard, 2018). Both call timing and song tempo are mediated by an interplay between inhibitory interneurons and excitatory projection neurons in HVC (Benichov and Vallentin, 2020; Kosche et al., 2015), and disrupting HVC activity disrupts both call timing and song timing (Benichov et al., 2016; Benichov and Vallentin, 2020; Long and Fee, 2008; Jaffe and Brainard, 2020). Moreover, the same HVC cells are active during both call responses and during song production (Benichov and Vallentin, 2020). Recent work suggests that birds with faster song tempos may have a more excitable HVC (Mets et al., 2021). Whether an innately more excitable HVC also underlies a lower call response latency is unknown, but artificially disinhibiting HVC does reduce call latency (Benichov and Vallentin, 2020). Given these findings, I hypothesized that - due to a reliance on shared circuitry - call latency and song tempo are anticorrelated in male Bengalese finches.

To determine whether call response latency is anticorrelated with song tempo, adult male Bengalese finches were moved from the colony room to a sound isolation chamber in which they were singly-housed overnight. The next day, each bird was moved to a testing sound isolation chamber outfitted with a cage, microphone, and speaker. Each bird was habituated to the

chamber for 10 minutes. After habituation, the distance call of an adult male conspecific was played through a speaker every 5-11 seconds (determined pseudorandomly for each trial), or 2 seconds after the bird stopped singing, calling, or making other loud sounds, whichever came later. After this session, the bird was returned to its original sound isolation chamber. Four hours later, a second habituation session and call playback session was conducted; this second session allowed me to determine whether various call parameters were stable within-bird across multiple sessions (Figure 2.1 A, see Methods, Experiment 1).

Birds responded with single calls, call trains, or songs to the call stimuli (Figure 2.1 B). These responses were clearly elicited rather than spontaneous, as most birds (8 out of 10) did not vocalize at all during habituation, and the two birds that vocalized during habituation did so in fewer than 10% of audio files (data not shown). During the playback session, each bird responded with a stereotypic latency, and the distribution of latencies for each bird had a clearly identifiable peak (Figures 2.1 C-D). The latency to return the call was stable across sessions (Figure 2.1 E; see figure legend for statistics). Percent successful trials (i.e., the percent of trials in which the bird produced at least one vocalization) for each bird in Session 1 was positively correlated with the percent successful trials in Session 2 (Figure 2.1 F, left panel). However, as a population, birds responded to more trials in Session 1 than in Session 2 (Figure 2.1 F, right panel). Similarly, there was a positive correlation between each bird's mean number of vocalizations in Session 1 vs. Session 2 (Figure 2.1 G, left panel) although there were significantly fewer call responses during Session 2 across the population (Figure 2.1 G, right panel). These results suggest that the relative value of these three quantified metrics - latency to respond to the call stimulus, percent successful trials, and the mean number of vocalizations per successful trial - are stable within-bird phenotypes that have a wide range across the population.

It appears that these metrics may shift across the population due to factors such as the novelty of the paradigm. These observations are reminiscent of song tempo, which is highly stereotyped within-bird but reliably increases in certain contexts such as during courtship (Cooper and Goller, 2006; Sossinka and Böhner, 1980; Sakata et al., 2008).

Given that these call metrics and song tempo are both highly stable within-bird and that calls and song rely on shared neural circuitry, I next examined whether call latency, percent successful trials, or the mean number vocalizations per successful trial correlated with song tempo. Indeed, latency to return the call was negatively correlated with song tempo (Figure 2.2 A) and percent successful trials was positively correlated with song tempo (Figure 2.2 B). There was a non-significant positive correlation ($p = 0.054$) between mean number of vocalizations per successful trial and song tempo (Figure 2.2 C).

Male Bengalese finches from the same nest tend to have similar song tempos as they are all tutored by their father and share genetic backgrounds (Mets and Brainard, 2018). As expected, mean song tempo clustered within-nest ($p = 0.0022$, one-tailed t-test). Latency to respond also clustered within-nest (Figure 2.2 D). Within-nest clustering of percent successful trials did not reach significance ($p = 0.076$, one-tailed t-test, Figure 2.2 E), and the mean number of vocalizations per successful trial did not cluster within-nest (Figure 2.2 F).

These call features also correlated with one another. In particular, call response latency was significantly anticorrelated with percent successful trials (Figure 2.3 A). There was a non-significant ($p = 0.095$) anticorrelation between the mean number of responses per successful trial and call response latency (Figure 2.3 B). Additionally, percent successful trials and the mean number of responses per successful trial were strongly correlated with one another (Figure 2.3

C). Thus, fast-singing birds tend to respond with short latency and many vocalizations in many trials (and vice versa).

This first experiment (hereon referred to as Experiment 1) was intended as a pilot experiment, and had several caveats. Firstly, the average inter-trial interval was variable across-birds because call playback was triggered only once the bird fell silent (see Methods). Additionally, two of the ten birds were exposed to various experimental paradigms including electrophysiological recordings from HVC prior to this experiment. Analyses that include only the eight experimentally-naïve birds are shown in Supplementary Figure 2.1. The directionality of all correlations reported thus far were maintained, although in some cases only a marginally significant ($p < 0.1$) relationship was observed; this was likely due to the smaller sample size and tempo range included in these analyses.

To address these caveats, a second experiment was conducted in experimentally-naïve males. In this experiment, the inter-trial interval was between 5 and 11 seconds (determined pseudorandomly for each trial), regardless of whether the bird was silent or making noise. Also, to remain consistent with data derived from females (see following sections), birds were habituated to a sound isolation chamber for 18 hours and were not moved to a separate testing chamber prior to the call playback test. Ten minutes prior to the call playback test, a baseline session was conducted in which no call stimuli were played but 3 seconds of audio was recorded every 5-11 seconds. After this baseline session, the distance call used in Experiment 1 was played through a speaker every 5-11 seconds (Figure 2.4 A, see Methods).

As in Experiment 1, birds responded to the call stimulus with single calls, call trains, or songs. Most birds (11 out of 12) vocalized during the baseline session. During the baseline session, there was no relationship between mean song tempo and call latency, percent successful

trials, nor number of call responses per successful trial (Supplemental Figure 2.2). Here, call "latency" was defined as the time from the start of the trial to the onset of the first call response. Thus, any observed relationships between tempo and calling behavior are specific to a social call exchange, rather than non-social, spontaneous calling.

Next, the stability of call responses between the two playback sessions was evaluated. The latency to return the call was stable across sessions (Figure 2.4 B). Percent successful trials for each bird in Session 1 was positively correlated with percent successful trials in Session 2 (Figure 2.4 C, left panel), although as a population, birds responded to more trials in Session 1 than in Session 2 (Figure 2.1 C, right panel). There was no significant correlation between the mean number of responses per successful trial in Session 1 vs. Session 2 (Figure 2.4 D, left panel). Notably, data from 10 out of the 12 males fell close to the unity line, indicating within-bird stability of this metric, while the two males with the highest mean number of call responses in the first session showed much lower mean number of call responses in the second session. There was a significant reduction in number of call responses per successful trials across the population from Session 1 to Session 2 (Figure 2.4 D, right panel). Taken together, these results suggest that latency and percent successful trials are stable within-bird phenotypes. Meanwhile, the mean number of responses per successful trial may be less stable particularly in birds that respond vigorously during the first session.

Consistent with Experiment 1, latency to return the call was negatively correlated with mean song tempo (Figure 2.5 A). The percent successful trials was non-significantly ($p = 0.050$) correlated with song tempo (Figure 2.5 B). In contrast to Experiment 1, there was no significant relationship between the mean number of responses per successful trial and song tempo (Figure 2.5 C). Mean song tempo clustered within-nest, as expected ($p = 0$, one-tailed t-test). Latency to

respond clustered within-nest (Figure 2.5 D), but percent successful trials did not (Figure 2.5 E) and nor did number of responses per successful trial (Figure 2.5 F).

Latency to respond was anticorrelated with percent successful trials - as in Experiment 1 - but this effect was non-significant (Figure 2.6 A). Unlike Experiment 1, there was no observed correlation between the mean number of responses per successful trial with latency (Figure 2.6 B), nor between the mean number of responses per successful trial with percent successful trials (Figure 2.6 C).

Several discrepancies were observed between Experiment 1 and Experiment 2. First, why did the mean number of responses per successful trial correlate with song tempo in Experiment 1 but not Experiment 2? This may be due to the fact that a number of trials were excluded in Experiment 2 when birds were mid-song or mid-call at the time of stimulus playback. Such exclusions were avoided in Experiment 1, since the call stimulus was only played when the bird was silent (see Methods). Trial exclusions in Experiment 2 were particularly prevalent at the start of the behavioral session, when birds tended to respond most vigorously with calls and songs. This is important, since there was an anticorrelation between number of responses per trial and the trial number in most birds in both Experiment 1 and 2 (data not shown). Thus, trials in which the birds may have responded with many vocalizations were excluded in Experiment 2; this likely dampened any correlation between song tempo and number of callbacks. Second, why was the range of percent successful trials larger in Experiment 1 (22.4% - 99.1%) than in Experiment 2 (92.4% - 100%)? This may be explained by the difference in habituation time: 10 minutes in Experiment 1 vs. 18 hours in Experiment 2. The long habituation time in Experiment 2 may have led to reduced stress levels, which has been linked to increased rates of vocalization (Yamahachi et al., 2020). This is supported by the observation that females responded to very few trials or not

at all when habituated to the testing chamber for only 10 minutes but responded to many trials when habituated for 18 hours (see following sections). Despite these differences, the combined results of Experiments 1 and 2 in males suggest a robust relationship between song tempo and calling behavior.

2.1.B. Call response features are not correlated with sire's song tempo in naturally-reared female Bengalese finches.

Although female finches do not sing, they do learn their sire's song (Miller, 1979). Much like males, female finches call back with stereotypic latency to call stimuli. In both males and females, the precision of call timing is abolished by lesion of the robust nucleus of the archipallium, which is immediately downstream of HVC (Benichov et al., 2016). These findings raise the possibility that similar mechanisms underlie call timing in males and females. If this was true, call features in the female would be expected to correlate with the sire's song tempo since this tempo reflects both the female's genetic background and her experience (i.e., the song that she learned). To evaluate this, the calling behavior of experimentally-naïve adult female Bengalese finches was examined as a function of sire's song tempo.

When Experiment 1 was conducted in females (Figure 2.7 A), most birds (6 of 9) did not respond in any trials (Figures 2.7 B-C). The three females that responded did so in a minority of trials (Figures 2.7 B-C) and the percent of successful trials was inconsistent across sessions (Figure 2.7 B). It was not feasible to examine response latency nor mean number of responses per trial as a function of sire's song tempo with so few female responders. Therefore, only the relationship between percent successful trials and sire's song tempo was examined. Surprisingly, there was a non-significant ($p = 0.062$) negative correlation between the sire's song tempo and percent successful trials (Figure 2.7 C).

To determine whether this relationship between percent successful trials and sire's song tempo was meaningful and replicable, a separate cohort of females was tested after 18 hours of habituation (Experiment 2; Figure 2.8 A). Most females vocalized in the 10 minutes before the call playback session, suggesting successful habituation to the sound isolation chamber (Supplementary Figure 2.3).

Females produced single calls and calls with multi-trilled structure in response to call playback (Figure 2.8 B) and distributions of call latency had a clearly identifiable peak (Figure 2.8 C). Call features were moderately stable over multiple sessions. There was a strong positive correlation between callback latency in Session 1 vs. Session 2 (Figure 2.8 D, left panel). Notably, the linear fit of these data deviated far from the unity line, although this was driven by one bird whose latency more than doubled from the first session to the second session. Overall, latency was consistent across sessions and there was no significant difference between latency in Session 1 vs. Session 2 (Figure 2.8 D, right panel). Percent successful trials was stable across sessions (Figure 2.8 E). There was a significant correlation between the mean number of responses per successful trials in Session 1 vs. Session 2 (Figure 2.8 F, left panel), although as a population birds responded with fewer vocalizations in the second session (Figure 2.8 F, right panel). Taken together, these findings suggest that - as in males - call latency, percent successful trials, and the mean number of responses per successful trials are stable within-bird phenotypes that have a wide range across the population. It appears that these metrics may shift across the population due to the novelty of the paradigm.

There was no relationship between call latency, percent successful trials, nor the mean number of responses per successful trials with sire's song tempo (Figure 2.9 A-C). Moreover, no within-nest clustering of any call metric was observed (Figure 2.9 D-F). These findings suggest

that in naturally-reared females - unlike males - calling behavior is not dependent on genetic nor experience-dependent factors that are consistent within a family. Thus, it seems unlikely that calling behavior in females is primarily dependent on learning of the sire's song, genetically-determined excitability of song circuitry, or other mechanisms that are expected to be consistent within-nest.

However, call parameters were strongly correlated with one another. In particular, call response latency was significantly anticorrelated with percent successful trials (Figure 2.10 A). There was also a significant anticorrelation between call response latency and the mean number of responses per successful trial (Figure 2.10 B). Finally, percent successful trials and the mean number of responses per successful trial were strongly correlated with one another (Figure 2.10 C). These findings suggest that a unique factor for each bird that is inconsistent within the nest dictates calling behavior. Such factors may include stress, which is known to impact the number and type of vocalizations in zebra finches (Yamahachi et al., 2020), or preference for the call stimulus, which may be indicated by number of call responses (Dunning et al., 2014; Dunning et al., 2020).

2.1.C. Call response features are correlated with the sire's song tempo in non-tutored female Bengalese finches.

To isolate the genetic contributions to calling behavior, a cohort of non-tutored birds was tested in the Experiment 2 call playback paradigm (Figure 2.11 A). Pairs of eggs were transferred from their home nests to be raised to independence by a female foster pair. At independence, each pair was moved into a sound isolation chamber where they remained until adulthood. This protocol ensured that each pair was not tutored with any song. Only same-sex pairs were included in analysis (see Methods for further detail). Importantly, the calling behavior of only four non-

tutored females from two nests has been examined thus far, and as such, results described here should be considered preliminary.

Non-tutored females responded to the call stimulus with calls that were similar in acoustic structure to that of naturally-reared females (Figure 2.11 B). Each bird responded with a stereotypic latency, and the distribution of latencies for each bird had a clearly identifiable peak (Figure 2.11 C). Although there were no positive correlations between each call metric in Session 1 vs. Session 2 (Figure 2.11 D-F, left panels), the spread of the data from the unity line was actually similar to that observed in naturally-reared males or females (see Figures 2.1, 2.4, 2.8). There were no differences between Session 1 vs. Session 2 in any call metric across the population (Figure 2.11 D-F, right panels). The small sample size makes it difficult to determine whether the call metrics were stable within-bird across multiple sessions.

There was a strong negative correlation between callback latency and sire's song tempo in these non-tutored females (Figure 2.12 A). There was no relationship between percent successful trials and sire's song tempo (Figure 2.12 B) and a non-significant ($p = 0.077$) positive correlation between the mean number of responses in successful trials with sire's song tempo (Figure 2.12 C). There was also a non-significant ($p = 0.086$) anticorrelation between latency and the mean number of responses per successful trial (Figure 2.13 B) and a non-significant ($p = 0.066$) positive correlation between the mean number of responses per successful trial and percent successful trials (Figure 2.13 C). The non-tutored sibling pairs had remarkably similar callback latencies. One sibling pair had typical latencies of 248 ms and 255 ms while the second sibling pair had latencies of 181 ms and 185 ms. This tight within-nest clustering of latency may be due to high coordination between the pair due to prolonged exposure, as has been observed in

paired songbirds (Prior et al., 2020), or to high heritability of callback latency in deprived social and learning conditions (discussed below).

These initial results are intriguing, although a larger sample of individuals and nests is required before making further conclusions. If future data were to follow this initial trend, it would seem that - unlike naturally-reared females - the calling behavior of non-tutored females is well-predicted by the sire's song tempo. Such a finding would be reminiscent of earlier work (Mets and Brainard, 2018) showing that the genetic contribution to song tempo is higher when the tutoring environment is poor (when the bird is computer-tutored through song playbacks) than when it is rich (when the bird is live-tutored by an unrelated adult male).

2.2 Methods

2.2.A. Subjects

Data were collected from 22 naturally-reared male Bengalese finches (*Lonchura striata domestica*; call playback Experiment 1: n = 10 individuals from 4 nests; call playback Experiment 2: n = 12 individuals from 5 nests), 21 naturally-reared female Bengalese finches (call playback Experiment 1: n = 9 individuals from 5 nests; call playback Experiment 2: n = 12 individuals from 6 nests), and 4 non-tutored female Bengalese finches (n = 4 individuals from 2 nests). At the time of the call playback experiment, males in Experiment 1 were between ages 137 and 145 days post-hatch (dph; mean of 137.9 +- STD 8.8 dph) and males in Experiment 2 were between ages 94 and 165 dph (mean of 115.0 +- STD 20.2 dph). Females in Experiment 1 were between ages 116 and 256 dph (mean of 152.2 +- STD 48.5 dph) and females in Experiment 2 were between ages 95 and 144 dph (mean of 113.8 +- STD 18.5 dph). Non-tutored females were between ages 96 and 104 dph (mean of 100 +- STD 4.6 dph). Two males in Experiment 1 had been used previously in electrophysiology experiments involving recordings

from HVC. Data excluding these two birds are presented in Supplementary Figure 2.1. All males in Experiment 2 and all females were naïve to experimentation.

All birds were bred in-house, with the exception of one male in Experiment 1 that was purchased from an outside vendor but was excluded from analysis due to its highly unusual song. All naturally-reared birds were kept in their home nest in the colony room until adulthood (~90 dph). Birds were then transferred to same-sex flight cages in the colony room, where they could hear and see hundreds of conspecifics.

Non-tutored females were removed as eggs from their parents ~36 hours after laying. Eggs were transferred to a pair of adult foster females housed in sound isolation chambers (Acoustic Systems), who raised the juveniles until independence (35-40 dph). At this time, juveniles were moved together into a separate sound isolation chamber where they remained until the call playback paradigm was conducted in adulthood (>90 dph). This ensured that each pair was not tutored with any song. Only same-sex pairs were included in analysis.

2.2.B. Song recording

Vocalizations were recorded in early adulthood (~90 dph) from birds temporarily singly-housed in sound isolation chambers. Audio was recorded in a custom-written Labview program (National Instruments; sampling frequency of 32 kHz) using a microphone (Countryman) and AM Systems pre-amplifier. Audio files were recorded when an amplitude threshold was crossed multiple times within a 2-second window. Audio from the 2 seconds prior to the amplitude threshold crossing and 1 second after noise cessation were recorded and saved. In post-processing, a custom MATLAB script was used to identify putative song files, based on the number and duration of amplitude threshold crossings.

2.2.C. Song tempo analysis

Song tempo was defined as the mean number of syllables per second of song. Putative song files were read into a custom MATLAB program that created an amplitude envelope by rectifying and smoothing the audio waveform. Amplitude threshold crossings of more than 10 ms were identified as putative syllables, which were then confirmed by eye to contain song-like spectral information. Song syllables separated from one another by no more than 250 ms were considered part of the same song bout. For each bird, tempo was quantified as the number of syllables in a song bout divided by the duration of the song bout in seconds, averaged across at least 15 song bouts.

2.2.D. Call playback paradigm

Call playback Experiment 1

The evening before the call playback experiment (~18:00), a cohort of birds (n = 4 per cohort) were transferred from the colony room to be singly-housed in sound isolation chambers. Each bird was housed in a wire cage with *ad libitum* access to food, water, and grit. The next day (~12:00), each bird was moved to a separate sound isolation chamber for testing. The cage in this chamber had one clear acrylic side which allowed for video monitoring and recording via a video camera (sampling rate of 30 frames per second; audio sampling rate of 16 kHz) during the baseline and habituation sessions. A microphone was secured atop the cage in a fixed location for all birds. No food, water, or grit was available in the testing chamber.

Birds were habituated to the testing chamber for 10 minutes, during which time audio was recorded via a custom-written LabView program (sampling rate of 32 kHz), and audiovisual data was recorded via a video camera. At the start of the habituation session, the experimenter placed their hand in the camera's field of view and lightly tapped on the testing cage; this served

as a landmark by which to synchronize audio recorded in LabView with audiovisual data recorded by the camera. The LabView program recorded 3 seconds of audio every 5-11 seconds. Additionally, if the bird vocalized or produced other noises that crossed an amplitude threshold (e.g., from hitting the cage or beating its wings), 5 seconds of audio was recorded (the 2 seconds prior to and the 3 seconds after the amplitude threshold crossing).

After the 10-minute habituation period, the call playback session began. A distance call recorded from a Bengalese finch male was played at 75 dB relative to the center of the bird's cage. The 2 seconds prior to the stimulus playback and the 3 seconds after the stimulus playback were recorded and saved. Additionally, if the bird vocalized or produced other noises that crossed an amplitude threshold, then 5 seconds of audio were recorded (the 2 seconds prior to and the 3 seconds after the amplitude threshold crossing). A call was played every 5-11 seconds (determined pseudorandomly for each trial) or 2 seconds after the bird ceased making noise, whichever came later. This ensured that no call playback occurred while the bird was singing or calling. After the 10-minute call playback session, the bird was transferred back to its original sound isolation chamber. Before the next bird was transferred to the testing chamber, a clean piece of newspaper was laid on the floor of the cage and the acrylic wall was cleaned with ethanol.

Four hours later (~16:00), birds were again habituated in the testing chamber and tested in the call playback paradigm as described above. After the call playback session, the bird was transferred back to its original sound isolation chamber. After all birds in the cohort were tested in the second call playback session, all birds were returned to the colony room.

Call playback Experiment 2

This experiment was similar to Experiment 1, but with two exceptions. First, rather than being transferred from one sound isolation chamber to the testing chamber 10 minutes prior to the playback session, birds were habituated to the sound isolation chamber for 18 hours. Second, the inter-trial interval was set to 5-11 seconds, regardless of whether the bird was vocalizing or making other noises. This ensured that all birds had equal distributions of inter-trial interval, regardless of how much they sang, called, or otherwise made noise.

The evening before the call playback experiment (~18:00), birds were transferred from the colony room to be singly-housed in sound isolation chambers. Each bird was housed in a wire cage with *ad libitum* access to food, water, and grit. The next day (~12:00), a 10-minute baseline session was conducted in which no calls were played. The LabView program recorded 3 seconds of audio every 5-11 seconds. Next, the call playback session began. A distance call recorded from a Bengalese finch male was played at 75 dB relative to the center of the bird's cage. The 2 seconds prior to the stimulus playback and the 3 seconds after the stimulus playback were recorded and saved. A call was played every 5-11 seconds (determined pseudorandomly for each trial). During the baseline and call playback sessions, behavior was monitored and recorded via a video camera (video sampling rate of 60 frames per second; audio sampling rate of 44 kHz). Four hours later (~16:00), a 10-minute baseline session was conducted again followed by a 10-minute playback session. After all birds in the cohort were tested in the second call playback session, all birds were returned to the colony room.

2.2.E. Call playback analysis

Files from the call playback session were read into a custom MATLAB script that created an amplitude envelope by rectifying and smoothing the audio waveform. Only files that contained a call stimulus were analyzed; in Experiment 1, this excluded any files triggered by noise made by

the bird between trials but in Experiment 2, this did not exclude any files. Amplitude threshold crossings of more than 5 ms were identified as putative calls or song syllables, which were then confirmed by eye to contain call-like or song-like spectral information. In rare cases that vocalizations could not be easily distinguished by eye (e.g., if a putative call occurred during a period of high cage noise), the audio file was played via Audacity to aurally confirm the presence or absence of the vocalization. The call playback stimulus, calls produced by the experimental bird, and songs produced by the experimental bird were assigned unique identifiers. All song syllables within a song bout (i.e., separated by ≤ 250 ms) were coded as a single unit (i.e., one song). Any files in which the bird's call or song began before the call stimulus and overlapped with the call stimulus were excluded from analysis.

After all files were scored in this manner, the minimum number of included trials across all birds in the experimental group was determined. All analysis thereafter included only this minimum number of trials, to ensure that an equal number of trials was compared across birds. For instance, if the minimum number of included trials for Experiment 1 males was 100, then the first 100 trials were examined for all Experiment 1 males.

Features of spontaneous calls were examined by scoring and analyzing calls produced during the habituation (Experiment 1) or baseline (Experiment 2) sessions. Call files from the first habituation or baseline session were scored identically to files from the playback sessions, except that no call stimulus was labeled. The relationship between each of the call metrics described below and mean song tempo was then examined via linear regression (see Statistics section).

Call Latency

Call latency for each trial was defined as the time in milliseconds from the onset of the call stimulus to the onset of the first call response from the experimental bird. For habituation or baseline (no call playback) sessions, latency was defined as the time in milliseconds from the onset of the trial to the onset of the first call. To find the typical call latency, a histogram with 20-ms bins was fit with spline interpolation. The latency value associated with the peak of this fit was defined as the typical latency.

Percent Successful Trials

Successful trials were defined as any in which the bird responded to the call stimulus at least once. The percent successful trials was calculated as

$$[(\#\ \text{Successful Trials}) / (\text{Total}\ \#\ \text{Trials})] * 100.$$

Mean Number of Calls Per Successful Trial

Successful trials were defined as any in which the bird responded to the call stimulus at least once. The number of call responses within each successful trial was found and averaged across all successful trials.

Statistics

To examine the relationship between each call response metric in Session 1 vs. Session 2, a linear model was fit to the Session 1 x Session 2 data. An F-test was used to determine whether the slope of the regression line was significantly different from zero, and an R^2 value was found to determine the goodness of model fit. A Wilcoxon signed-rank test was used to compare the paired Session 1 vs. Session 2 data for each call metric to evaluate any upward or downward shifts across a given population.

To examine the relationship between song tempo and each call response metric, a linear model was fit to the tempo x call metric data. An F-test was used to determine whether the slope

of the regression line was significantly different from zero, and an R^2 value was found to determine the goodness of model fit.

Within-Nest Clustering

Monte Carlo simulations were used to determine whether each call metric clustered within-nest. The observed data consisted of the call metric for each bird (e.g., call latency for each bird) and the nest identity of each bird. To determine the extent to which the call metric clustered within-nest, the distance between each bird's datapoint and the mean for the nest was calculated. Then, the distance of each bird's metric from the nest mean was averaged across the population to identify the observed average within-nest distance. To determine the clustering of the call metric by chance alone, the relationship between call metric and nest identity was shuffled 10,000 times. In each of the 10,000 simulations, the within-nest distance (using the shuffled data-nest relationships) was calculated. A call metric was determined to cluster by nest if the chance-level average within-nest distance was higher than the observed average within-nest clustering more than 95% of the time.

2.3 Figures

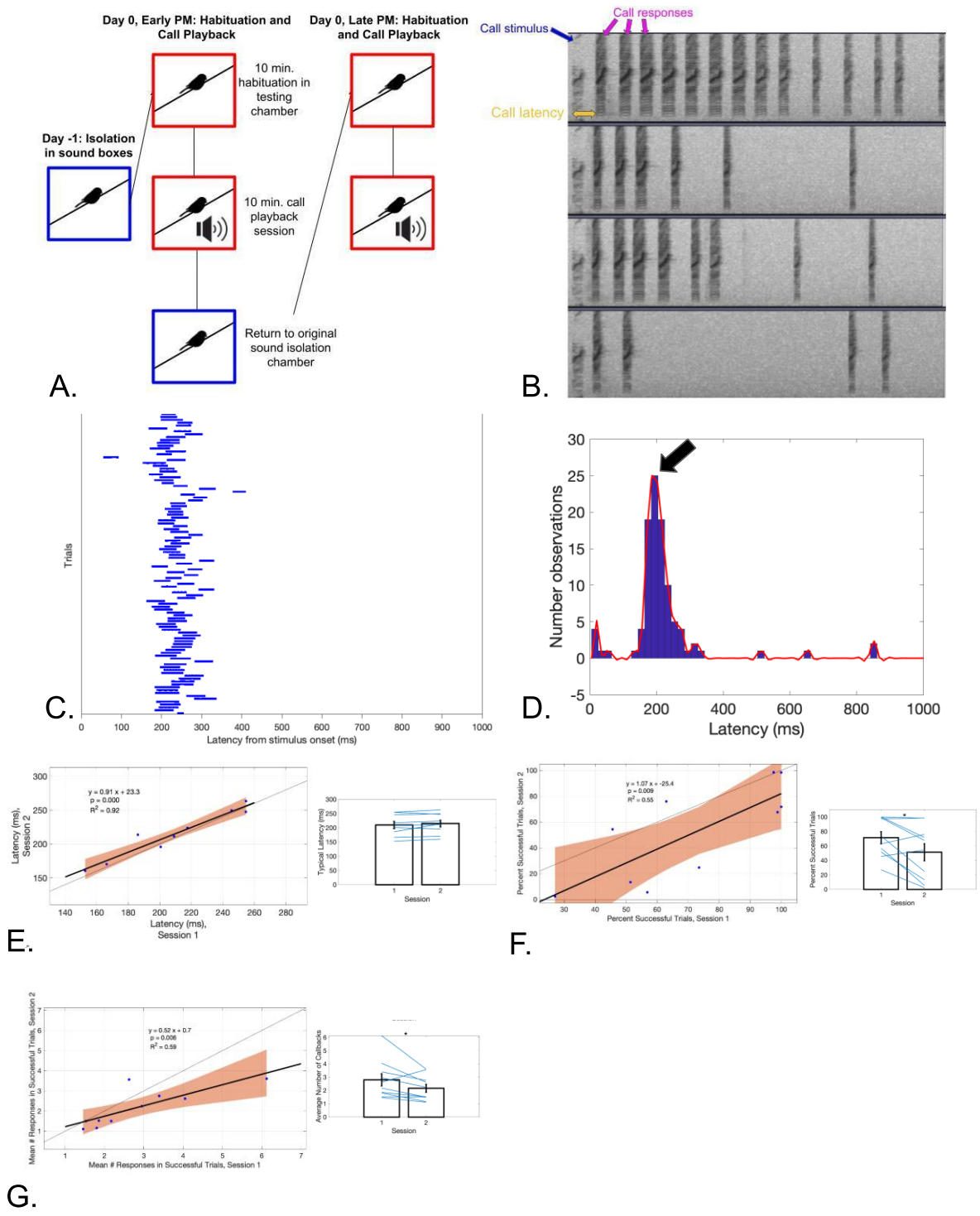


Figure 2.1. Call latency and other call features are stable within-bird phenotypes (Males, Experiment 1). A. Experimental design for Experiment 1. The day prior to the call playback test, birds were moved from the colony to be singly-housed in a sound isolation chamber (blue box). The following day, each bird was habituated to the testing chamber (red box) for 10 minutes

prior to the start of the 10-minute playback session. During the playback session, the distance call of an adult male conspecific was presented to the experimental bird every 5-11 seconds, or 2 seconds after the experimental male stopped making noise, whichever came later. The bird was then returned to its original sound isolation chamber for four hours. The habituation and testing session were then repeated. B. Four representative trials from a single bird. The first call is the call stimulus. All subsequent calls were produced by the experimental bird. The time between the onset of the call stimulus and the onset of the first call response was considered the call latency. C. Raster plot of first call responses from one representative bird. Blue lines indicate the first call response generated by the experimental bird. The length of the blue line indicates the length of the call. D. Histogram of call latencies (bin size = 20 ms) overlaid with a spline fit. The typical latency was defined as the call latency at the peak of the spline fit, marked with a black arrow. E. *Left*: Significant positive correlation between typical latency value in the first session versus second session (n = 10 individuals from 4 nests, slope = 0.64, $R^2 = 0.74$, p = 2.7E-5). The unity line is shown as a thin black line and a linear fit of the data is shown as a thick black line. The 95% confidence intervals of the linear fit are shown in orange. A perfect correlation between Session 1 values and Session 2 values would be indicated by perfect alignment of the linear fit with the unity line. *Right*: Latencies in Session 1 were not significantly different from latencies in Session 2 (p = 0.16, Wilcoxon signed-rank test). F. *Left:* Significant positive correlation between percent successful trials in Session 1 vs. Session 2 (n = 10 individuals from 4 nests, slope = 1.07, $R^2 = 0.55$, p = 0.009). *Right:* Percent successful trials was significantly lower in Session 2 vs. Session 1 (p = 0.049, Wilcoxon signed-rank test). G. *Left*: Significant positive correlation between mean number of call responses in successful trials in Session 1 vs. Session 2 (n = 10 individuals from 4 nests, slope = 0.52, $R^2 = 0.59$, p = 0.006). *Right:* Lower mean number of call responses in successful trials in Session 2 vs. Session 1 (p = 0.049, Wilcoxon signed-rank test). All bar plots display mean ± SEM.

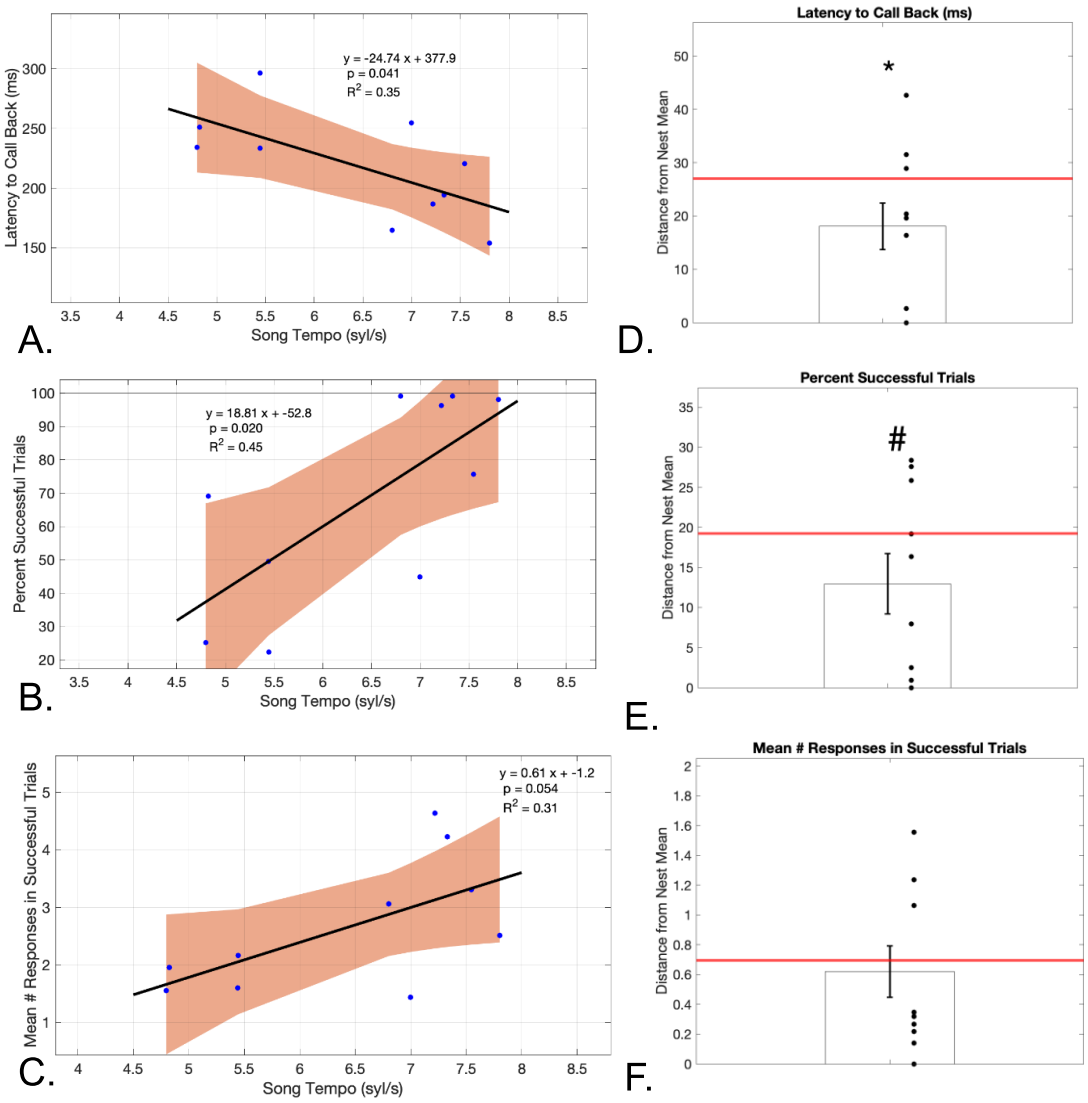


Figure 2.2. Call features correlate with song tempo (Males, Experiment 1). A. Significant anticorrelation between typical call response latency and mean song tempo (n = 10 individuals from 4 nests, slope = -24.7, $R^2 = 0.35$, p = 0.041). The 95% confidence intervals of the linear fit are shown in orange. B. Significant positive correlation between percent successful trials and mean song tempo (n = 10 individuals from 4 nests, slope = 18.8, $R^2 = 0.45$, p = 0.020). C. Non-significant positive correlation between mean number of responses per successful trial and mean song tempo (n = 10 individuals from 4 nests, slope = -0.61, $R^2 = 0.31$, p = 0.054). D. Significant within-nest clustering of call response latency (p = 0.049, one-tailed t-test). E. Non-significant within-nest clustering of percent successful trials (p = 0.076, one-tailed t-test). F. No within-nest clustering of number call responses per successful trial (p = 0.26, one-tailed t-test). For plots in D, E, and F: Each black data point indicates the difference between the bird's value for the call metric and the mean value of the metric within the bird's nest. The red line shows the average distance from the nest mean by chance across 10,000 simulations. All bar plots display mean ± SEM.

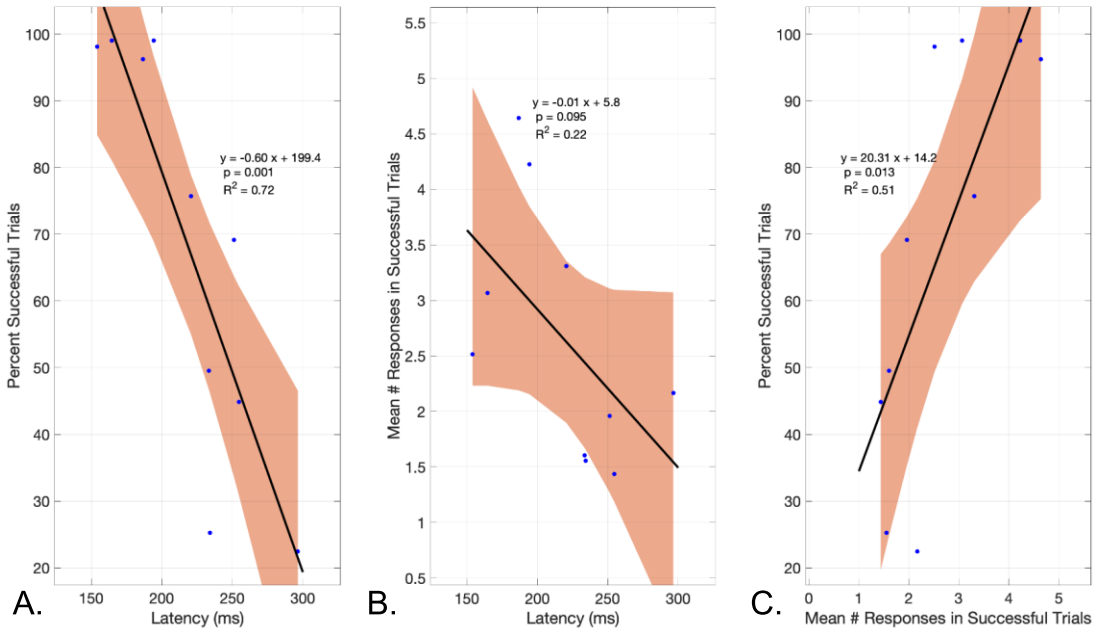


Figure 2.3. Call features correlate with one another (Males, Experiment 1). A. Significant anticorrelation between percent successful trials and typical call response latency (n = 10 individuals from 4 nests, slope = -0.60, $R^2 = 0.72$, p = 0.001). The 95% confidence intervals of the linear fit are shown in orange. B. Anticorrelation (trend) between mean number of responses per successful trial and typical call response latency (n = 10 individuals from 4 nests, slope = -0.01, $R^2 = 0.22$, p = 0.095). C. Significant positive correlation between mean number of responses per successful trial and percent successful trials (n = 10 individuals from 4 nests, slope = 20.31, $R^2 = 0.51$, p = 0.013).

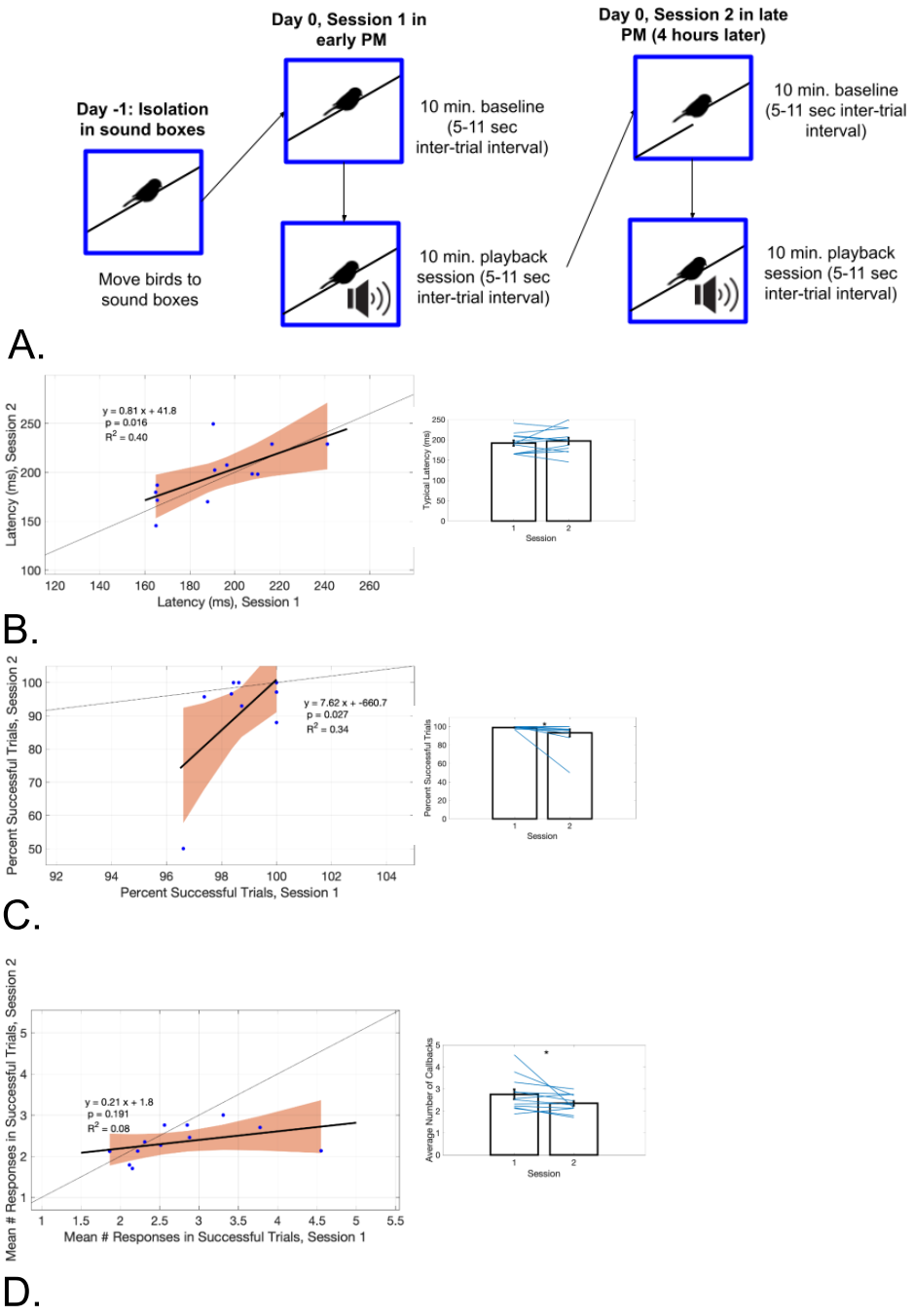


Figure 2.4. Stability of call latency and other call features across multiple sessions (Males, Experiment 2). A. Experimental design for Experiment 2. The day prior to the call playback test, birds were moved from the colony to be singly-housed in a sound isolation chamber (blue box). The following day, a 10-minute session without call playbacks was conducted to examine baseline calling levels. During the playback session, the distance call of an adult male conspecific was presented to the experimental bird every 5-11 seconds. Four hours later, the baseline and testing sessions were repeated. B. *Left:* Significant positive correlation between the typical latency in Session 1 vs. typical latency in Session 2 (n = 12 individuals from 5 nests,

slope = 0.81, R^2 = 0.40, p = 0.016). The unity line is shown as a thin black line and a linear fit of the data is shown as a thick black line. The 95% confidence intervals of the linear fit are shown in orange. A perfect correlation between Session 1 values and Session 2 values would be indicated by perfect alignment of the linear fit with the unity line. *Right*: Latencies in Session 1 were not significantly different from latencies in Session 2 (p = 0.62, Wilcoxon signed-rank test). C. *Left:* Significant positive correlation between percent successful trials in Session 1 vs. percent successful trials in Session 2 (n = 12 individuals from 5 nests, slope = 7.62, R^2 = 0.34, p = 0.027). *Right:* Higher percent successful trials in Session 1 vs. Session 2 (p = 0.039, Wilcoxon signed-rank test). D. *Left:* No correlation between the mean number responses per successful trial in Session 1 vs. the mean number responses per successful trial in Session 2 (n = 12 individuals from 5 nests, slope = 0.21, R^2 = 0.08, p = 0.19). *Right:* The mean number of responses per successful trial was significantly higher in Session 1 vs. Session 2 (p = 0.027, Wilcoxon signed-rank test). All bar plots display mean ± SEM.

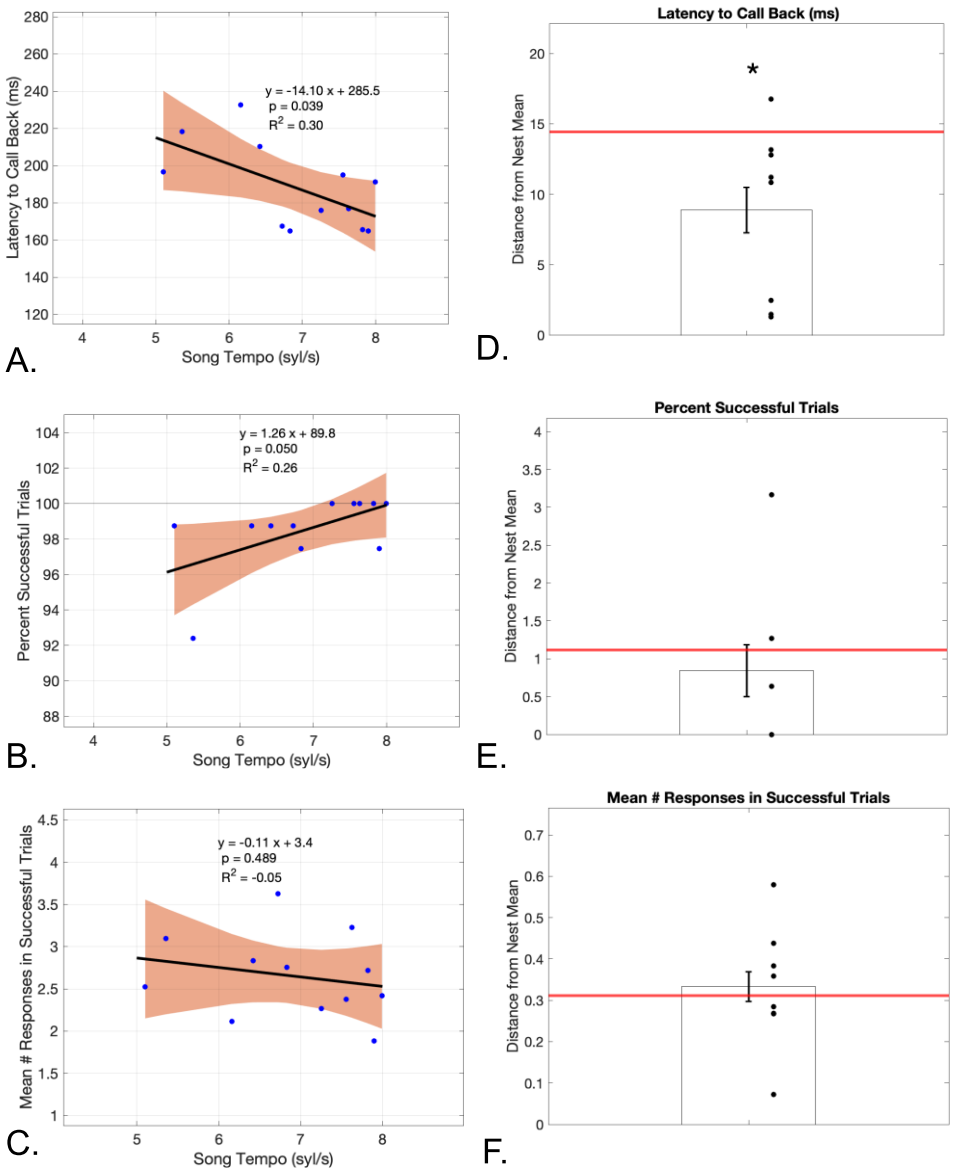


Figure 2.5. Call latency correlates with song tempo (Males, Experiment 2). A. Significant anticorrelation between typical call response latency and mean song tempo (n = 12 individuals from 5 nests, slope = -14.1, R^2 = 0.30, p = 0.039). The 95% confidence intervals of the linear fit are shown in orange. B. Non-significant positive correlation between percent successful trials and mean song tempo (n = 12 individuals from 5 nests, slope = 1.26, R^2 = 0.26, p = 0.050). C. No correlation between the mean number of responses per successful trial and mean song tempo (n = 12 individuals from 5 nests, slope = -0.11, R^2 = -0.05, p = 0.489). D. Significant within-nest clustering of call response latency (p = 0.038, one-tailed t-test). E. No within-nest clustering of percent successful trials (p = 0.11, one-tailed t-test). F. No within-nest clustering of number call responses per successful trial (p = 0.63, one-tailed t-test). For plots in D, E, and F: Each black data point indicates the difference between the bird's value for the call metric and the mean value of the metric within the bird's nest. The red line shows the average distance from the nest mean by chance across 10,000 simulations. All bar plots display mean ± SEM.

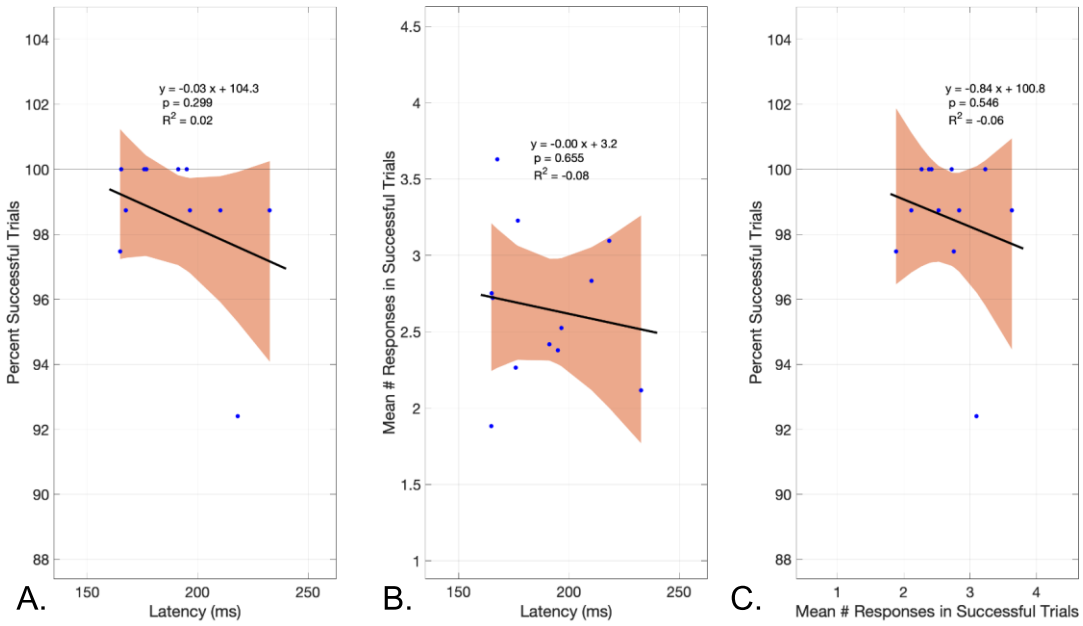


Figure 2.6. Call features do not correlate with one another (Males, Experiment 2). A. No correlation between percent successful trials and typical call response latency (n = 12 individuals from 5 nests, slope = -0.03, R^2 = 0.02, p = 0.30). The 95% confidence intervals of the linear fit are shown in orange. B. No correlation between mean number of responses per successful trial and typical call response latency (n = 12 individuals from 5 nests, slope = -0.0031, R^2 = -0.08, p = 0.66). C. No correlation between mean number of responses per successful trial and percent successful trials (n = 12 individuals from 5 nests, slope = -0.84, R^2 = -0.06, p = 0.55).

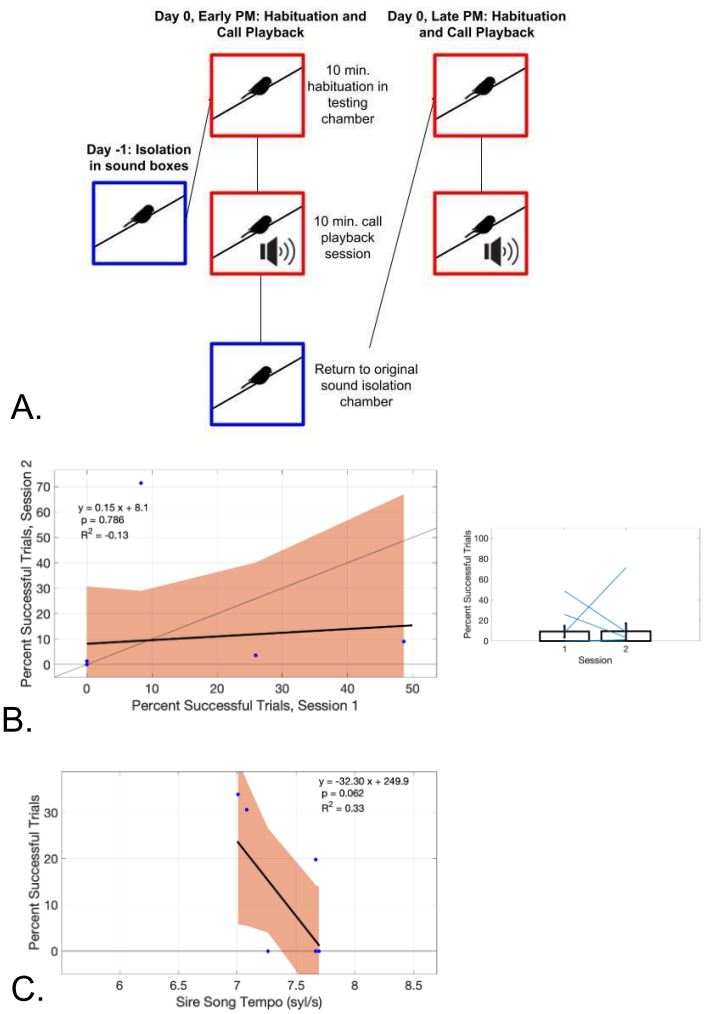


Figure 2.7. Females in Experiment 1 rarely respond to the call stimulus (Females, Experiment 1). A. Experimental design for Experiment 1. The day prior to the call playback test, birds were moved from the colony to be singly-housed in a sound isolation chamber (blue box). The following day, each bird was habituated to the testing chamber (red box) for 10 minutes prior to the start of the 10 minute playback session. During the playback session, the distance call of an adult male conspecific was presented to the experimental bird every 5-11 seconds, or 2 seconds after the experimental female stopped making noise, whichever came later. The bird was then returned to its original sound isolation chamber for four hours. The habituation and testing session were then repeated. B. *Left*: No correlation between percent successful trials in Session 1 vs. percent successful trials in Session 2 (n = 9 individuals from 5 nests, slope = 0.15, R^2 = -0.13, p = 0.79). The unity line is shown as a thin black line and a linear fit of the data is shown as a thick black line. The 95% confidence intervals of the linear fit are shown in orange. A perfect correlation between Session 1 values and Session 2 values would be indicated by perfect alignment of the linear fit with the unity line. *Right:* Percent successful trials in Session 1 was not significantly different from percent successful trials in Session 2 (p = 1, Wilcoxon signed-rank test). C. Non-significant negative correlation between sire's song tempo and percent successful trials (n = 9 individuals from 5 nests, slope = -32.3, R^2 = 0.33, p = 0.062). All bar plots display mean ± SEM.

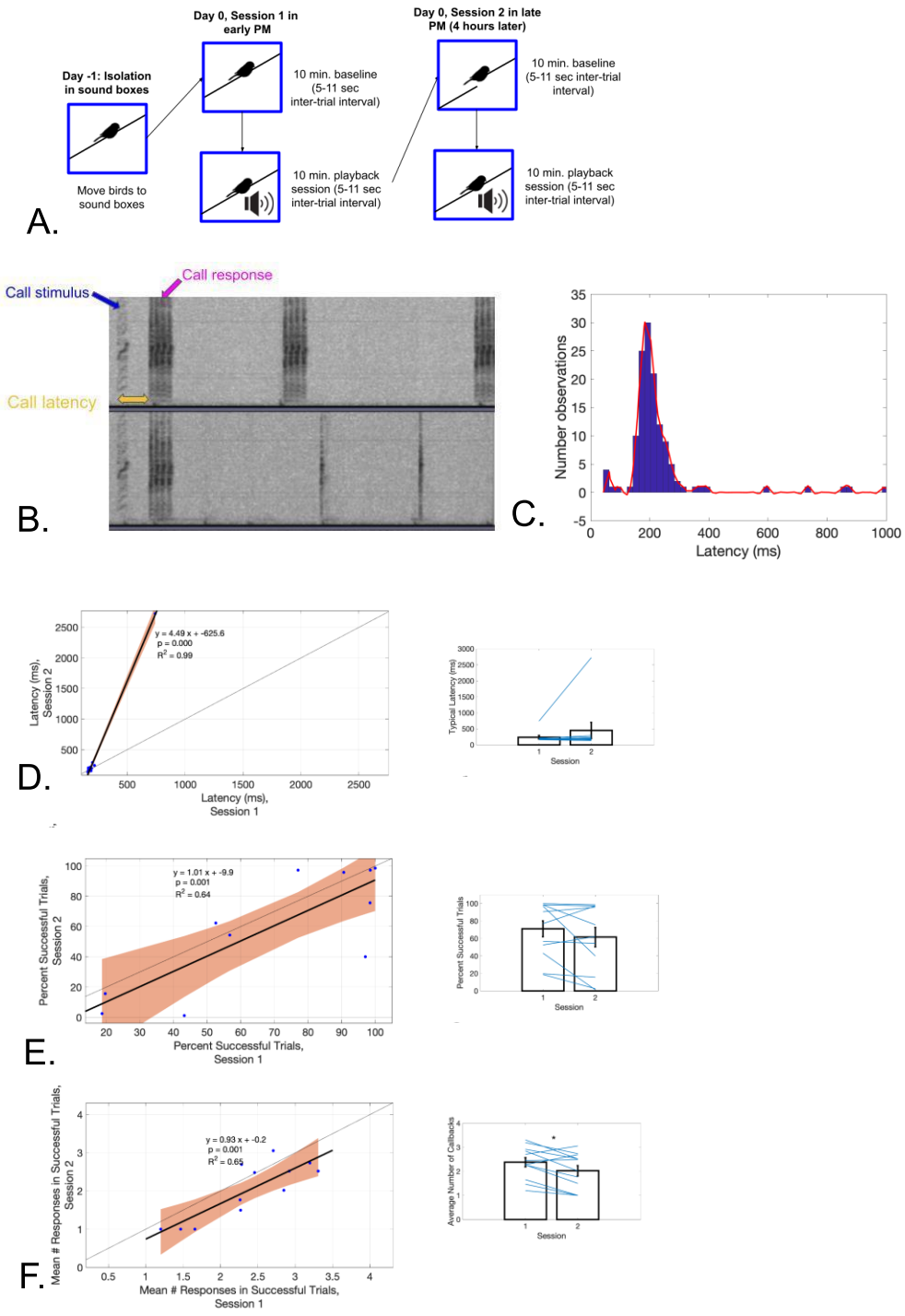


Figure 2.8. Stability of call latency and other call features across multiple sessions (Females, Experiment 2). A. Experimental design for Experiment 2. The day prior to the call playback test, birds were moved from the colony to be singly-housed in a sound isolation chamber (blue box). The following day, a 10-minute session without call playbacks was conducted to examine baseline calling levels. During the playback session, the distance call of an adult male conspecific was presented to the experimental bird every 5-11 seconds. Four hours later, the

baseline and testing session were repeated. B. Two representative trials from a single bird. The first call is the call stimulus. All subsequent calls were produced by the experimental bird. The time between the onset of the call stimulus and the onset of the first call response was considered the call latency. C. Histogram of call latencies (bin size = 20 ms) overlaid with a spline fit. The typical latency was defined as the call latency value at the peak of the spline fit. D. *Left:* Significant correlation between the typical latency in Session 1 vs. typical latency in Session 2 (n = 12 individuals from 6 nests, slope = 4.49, $R^2 = 0.99$, p = 2.4E-10). The unity line is shown as a thin black line and a linear fit of the data is shown as a thick black line. The 95% confidence intervals of the linear fit are shown in orange. A perfect correlation between Session 1 values and Session 2 values would be indicated by perfect alignment of the linear fit with the unity line. *Right*: Latencies in Session 1 were not significantly different from latencies in Session 2 (p = 0.19, Wilcoxon signed-rank test). E. *Left:* Significant positive correlation between percent successful trials in Session 1 vs. percent successful trials in Session 2 (n = 12 individuals from 6 nests, slope = 1.01, $R^2 = 0.64$, p = 0.001). *Right*: Percent successful trials in Session 1 was not significantly different from percent successful trials in Session 2 (p = 0.20, Wilcoxon signed-rank test). F. *Left:* Significant positive correlation between the mean number responses per successful trial in Session 1 vs. the mean number responses per successful trial in Session 2 (n = 12 individuals from 6 nests, slope = 0.93, $R^2 = 0.65$, p = 0.001). *Right*: The mean number of responses per successful trial was higher in Session 1 than in Session 2 (p = 0.02, Wilcoxon signed-rank test). All bar plots display mean ± SEM.

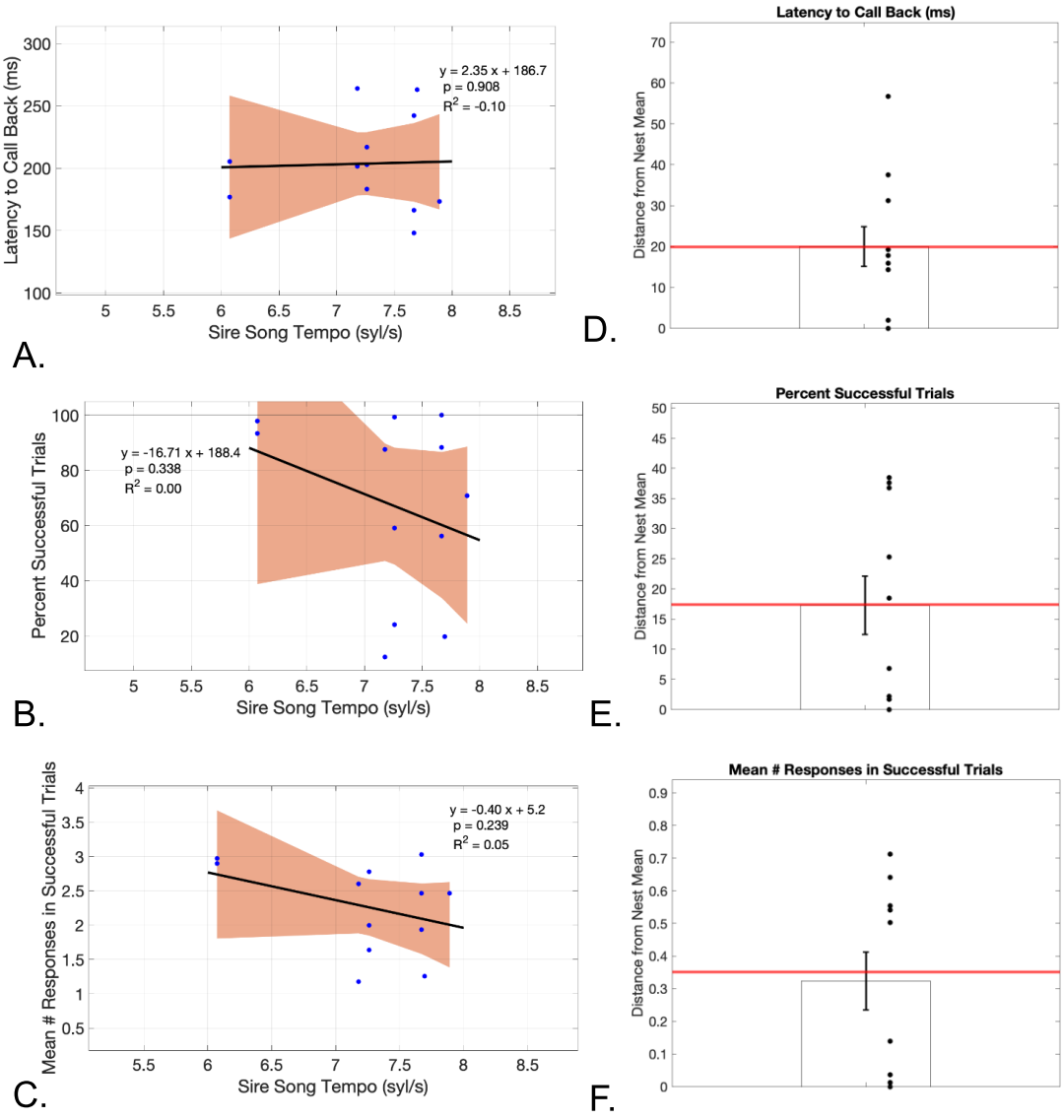


Figure 2.9. In females, there is no relationship between sire's song tempo and calling behavior (Females, Experiment 2). A. No correlation between typical call response latency and sire's mean song tempo (n = 12 individuals from 6 nests, slope = 2.35, R^2 = -0.10, p = 0.91). The 95% confidence intervals of the linear fit are shown in orange. B. No correlation between percent successful trials and sire's mean song tempo (n = 12 individuals from 6 nests, slope = -16.71, R^2 = 0.0012, p = 0.34). C. No correlation between the mean number of responses per successful trial and sire's mean song tempo (n = 12 individuals from 6 nests, slope = -0.40, R^2 = 0.05, p = 0.24). D. No within-nest clustering of call response latency (p = 0.47, one-tailed t-test). E. No within-nest clustering of percent successful trials (p = 0.47, one-tailed t-test). F. No within-nest clustering of number call responses per successful trial (p = 0.35, one-tailed t-test). For plots in D, E, and F: Each black data point indicates the difference between the bird's value for the call metric and the mean value of the metric within the bird's nest. The red line shows the average distance from the nest mean by chance across 10,000 simulations. All bar plots display mean ± SEM.

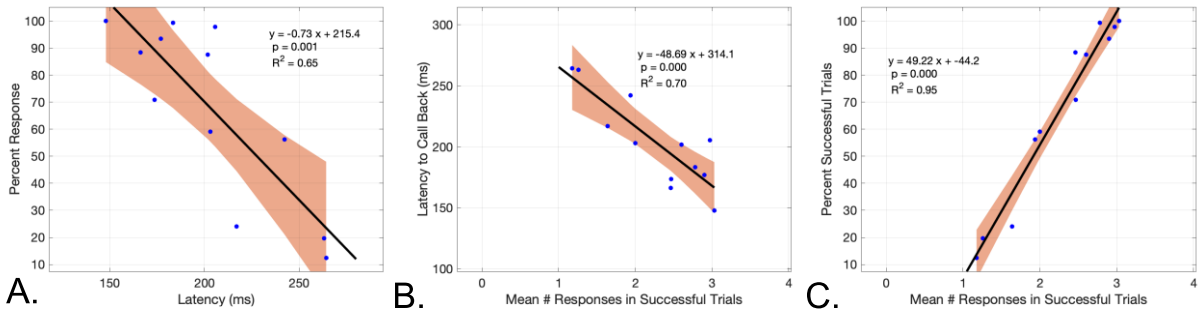


Figure 2.10. Call features correlate strongly with one another (Females, Experiment 2). A. Significant negative correlation between percent successful trials and typical call response latency (n = 12 individuals from 6 nests, slope = -0.73, R^2 = 0.65, p = 0.001). The 95% confidence intervals of the linear fit are shown in orange. B. Significant negative correlation between mean number of responses per successful trial and typical call response latency (n = 12 individuals from 6 nests, slope = -48.7, R^2 = 0.70, p = 4.5E-4). Significant positive correlation between mean number of responses per successful trial and percent successful trials (n = 12 individuals from 6 nests, slope = 49.2, R^2 = 0.95, p = 3.3E-8).

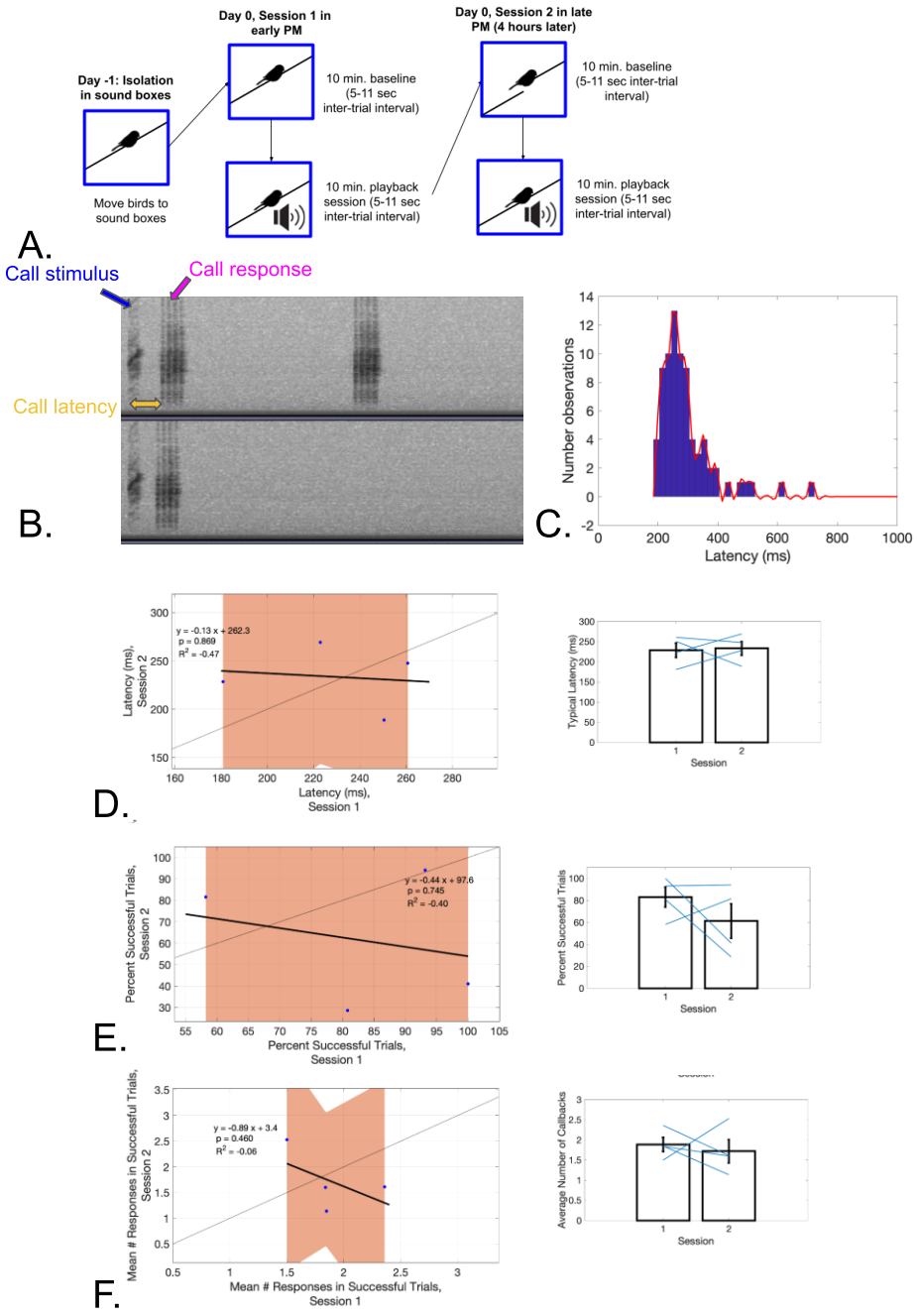


Figure 2.11. Stability of call features in non-tutored females (Non-Tutored Females, Experiment 2). A. Experimental design for Experiment 2. The day prior to the call playback test, birds were moved from the colony to be singly-housed in a sound isolation chamber (blue box). The following day, a 10-minute session without call playbacks was conducted to examine baseline calling levels. During the playback session, the distance call of an adult male conspecific was presented to the experimental bird every 5-11 seconds. Four hours later, the baseline and testing sessions were repeated. B. Two representative trials from a single bird. The first call is the call stimulus. All subsequent calls were produced by the experimental bird. The time between the onset of the call stimulus and the onset of the first call response was considered the call latency.

C. Histogram of call latencies (bin size = 20 ms) overlaid with a spline fit. The typical latency was defined as the call latency value at the peak of the spline fit. D. *Left:* No correlation between typical latency in Session 1 vs. typical latency in Session 2 (n = 4 individuals from 2 nests, slope = -0.13, R^2 = -0.47, p = 0.87). The unity line is shown as a thin black line and a linear fit of the data is shown as a thick black line. The 95% confidence intervals of the linear fit are shown in orange. A perfect correlation between Session 1 values and Session 2 values would be indicated by perfect alignment of the linear fit with the unity line. *Right*: Typical latency in Session 1 was not significantly different from typical latency in Session 2 (p = 0.99, Wilcoxon signed-rank test). E. *Left:* No correlation between percent successful trials in Session 1 vs. percent successful trials in Session 2 (n = 4 individuals from 2 nests, slope = -0.44, R^2 = -0.40, p = 0.75). *Right*: Percent successful trials in Session 1 was not significantly different from percent successful trials in Session 2 (p = 0.63, Wilcoxon signed-rank test). F. *Left:* No correlation between the mean number of responses per successful trial in Session 1 vs. mean number of responses per successful trial in Session 2 (n = 4 individuals from 2 nests, slope = -0.89, R^2 = -0.06, p = 0.46). *Right*: The mean number of responses per successful trials in Session 1 was not significantly different from the mean number of responses per successful trial in Session 2 (p = 0.88, Wilcoxon signed-rank test). All bar plots display mean ± SEM.

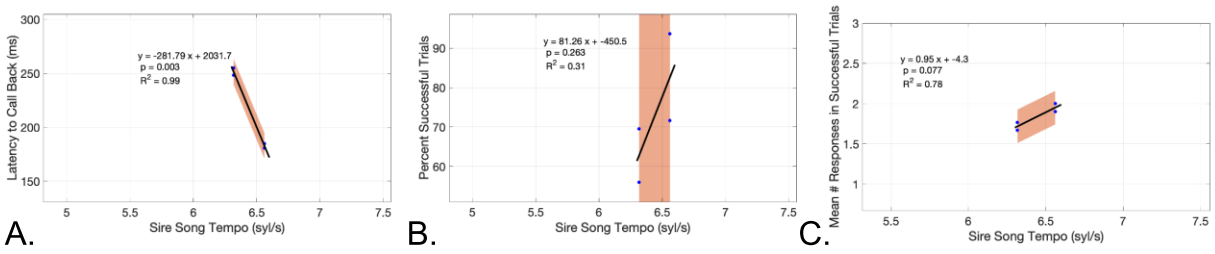


Figure 2.12. Call latency correlates with the sire's mean song tempo in non-tutored females (Non-Tutored Females, Experiment 2). A. Significant anticorrelation between typical call response latency and sire's mean song tempo (n = 4 individuals from 2 nests, slope = -281.8, R^2 = 0.99, p = 0.003). The 95% confidence intervals of the linear fit are shown in orange. B. No correlation between percent successful trials and sire's mean song tempo (n = 4 individuals from 2 nests, slope = 81.26, R^2 = 0.31, p = 0.26). C. Non-significant positive correlation between the mean number of responses per successful trial and sire's mean song tempo (n = 4 individuals from 2 nests, slope = 0.95, R^2 = 0.78, p = 0.077).

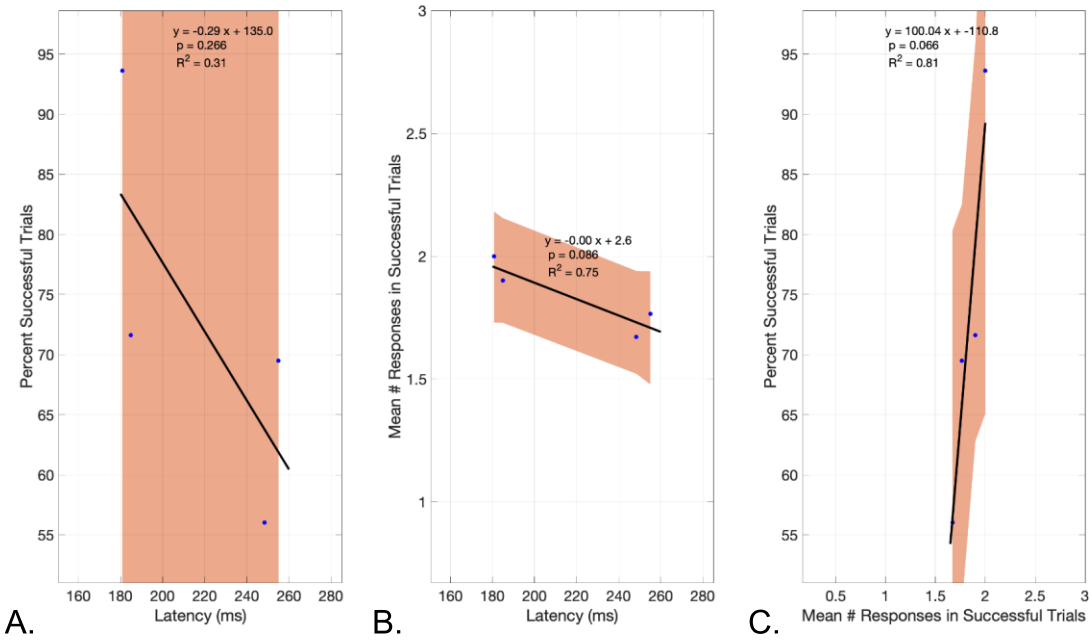
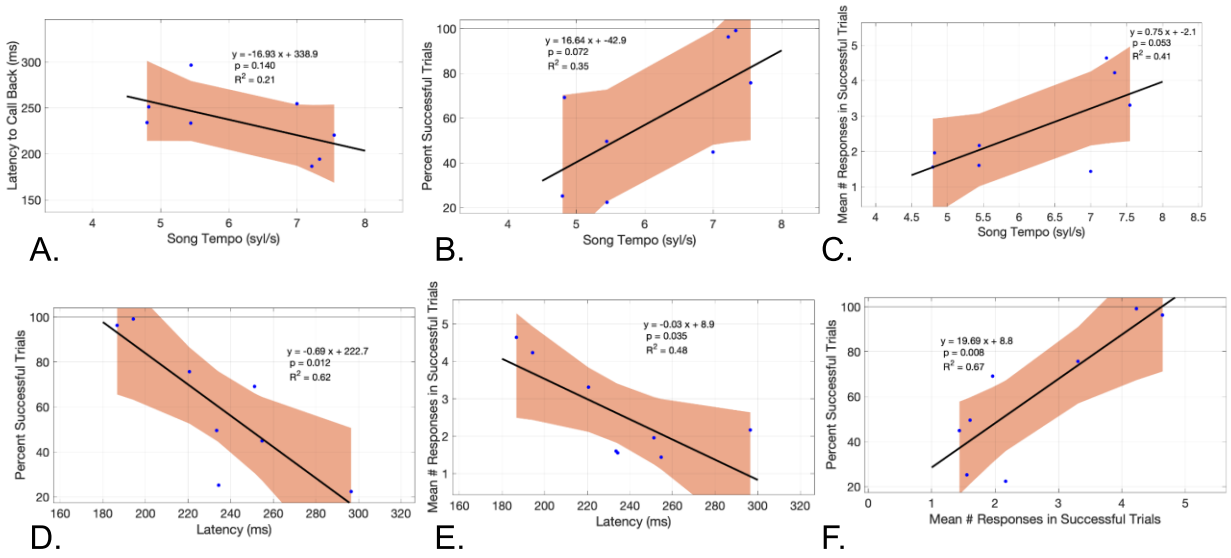
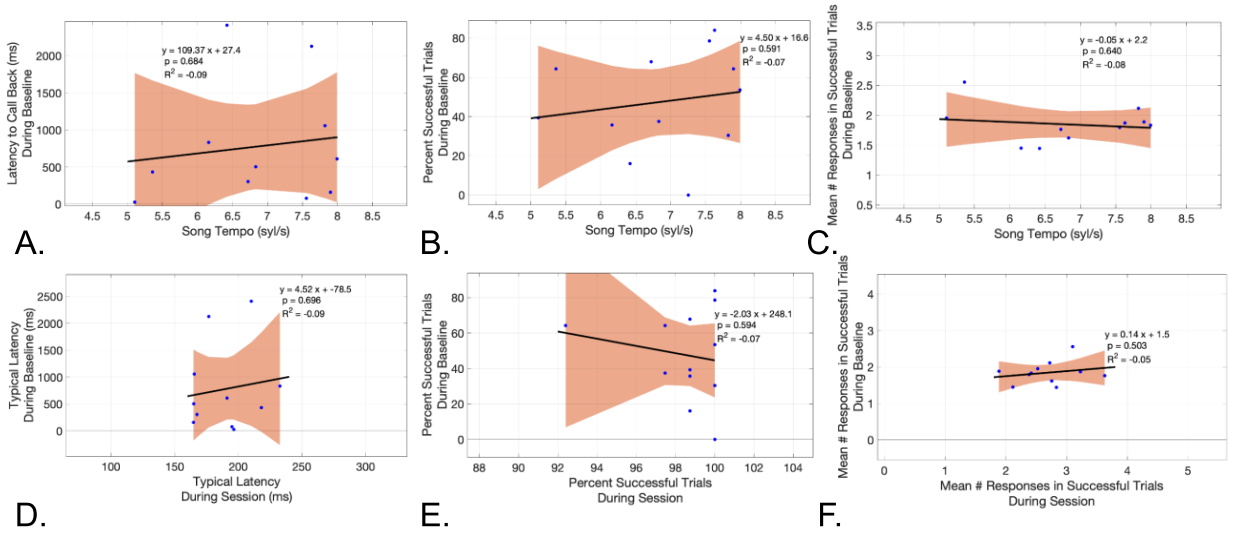


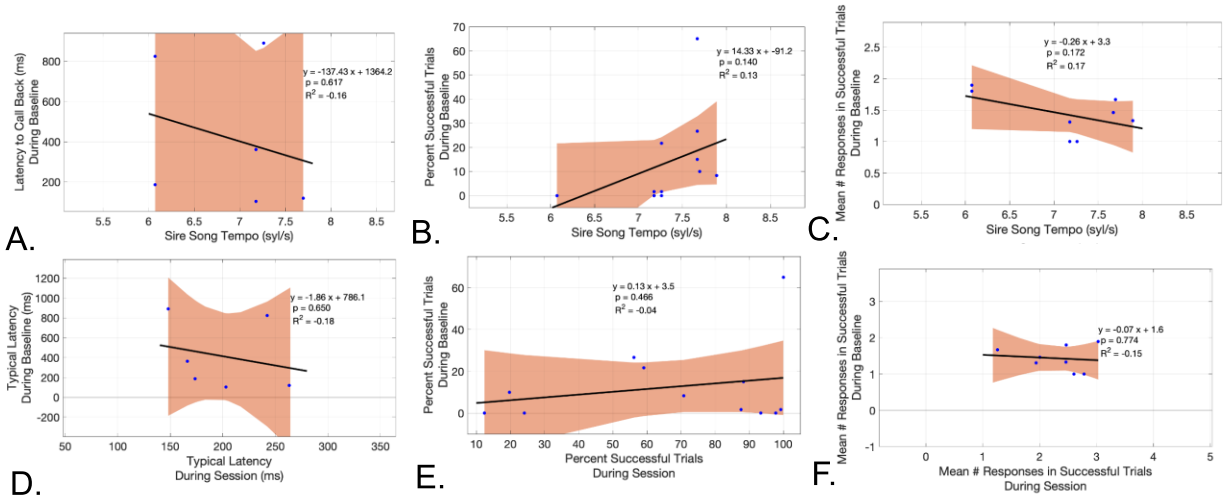
Figure 2.13. Call features appear to correlate with one another in non-tutored females (Non-Tutored Females, Experiment 2). A. No significant correlation between typical latency and percent successful trials (n = 4 individuals from 2 nests, slope = -0.29, R^2 = 0.31, p = 0.27). The 95% confidence intervals of the linear fit are shown in orange. B. Non-significant anticorrelation between typical latency and mean number of responses per successful trial (n = 4 individuals from 2 nests, slope = -0.003, R^2 = 0.75, p = 0.086). C. Non-significant correlation between mean number of responses per successful trial and percent successful trials (n = 4 individuals from 2 nests, slope = 100, R^2 = 0.81, p = 0.066).



Supplemental Figure 2.1. Results from Experiment 1 including only experimentally-naïve males. A. Non-significant negative correlation between typical call latency and mean song tempo (n = 8 individuals from 3 nests, slope = -16.93, $R^2 = 0.21$, p = 0.14). The 95% confidence intervals of the linear fit are shown in orange. B. Non-significant positive correlation between percent successful trials with mean song tempo (n = 8 individuals from 3 nests, slope = 16.64, $R^2 = 0.35$, p = 0.072). C. Non-significant positive correlation between mean number responses per successful trial with mean song tempo (n = 8 individuals from 3 nests, slope = 0.75, $R^2 = 0.41$, p = 0.053). D. Significant negative correlation between latency to respond with percent successful trials (n = 8 individuals from 3 nests, slope = -0.69, $R^2 = 0.62$, p = 0.012). E. Significant negative correlation between typical call response latency with the mean number of responses per successful trial (n = 8 individuals from 3 nests, slope = -0.03, $R^2 = 0.48$, p = 0.035). F. Significant positive correlation between mean number of responses in successful trials with percent successful trials (n = 8 individuals from 3 nests, slope = 19.69, $R^2 = 0.67$, p = 0.008).



Supplemental Figure 2.2. No relationship between spontaneous call features and song tempo (Males, Experiment 2). A. No correlation between typical "latency" during the baseline session (i.e., time of first call relative to start of trial) and mean song tempo (n = 11 individuals from 5 nests, slope = 109.4, R^2 = -0.09, p = 0.68). The 95% confidence intervals of the linear fit are shown in orange. B. No correlation between percent successful trials in the baseline session and mean song tempo (n = 12 individuals from 5 nests, slope = 4.50, R^2 = -0.07, p = 0.59). C. No correlation between the mean number of responses per successful trial in the baseline session and mean song tempo (n = 11 individuals from 5 nests, slope = -0.05, R^2 = -0.08 p = 0.64). D. No correlation between typical "latency" during the baseline session (i.e., time of first call relative to start of trial) with typical latency during the playback session (i.e., time of first call relative to call stimulus onset) (n = 11 individuals from 5 nests, slope = 4.52, R^2 = -0.09, p = 0.70). E. No correlation between percent successful trials during the baseline session with percent successful trials during the playback session (n = 12 individuals from 5 nests, slope = -2.03, R^2 = -0.07, p = 0.59). F. No correlation between mean number of responses per successful trial during the baseline session with mean number of responses per successful trial during the playback session (n = 11 individuals from 5 nests, slope = 0.14, R^2 = -0.05, p = 0.50).



Supplementary Figure 2.3. No relationship between spontaneous call features and song tempo (Females, Experiment 2). A. No correlation between typical "latency" during the baseline session (i.e., time of first call relative to start of trial) and sire's mean song tempo (n = 6 individuals from 4 nests, slope = -137.43, R^2 = -0.16, p = 0.62). The 95% confidence intervals of the linear fit are shown in orange. B. No correlation between percent successful trials during the baseline session and sire's mean song tempo (n = 12 individuals from 6 nests, slope = 14.33, R^2 = 0.13, p = 0.14). C. No correlation between mean number of responses per successful trial during the baseline session and sire's mean song tempo (n = 6 individuals from 4 nests, slope = -0.26, R^2 = 0.17, p = 0.17). D. No correlation between typical "latency" during the baseline session (i.e., time of first call relative to start of trial) vs. typical latency during the playback session (i.e., time of first call relative to call stimulus onset) (n=6 individuals from 4 nests, slope = -1.86, R^2 = -0.18, p = 0.65). E. No correlation between percent successful trials during the baseline session and percent successful trials during session (n=12 individuals from 6 nests, slope = 0.13, R^2 = -0.04, p = 0.47). F. No correlation between mean number of responses per successful trial during the baseline session and mean number of responses per successful trial during the playback session (n=8 individuals from 5 nests, slope = -0.07, R^2 = -0.15, p = 0.77).

CHAPTER 3: Features of spontaneous calls do not significantly cluster within-nest in females

3.1 Results

In Chapter 2, I described that call features in females do not cluster within-nest. This result was surprising, since genetic factors or nest-wide experiences (e.g., learning during early life) may be expected to explain behaviors that are highly stereotyped within-bird but variable across the population. I therefore examined whether call features may cluster within nest when calls were sampled more broadly from many birds across a variety of nests.

The lab maintains a database that contains hundreds of thousands of spontaneous vocalizations from a wide variety of nests. Typically, this database is used as a way to determine the sex of the birds, since males produce song and short calls while females produce no song and multi-lobed trills. However, this database can also be used as a source of vocalization data for many birds from many nests.

To determine whether features of spontaneous calls cluster within-nest, I analyzed spontaneous calling data from 29 females across 13 nests and quantified inter-lobe interval and inter-call interval. Inter-lobe interval refers to the time between trills within a single call (Figure 3.1 A, see Methods) and inter-call interval refers to the time between calls (Figure 3.1 B, see methods). The distributions of both inter-lobe interval and inter-call interval had a clearly identifiable peak, which was termed the typical inter-lobe interval or typical inter-call interval respectively (Figure 3.1 C-D). There was no statistically significant within-nest clustering of inter-lobe interval or inter-call interval (Figure 3.1 E-H).

These results provide further support for the finding that calling behavior in females - be it elicited or spontaneous - cannot be explained by nest-wide phenomena such as genetic background or early-life learning. Given that female calling behavior is highly stereotyped

within-bird and variable across the population, these behaviors may be underlaid by highly individual-specific factors that warrant further investigation.

3.2 Methods

3.2.A Subjects

Data were analyzed from 29 naturally-raised female Bengalese finches (*Lonchura striata domestica*) from 13 nests in early adulthood (~ 90 dph). All birds were bred in-house.

3.2.B Audio recording

Vocalizations were recorded in early adulthood (~90 dph) from birds temporarily singly-housed in sound isolation chambers. Audio was recorded in a custom-written Labview program (National Instruments; sampling frequency of 32 kHz) using a microphone (Countryman) and AM Systems pre-amplifier. Audio files were recorded when an amplitude threshold was crossed multiple times within a 6-second window. Audio from the 2-seconds prior to the amplitude threshold crossing until the 1-second after noise cessation was saved.

3.2.C Call labeling

Putative call files were read into a custom MATLAB program that created an amplitude envelope by rectifying and smoothing the audio waveform. Amplitude threshold crossings of more than 5 ms were identified as putative calls, which were then confirmed by eye to contain call-like spectral information. Each lobe of multi-lobed calls was labeled (e.g., a two-lobed call was labeled as ‘dd’ and a three-lobed call was labeled as ‘ttt’). As described previously (Yoneda and Okanoya, 1991), multi-lobed calls were clearly distinguishable from multiple single-lobed calls (see Figure 3.1 A-B). Most notably, inter-lobe intervals were very short (median of 46.4 ms, maximum value of 99.2 ms) compared to inter-call intervals (median of 802 ms, minimum value

of 366 ms); this effect was highly significant (Wilcoxon rank-sum test, p = 1.71 E-97). An average of 170.3 ± STD 97.3 calls was analyzed per bird.

3.2.D Analysis

Inter-Lobe Interval

The inter-lobe interval was found for each call with more than one lobe. First, a bandpass filter (FIR filter with Hanning window, passband between 0.5 kHz and 9.5 Hz) was applied to the section of audio associated with the call. The audio was then rectified and smoothed with a 1-ms moving-average filter. Finally, the largest peak (at a non-zero lag) of the autocorrelation was found; the time value associated with this peak was determined to be the inter-lobe interval for the call. To find the typical inter-lobe interval, a histogram of inter-lobe intervals with 1-ms bins was fit with spline interpolation. The inter-lobe interval value associated with the peak of this fit was defined as the typical inter-lobe interval. Only birds with more than 30 inter-lobe intervals were included in analysis; this ensured that a clear inter-lobe interval peak was identified.

Inter-Call Interval

Inter-call interval was calculated as the time between the onset of the first lobe of a given call to first lobe of the subsequent call. To find the typical inter-call interval, a histogram of inter-call intervals with 200-ms bins was fit with spline interpolation. The inter-call interval value associated with the peak of this fit was defined as the typical inter-call interval. Only birds with more than 30 inter-call intervals were included in analysis; this ensured that a clear inter-call interval peak was identified.

Within-Nest Clustering

Monte Carlo simulations were used to determine whether each call metric clustered within-nest. The observed data consisted of the call metric for each bird (e.g., inter-call interval for each bird)

and the nest identity of each bird. To determine the extent to which the call metric clustered within-nest, the distance between each bird's data point and the mean for the nest was calculated. Then, the distance of each bird's metric from the nest mean was averaged across the population to identify the observed average within-nest distance. To determine the clustering of the call metric by chance alone, the relationship between call metric and nest identity was shuffled 10,000 times. In each of the 10,000 simulations, the within-nest distance (using the shuffled data-nest relationships) was calculated. A call metric was determined to cluster by nest if the chance-level average within-nest distance was higher than the observed average within-nest clustering more than 95% of the time.

3.3 Figures

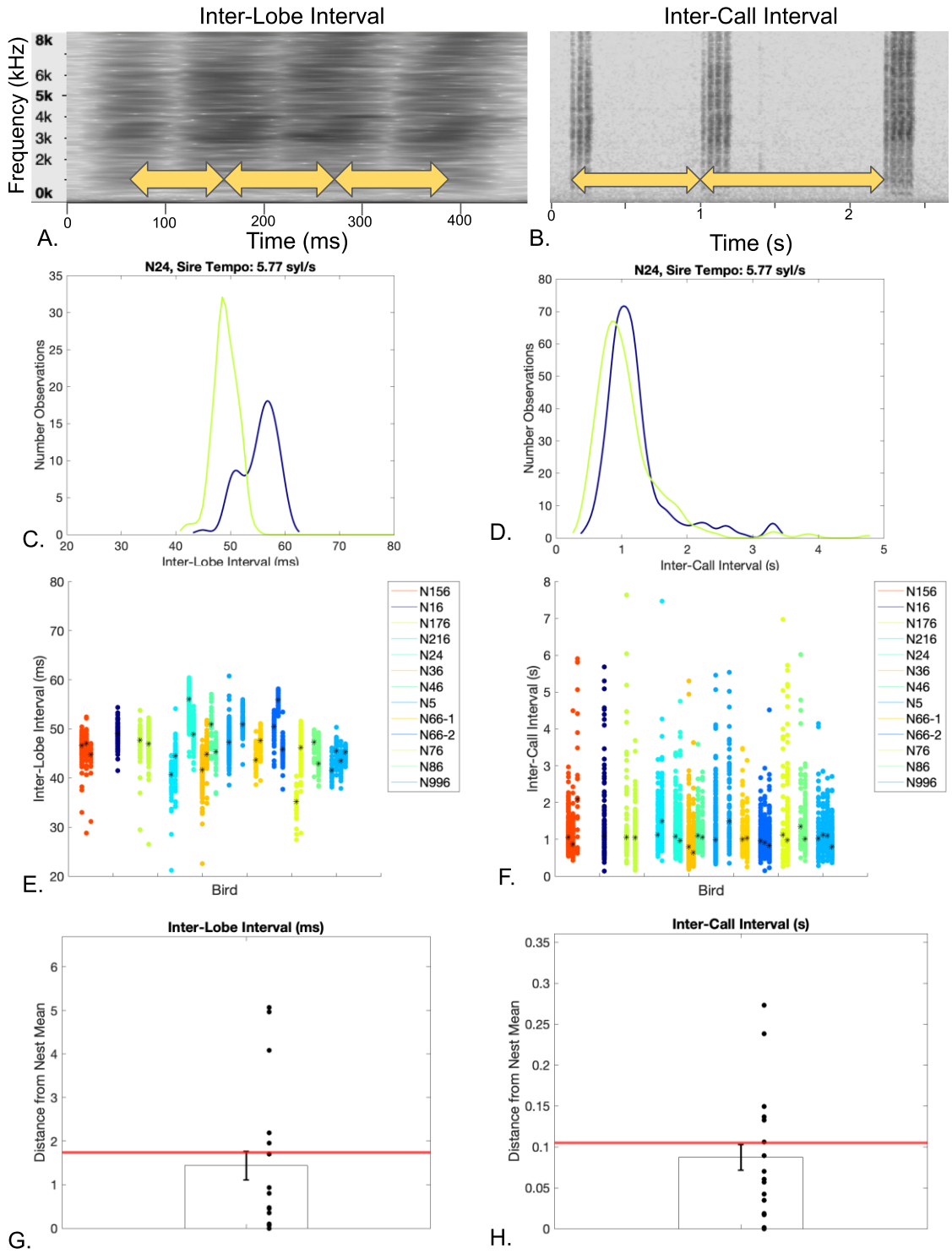


Figure 3.1. Spontaneous inter-lobe interval and inter-call interval in females does not cluster within-nest. A. Representative spectrogram showing a four-lobed call. Yellow arrows indicate

inter-lobe intervals, calculated as the time from the peak of one call lobe to the peak of the next call lobe (see Methods). B. Representative spectrogram showing a call train. Yellow arrows indicate inter-call intervals, calculated as the time between call onsets. C. Representative histograms (spline fit) of inter-lobe interval for two birds from one nest (in this representation, each bird in the nest was given a unique color identifier). The typical inter-lobe interval for each bird was defined as the peak of the spline fit. D. Same as C, but for inter-call interval. E. Inter-lobe intervals for each bird, color-coded by nest (n = 29 individuals from 13 nests). The median is plotted as a black asterisk for each bird. F. Inter-call intervals for each bird, color-coded by nest (n = 29 individuals from 13 nests). The median is plotted as a black asterisk for each bird. G. No within-nest clustering of inter-lobe ($p = 0.16$, one-tailed t-test). H. No significant within-nest clustering of inter-call intervall ($p = 0.097$, one-tailed t-test). For plots in G and H: Each black data point indicates the difference between the bird's value for the call metric and the mean value of the metric within the bird's nest. The red line shows the average distance from the nest mean by chance across 10,000 simulations. All bar plots display mean ± SEM.

CHAPTER 4: HVC is tuned to encode sounds presented at particular tempos

4.1 Results

The sensorimotor nucleus HVC plays an important role in song learning (Roberts et al., 2012; Vallentin et al., 2016) and song timing (Hahnloser et al., 2002; Long and Fee, 2008; Zhao et al., 2019). Across species, neural activity in the HVC of adult male birds increases selectively in response to playback of that bird's song (McCasland and Konishi, 1981; Margoliash, 1983; Margoliash, 1986; Margoliash and Fortune, 1992). In male Bengalese finches, this tuning is sensitive to the temporal structure of song playback (Bouchard and Brainard, 2016). Male Bengalese finches are constrained by a genetic predisposition to produce song at a particular tempo (Mets and Brainard, 2018) and they learn best from songs produced at this tempo (Mets and Brainard, 2019). One proposed mechanism for this predisposition is that HVC excitability is genetically determined (Mets et al., 2021). Natural or artificial changes to HVC activity or excitability drive changes in tempo (Long and Fee, 2008; Aronov and Fee, 2012; Mets et al., 2021) but whether innate differences in HVC excitability drive learning of song sung at particular tempos is unknown. If this was so, we may expect HVC to be preferentially activated by sounds produced at particular tempos, thereby enhancing encoding of this sound. To examine this hypothesis, these experiments tested whether the HVC of an adult male Bengalese finch was preferentially activated by sound presented at tempos that are similar to the bird's song tempo.

4.1.A. Validation of HVC's preferential response to playbacks of the bird's own song (BOS)

Previous literature has described that HVC activity is higher during playback of BOS than during playback of temporally reversed BOS (rBOS), conspecific songs, heterospecific song, or other natural sounds (McCasland and Konishi, 1981; Margoliash, 1983; Margoliash, 1986; Margoliash and Fortune, 1992). To replicate this result in our own hands, multi-unit activity (MUA) was

recorded from the HVC of an awake male Bengalese finch implanted with a multi-electrode array. Response strength was found by subtracting the baseline activity spike rate from the spike rate during stimulus playback (see Methods). As expected, preferential activation of HVC by BOS and not rBOS was observed (Figure 4.1). This preferential activation to BOS syllables over rBOS syllables was statistically significant at both Site 1 ($p = 5.47E-45$, t-statistic = 22.01, n = 63 BOS trials, n = 65 rBOS trials, two-tailed t-test) and Site 2 ($p = 1.41E-27$, t-statistic = 15.11, n = 19 BOS trials, n = 18 rBOS trials, two-tailed t-test). Also in line with previous work (Markowitz et al., 2015), the local field potential (LFP) showed a strong response to BOS. A strong LFP response to rBOS was also observed (Supplementary Figure 4.1). This strong response to rBOS is likely due to the fact that LFP reflects contributions from a much wider range of neural regions than MUA. While HVC responds only weakly to rBOS, nearby regions respond robustly to BOS and rBOS (Janata and Margoliash, 1999; Lewicki and Arthur, 1996). To quantify the LFP response to BOS vs. rBOS playback, a stimulus-triggered average aligned to the onset of BOS or rBOS syllables was created and the distance between the voltage peak and voltage trough during each syllable presentation was calculated (see Methods). At each of three sites, the peak-to-trough distance was much greater during BOS syllable playback than during rBOS syllable playback (Supplementary Figure 4.2).

4.1.B. HVC tuned to BOS tempo and similar tempos

To examine whether HVC is sensitive to the global tempo of BOS playback, a series of stimuli were created in which the inter-syllable gap (ISG) length was shortened or lengthened, thereby changing song tempo without altering syllable spectral information (Figure 4.2 A; see Methods). Visual inspection of spike histograms suggested that HVC activity during BOS syllables was highest when the playback tempo was similar to the naturalistic song tempo (Figure 4.2 B).

Indeed, plotting the response strength against BOS tempo revealed a tuning curve centered around the naturalistic BOS tempo (Figure 4.2 C; Supplementary Table 4.1). This tuning curve became particularly clear after controlling for the length of the stimulus (Figure 4.2 D; Supplementary Table 4.1). Notably, HVC MUA activity fell off more sharply when ISGs were shortened by a given percentage vs. lengthened by that same percentage (Figures 4.2 C-D). This may be because shortening ISGs by a given percentage causes more of a tempo change than lengthening ISGs by that same percentage; for instance, shortening ISGs by 50% resulted in a tempo increase of 1.9 syl/s, while lengthening ISGs by 50% resulted in a tempo decrease of 1.2 syl/s. This is because shortening ISGs by 50% results in a larger change in [Syllable Length : (Syllable Length + Gap Length)] ratio than does increasing ISGs by 50%.

LFP responses to BOS and tempo-manipulated BOS were also examined (Supplementary Figure 4.3). The peak-to-trough distance of the LFP trace during syllable presentation was not equal across stimuli. Rather, a weak tuning curve centered at the naturalistic BOS tempo was observed (Supplementary Figure 4.4; Supplementary Table 4.2). LFP likely reflects the congregate output of many neural networks rather than the activity of HVC alone, and some of these networks are less specific to temporal structure than HVC (Janata and Margoliash, 1999; Lewicki and Arthur, 1996). This may explain why the LFP-derived tuning curve was similar to, but less clear than, the MUA-derived tuning curve.

4.1.C. HVC tuned to bursts of white noise presented at 4-5 Hz

To evaluate whether HVC responds preferentially to particular tempos, even absent of the spectral information typical of song, MUA and LFP were recorded while spectrally-neutral "syllables" were presented. Specifically, 50-ms bursts of white noise were interspersed with periods of silence lasting between 41-ms and 950-ms, thereby creating a stimulus set with

tempos between one white noise “syllable” per second and 11 white noise “syllables” per second. The 50-ms duration was chosen because this was similar to the length of a typical song syllable (mean syllable length = 78 ms). Two stimulus sets were created: Stimulus Set 1: {White Noise “Syllables” at 1 Hz, 2 Hz, 3 Hz, …, 9 Hz}; Stimulus Set 2: {White Noise “Syllables” at 1 Hz, 4 Hz, 5 Hz, 6 Hz, …, 11 Hz}. These stimulus sets were chosen to fully sample tempos in the range of BOS global tempo (7 syl/s) and BOS introductory note tempo (5.3 syl/s). MUA and LFP were recorded from two sites in HVC during playback of these stimulus sets (one unit per stimulus set). Recording bulk activity in HVC through MUA or LFP is ideal for evaluating the efficacy of multiple auditory stimuli in eliciting HVC activity, as these techniques avoid biasing to particular HVC subpopulations (Volman, 1993).

HVC responded robustly to playbacks of white noise bursts at Site 1 (Figure 4.3). MUA was time-locked to the onset of white noise “syllables.” This was evident by eye (Figure 4.3 B), and power spectral density plots of spike trains showed peaks of power that matched the stimulus playback tempo (Supplementary Figure 4.5; see Methods). HVC activity was not equal across all white noise stimuli. Rather, plotting stimulus tempo by response strength revealed a tuning curve centered around the 5-Hz stimulus (Figure 4.3 C; Supplementary Table 4.3). To control for the fact that each stimulus contained a different number of white noise syllables, and that HVC activity decayed with subsequent syllable presentations (e.g., Figure 4.4 B) only the first 5 white noise “syllables” were included in a second analysis. Additionally, to control for the fact that each stimulus had a different ISG length, only the section of each ISG corresponding to the shortest ISG length across all stimuli was analyzed. After correcting for the number of white noise “syllables” and the ISG length, the general shape of the tuning curve remained (Figure 4.3 D; Supplementary Table 4.3).

In a separate site, which was recorded from on a different day using Stimulus Set 2 (Figure 4.4 A), HVC activity was again time-locked to the onset of white noise syllables (Figure 4.4 B). HVC activity was tuned to the 4-Hz white noise stimulus (Figure 4.4 C; Supplementary Table 4.4). After correcting for the number of white noise syllables, selective activation to 4-Hz and 5-Hz white noise stimuli was observed (Figure 4.4 D; Supplementary Table 4.4).

Examination of the LFP trace also revealed strong activation during white noise playback at both recording sites (Supplementary Figures 4.6 - 4.7). Analyzing the peak-to-trough distance of the LFP trace during white noise syllables presented at each tempo revealed a tuning curve centered at 5 Hz at Site 1 (Supplementary Figure 4.8; Supplementary Table 4.5) and 4 Hz at Site 2 (Supplementary Figure 4.9; Supplementary Table 4.5). These LFP results are highly consistent with the MUA results.

The mean song tempo for this bird was 7.0 syl/s (STD = 0.55 syl/s, n = 54 songs). If HVC was tuned to encode tempos similar to BOS tempo, we may expect the tuning curve to be centered around 7 Hz rather than 5 Hz. However, the mean tempo of the introductory notes was 5.3 syl/s (STD = 0.2 syl/s, n = 11 songs) and the largest peak in the power spectrum of the syllable-gap temporal pattern (see Methods) was found at 4.4 Hz (Supplementary Figure 4.10). Temporal tuning to the introductory note tempo may be advantageous for the learning and production of song. Introductory notes have been suggested to act as an “alerting signal” to indicate the onset of an important stimulus to a receiving organism (Richards, 1981; Ord and Stamps, 2008) and to prepare the singer’s motor circuitry for song (Rajan, 2018). The number of introductory note repeats (as well as their acoustic structure) is learned from the tutor (Kalra et al., 2021). If, during learning, the tutee’s HVC is tuned to encode the tutor’s introductory note tempo, this may enhance learning of tutor song.

4.2 Methods

4.2.A. Subjects

The subject was a single adult Bengalese finch male who was naïve to experimentation prior to implantation of a custom-built tungsten tetrode array integrated into a shuttle drive. The shuttle drive allowed the tetrodes to be raised and lowered following implantation. The shuttle drive was attached to a connector (Omnetics), which was in turn affixed to the female side of a RHS2116 headstage (Intan Technologies).

During surgery (conducted by Kurtis Swartz), the bird was sedated with ketamine and midazolam and anesthetized with 0.5% - 3% isoflurane in oxygen using a non-rebreathing anesthesia machine (VetEquip). After removing feathers and skin from the head, the top layer of skull was removed at the branchpoint of the midsagittal sinus. Y0 was then located and used as the reference point for HVC (0.4 mm anterior to and 2.5 mm lateral of Y0). Both layers of skull and the dura were removed around the HVC coordinates. Multiple recordings were made from the surrounding region, moving in 150 μm steps in the AP and ML directions, and lowering a 1 MΩ electrode (Carbostar) from -150 μm to -650 μm DV. The electrode was considered to be in HVC if characteristic bursting activity was observed. After HVC was fully mapped in this manner, the microelectrode array was placed to hover at the brain surface. A ground wire was implanted over the cerebellum, under the skull but above the dura. The Omnetics connector was affixed to the skull with dental cement. The bird was allowed to recover from surgery for 2 weeks prior to all experimentation and was habituated to the recording set-up by plugging and unplugging the implanted connector to the headstage. The RHS2116 headstage contains stimulator/amplifier channels, and the bird was used in HVC stimulation experiments prior to the

experiments described here. At the time of the experiments described here, the bird was between 297 dph and 364 dph.

4.2.B. Song collection

At 212 dph, the bird was placed in a sound isolation chamber (Acoustic Systems) for one day. Audio was recorded via a custom-written LabView program (sampling rate of 32 kHz) whenever the bird vocalized or otherwise made noises that crossed an empirically determined threshold multiple times within a 6-second window. Audio from the 2 seconds prior to the amplitude threshold crossing and 1 second after noise cessation were recorded and saved.

4.2.C. Stimuli

Bird's own song (BOS) and reverse BOS (rBOS)

The bird's own song (BOS) stimulus was selected from the audio files collected from the experimental bird (see Song Collection above). A custom program written in MATLAB was used to remove non-song audio files (consisting of calls or cage noise) from consideration. Song files were then examined to find a representative song file. The selected song was 12.9-seconds long and had a tempo of 6.2 syllables per second (syl/s), and the mean tempo across 54 songs was 7.0 syl/s.

The song file was read into a custom MATLAB program that created an amplitude envelope by rectifying and smoothing the audio waveform. Amplitude threshold crossings of more than 10 ms were identified as putative syllables, which were then confirmed by eye to contain song-like spectral information. After identifying all syllables within the song bout (separated from one another by no more than 250 ms), the audio between syllables was replaced with silence. This was done to ensure that no cage noise or other background noise was played back to the bird. This silent-gap version of BOS was designated as the BOS stimulus for

playback. To create the temporally reversed BOS (rBOS) stimulus, the silent-gap BOS audio file was reversed.

Two versions of BOS were used: in Stimulus Set 1 of white noise playback experiments, a shortened, 4.5-second version of BOS was used. This version was not truncated in the middle of a syllable nor in the middle of a song motif. A shortened version of rBOS was made by reversing this BOS audio file. In tempo-manipulated BOS experiments and in Stimulus Set 2 of white noise playback experiments, the full 12.9-second version of the BOS file was used. These and all stimuli contained a 25-ms high-amplitude "trigger" as a second channel of audio, which was used to start and stop electrophysiological recordings (see below).

BOS with shortened or lengthened inter-syllable gaps

To create faster and slower versions of BOS without altering the spectral content of syllables, the inter-syllable gaps (ISGs) were shortened or lengthened. For instance, by reducing the length of each ISG by 25%, the tempo of the song was 7.0 syl/s rather than 6.2 syl/s. Another benefit of this approach was that it mirrors naturalistic changes in song tempo, since circadian, social-context-induced, or age-related changes in song tempo are largely due to changes in ISG length (Cooper et al., 2012; Glaze and Troyer, 2006). A total of eight tempo-manipulated BOS stimuli were created: 25% shorter ISGs, 50% shorter ISGs, 75% shorter ISGs, 100% shorter ISGs (i.e., ISGs deleted), 25% longer ISGs, 50% longer ISGs, 75% longer gaps, and 100% longer gaps (i.e., ISG length doubled). Syllable onsets and offsets were labeled for later use in analysis. The shortest stimulus (BOS with 100% shorter ISGs) was 6.8 seconds long and the longest stimulus (BOS with 100% longer ISGs) was 18.9 seconds long. Unmanipulated versions of BOS and rBOS were presented in this stimulus set, to ensure that the recorded cells had a reliable response to BOS and not to rBOS.

White noise

Audio files containing bursts of white noise were created by creating a vector of randomly sampled values between -1 and 1 that lasted 50 ms at a 32 kHz sampling rate. These 50-ms white noise bursts were separated by periods of silence. Eleven audio files containing bursts of white noise were created: bursts of white noise presented at 1 Hz, 2 Hz, 3 Hz, …, 11 Hz. The length of each file in the first stimulus set (1-9 Hz) was between 4.5 seconds and 5 seconds, thereby approximately matching the length of the BOS file in that stimulus set. The length of each file in the second stimulus set (1 Hz, 4-11 Hz) was between 12.5 seconds and 13 seconds, thereby approximately matching the length of the BOS file in that stimulus set. White noise burst onsets and offsets were labeled for use in analysis. Unmanipulated versions of BOS and rBOS were presented in these stimulus sets, to ensure that the recorded cells had a reliable response to BOS and not to rBOS.

4.2.D. Electrophysiology and stimulus playback

The bird was placed in a wire cage within a sound isolation chamber (Acoustic Systems) and the Intan headstage was connected via a tether to a USB interface board (Intan Technologies). This allowed the bird to move and fly within the cage while neural activity was recorded. The light was turned off in the sound isolation chamber, which greatly reduced the bird's movement. A microphone (Countryman) was placed atop the cage to monitor stimulus playback and to ensure that singing bouts were excluded from analysis.

Stimuli were played via a speaker approximately 25 cm from the bird's cage with a 10-second inter-trial interval. A custom MATLAB script was used to pseudorandomly select a stimulus from the stimulus set. Each stimulus contained two channels: one audio channel and one trigger channel that triggered the Intan RHX Acquisition software to save data. Specifically,

during stimulus creation, a vector of zeros was created of equal length to the given stimulus. Then, 25-ms high-amplitude pulse triggers were placed at the start and end of the vector to signify the beginning and end of the audio file, respectively. The audio was split such that the audio data was fed to the speaker while the trigger data was fed to the Intan acquisition board. Electrophysiological data (amplified from the multi-electrode array) and audio data (amplified from the microphone) starting from 5 seconds before the first trigger was received, and ending 5 seconds after the second trigger was received, were then saved. Each stimulus was played at least 10 times. Electrophysiological data were amplified with an AM Systems amplifier, digitized at 30 kHz, and collected with Intan RHX Acquisition software. During acquisition, data was notch-filtered to exclude 60-Hz noise.

4.2.E. Analysis

Data reading and filtering

Neural data, trigger data, trial time data, and audio data were read into MATLAB using the Intan RHS file reader (Intan Technologies). Audio was bandpass filtered (Butterworth) between 0.5 kHz and 9.5 kHz. Neural data was high-pass filtered to detect spikes and low-pass filtered to detect local field potential (see below). There was a single channel of neural data, as the other channels were nonfunctional. Neural data was shifted backward by 15 ms relative to audio data to account for the response delay, as has been done previously (Bouchard and Brainard, 2016).

Movement artifact detection

Movement artifacts were identified by large deviations in voltage. Any datapoint in the voltage trace that was less than -200 mV was flagged (this threshold was found empirically). If the total number of flagged data points during the stimulus playback exceeded 0.1% of the stimulus length, the file was excluded from analysis. Excluded files were then manually examined to

ensure that only files containing movement artifacts were excluded. A subset (50%) of included files were manually examined to ensure that files containing movement artifacts were successfully identified and removed. Movement artifacts were rare, and their presence could be readily confirmed via examination of audio recorded from the microphone (since most movement was associated with wing flaps, which also make noise).

Alignment of neural data to syllables and white noise bursts

Each vector of neural data contained pre-stimulus data, during-stimulus data, and post-stimulus data. The start of the stimulus playback was defined as time zero in the time vector obtained from the RHS file. Time zero was used to align neural data to the onsets and offsets of song syllables or white noise bursts. For instance, if BOS Syllable 1 occurred from 0 to 200 ms and Syllable 2 occurred from 400 to 550 ms, then neural data from 15 to 215 ms was considered as the response to Syllable 1 and neural data from 415 to 565 ms was considered as the response to Syllable 2 (recall that a response delay of 15 ms was applied).

Spike detection

Raw multi-unit neural data was high-pass filtered with a cut-off frequency of 250 Hz. Data was thresholded to detect events that deviated farther than -7 SDs from the mean voltage. To ensure that spikes were not double-counted, a datapoint was only counted as a spike if the preceding datapoint was higher than the -7 SD threshold.

Spike rate and response strength

Spike rate was defined as the number of detected spikes per second. Response strength was calculated by subtracting the baseline spike rate (i.e., spike rate during the 5 seconds preceding stimulus playback) from the spike rate during stimulus playback.

Local field potential (LFP)

Local field potential was isolated from the raw neural trace by downsampling from 32 kHz to 1 kHz and applying a filter with a passband between 1 Hz and 100 Hz.

Stimulus-triggered average (STA) LFP traces were created by aligning to the onset of song syllables or white noise "syllables." For white noise stimuli, the 15 ms to 65 ms after the onset of each white noise burst in the trial was isolated (recall that a response delay of 15 ms was applied). For white noise stimuli, only the first 10 white noise bursts in each trial (or the total number of white noise bursts per trial, whichever was lower) was included in analyses. For BOS and rBOS stimuli, LFP traces from 15 ms after the onset of each BOS syllable through 15 ms after the average BOS syllable duration (15 ms + 78 ms) was isolated (recall that a response delay of 15 ms was applied). Next, the traces from each stimulus condition were averaged to create the STA trace.

To quantify the change in voltage during the syllable presentation, the difference between the maximum voltage and the minimum was quantified for each syllable presentation (referred to as the peak-to-trough distance in the text). The peak-to-trough distance was then averaged across all presentations of the syllable for each stimulus type.

Power spectrum density of song

A power spectrum density plot was used to identify the tempos present in the song, following a protocol adapted from Kozhevnikov and Fee (2007). First, song was downsampled from 32 kHz to 1 kHz. Next, a binary vector was created to describe the temporal patterning of song syllables. For each data point in the song, a 1 was added to the vector if the point was part of a song syllable and a 0 was added if the point was part of an inter-syllable gap. Next, power at each frequency from 0 Hz to the Nyquist frequency (500 Hz) was found using a discrete Fourier

Transform with 0.08 Hz resolution. Finally, the power at each frequency was normalized by dividing the power vector by the maximum value in the vector.

Power spectrum density of spike trains

A vector describing the temporal patterning of spike times in each trial was made by making a histogram of spike times with 1-ms bins. Then the power spectrum density of this vector was calculated as described in the previous section. To obtain a single power spectrum density plot for each stimulus, the power spectrum density vectors for each trial within a given stimulus type was averaged.

Statistics

All comparisons between two groups were done via a two-tailed t-test. All comparisons between three or more groups were done via a one-way ANOVA with multiple comparisons using Bonferroni correction.

4.3 Figures

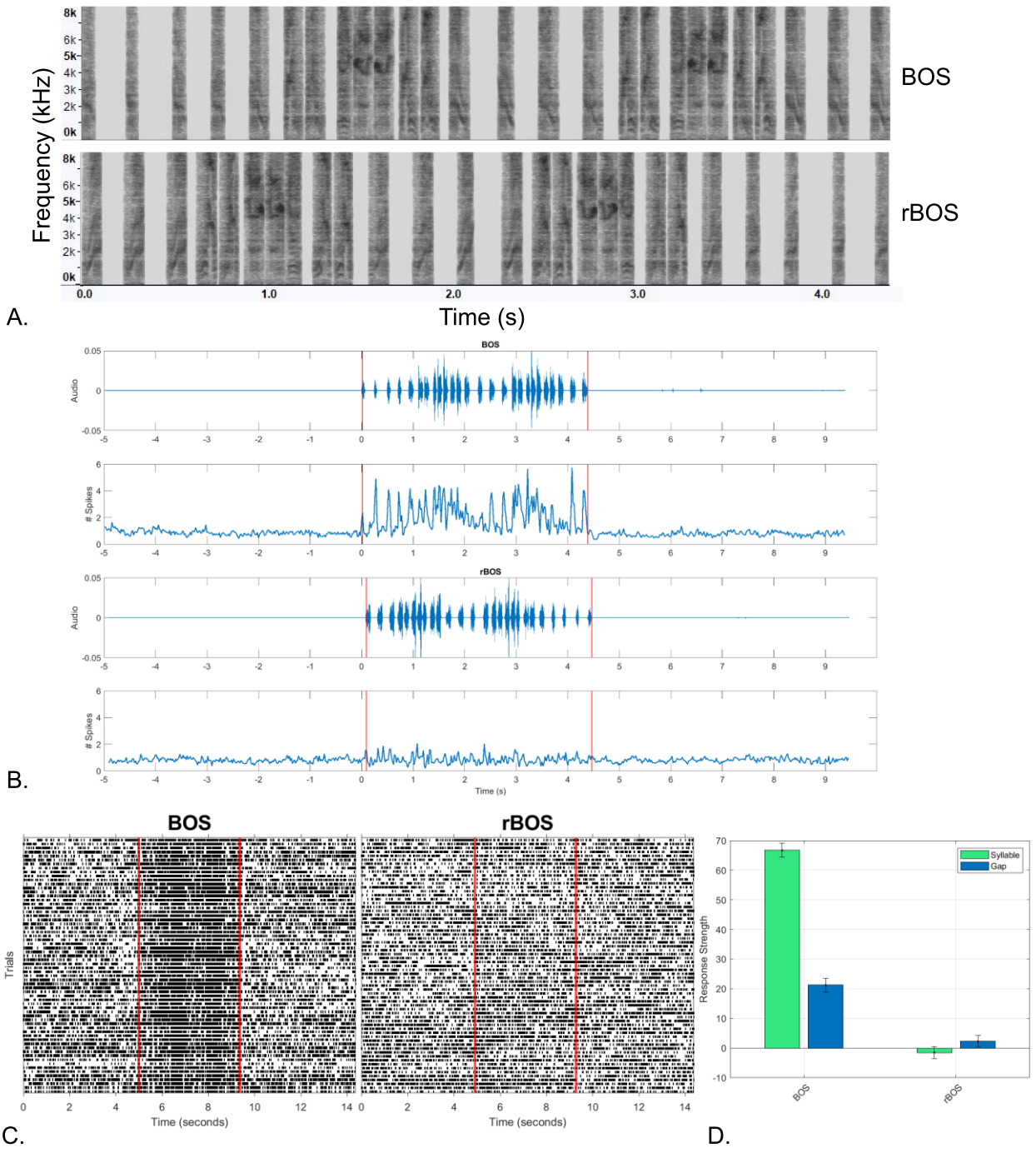


Figure 4.1. Multi-unit response to bird's own song (BOS) vs. temporally-reversed BOS (rBOS). A. *Upper panel:* Spectrogram of BOS with silent inter-syllable gaps (ISGs). *Lower panel:* Spectrogram of rBOS with silent ISGs. B. *Upper panels:* Average audio oscillogram and average spike histogram (bin size = 20 ms) during BOS playback. *Lower panels:* Average audio oscillogram and average spike histogram (bin size = 20 ms) during rBOS playback. Red lines indicate the start and end of stimulus playback. C. Spike raster plot from BOS trials (left) and

rBOS trials (right). Red lines indicate the start and end of stimulus playback. D. Response strength (spike rate during playback minus spike rate before playback) during syllables (green) and ISGs (blue) in BOS and rBOS trials. The response strength during BOS syllables was significantly higher than during rBOS syllables ($p = 5.47E-45$, t-statistic = 22.01, $n = 63$ BOS trials, $n = 65$ rBOS trials, two-tailed t-test comparing syllable response strength). All bar plots display mean ± SEM.

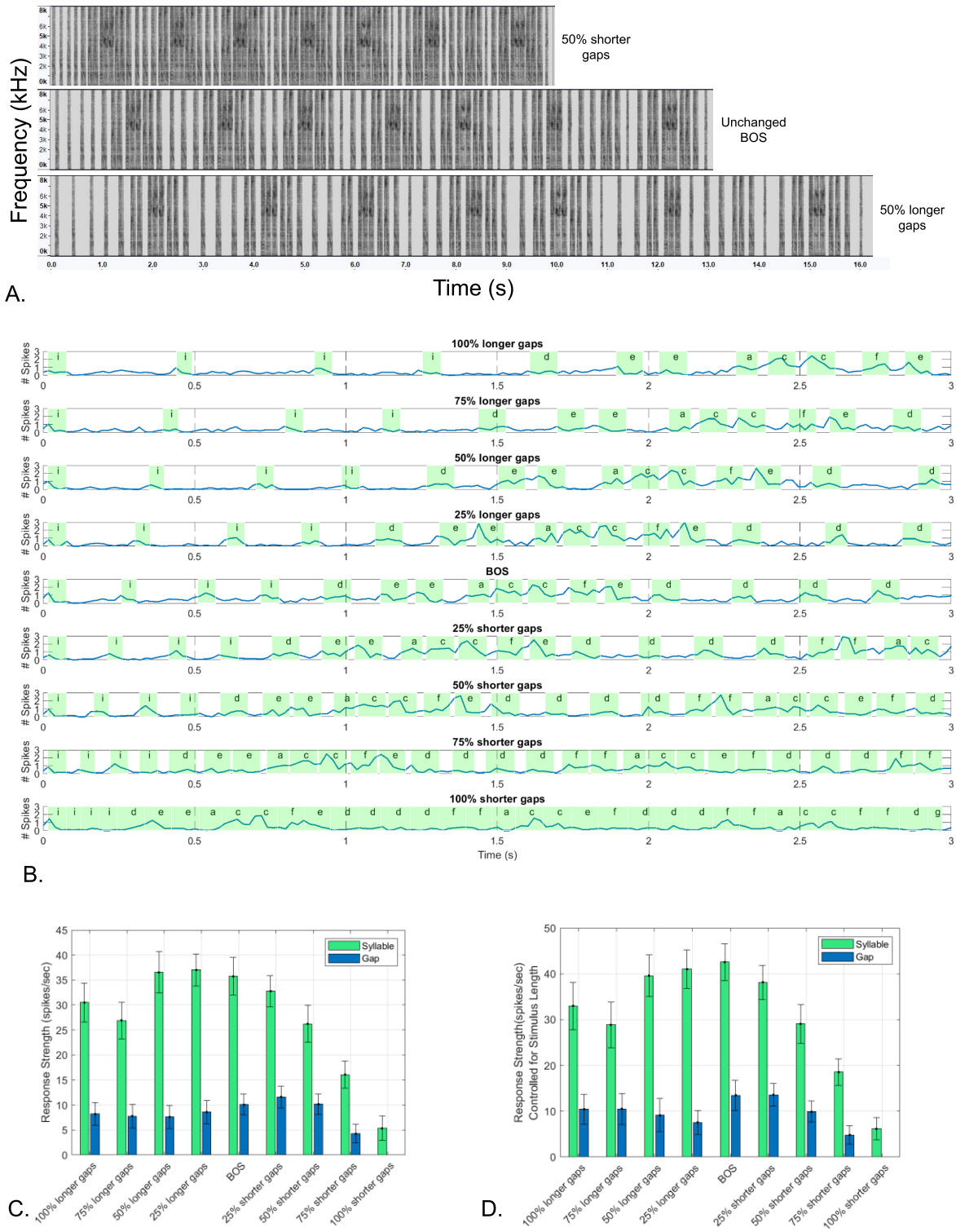


Figure 4.2. HVC is sensitive to the global tempo of BOS playback. A. Spectrogram of BOS with 50% shortened gap lengths (top panel), unchanged gap lengths (middle panel), and 50% lengthened gaps lengths (lower panel). B. Average spike histogram (bin size = 20 ms) during the

first 3 seconds of playback of tempo-manipulated BOS and unchanged BOS. Syllable playback is denoted by green boxes with the associated syllable label above. C. Response strength (spike rate during playback minus spike rate before playback) during syllables (green) and ISGs (blue) in tempo-manipulated BOS playback trials. The response strength was significantly different across stimulus playback condition ($p = 7.14$ E-9, $df = 8$, F-statistic = 7.7, one-way ANOVA comparing syllable response strength; See Supplementary Table 4.1 for multiple comparisons). D. Response strength during syllables (green) and ISGs (blue) when comparing the first 5 seconds of each stimulus. The response strength was significantly different across stimulus playback condition ($p = 2.19$ E-8, $df = 8$, F-statistic = 7.29, one-way ANOVA comparing syllable response strength; See Supplementary Table 4.1 for multiple comparisons). All bar plots display mean ± SEM.

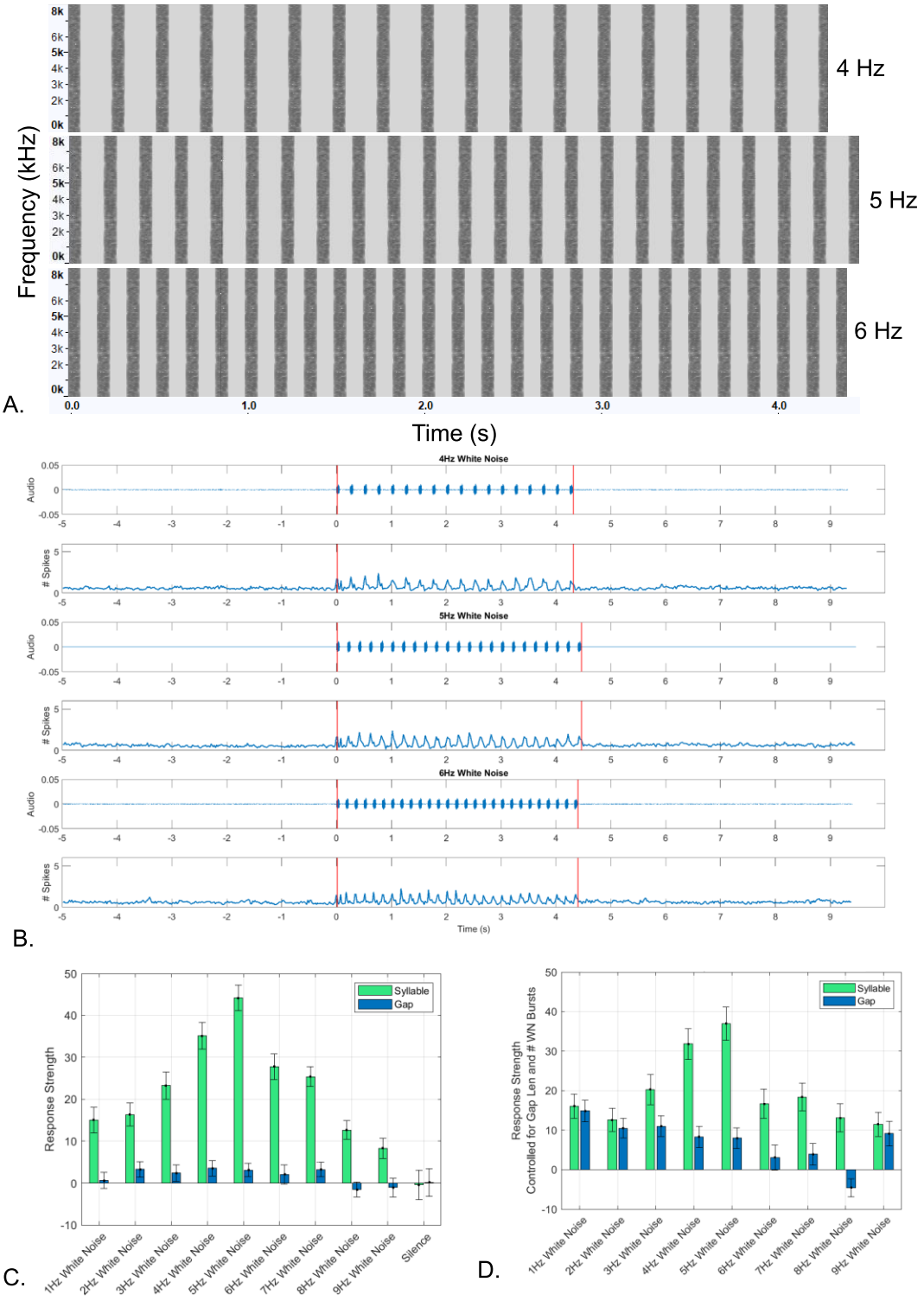


Figure 4.3. Tuning of HVC to 5 Hz tempo, multi-unit Site 1. A. Spectrogram of white noise bursts presented at 4 Hz (top panel), 5 Hz (middle panel), and 6 Hz (lower panel) with silent inter-syllable gaps (ISGs). B. Average audio oscillogram and average spike histogram (bin size = 20 ms) during playback of 4-Hz white noise (top panels), 5-Hz white noise (middle panels), and

6-Hz white noise (lower panels). Red lines indicate the start and end of stimulus playback. C. Response strength (spike rate during playback minus spike rate before playback) during syllables (green) and ISGs (blue) in white noise playback trials. The response strength was significantly different across stimulus playback condition ($p = 9.92E-22$, $df = 8$, F-statistic = 16.05, one-way ANOVA comparing syllable response strength; See Supplementary Table 4.3 for multiple comparisons). D. Response strength during syllables (green) and ISGs (blue) when controlling for the number of white noise bursts and ISG length across playback conditions. The response strength was significantly different across stimulus playback condition ($p = 9.14E-8$, $df = 8$, F-statistic = 6.22, one-way ANOVA comparing syllable response strength; See Supplementary Table 4.3 for multiple comparisons). All bar plots display mean ± SEM.

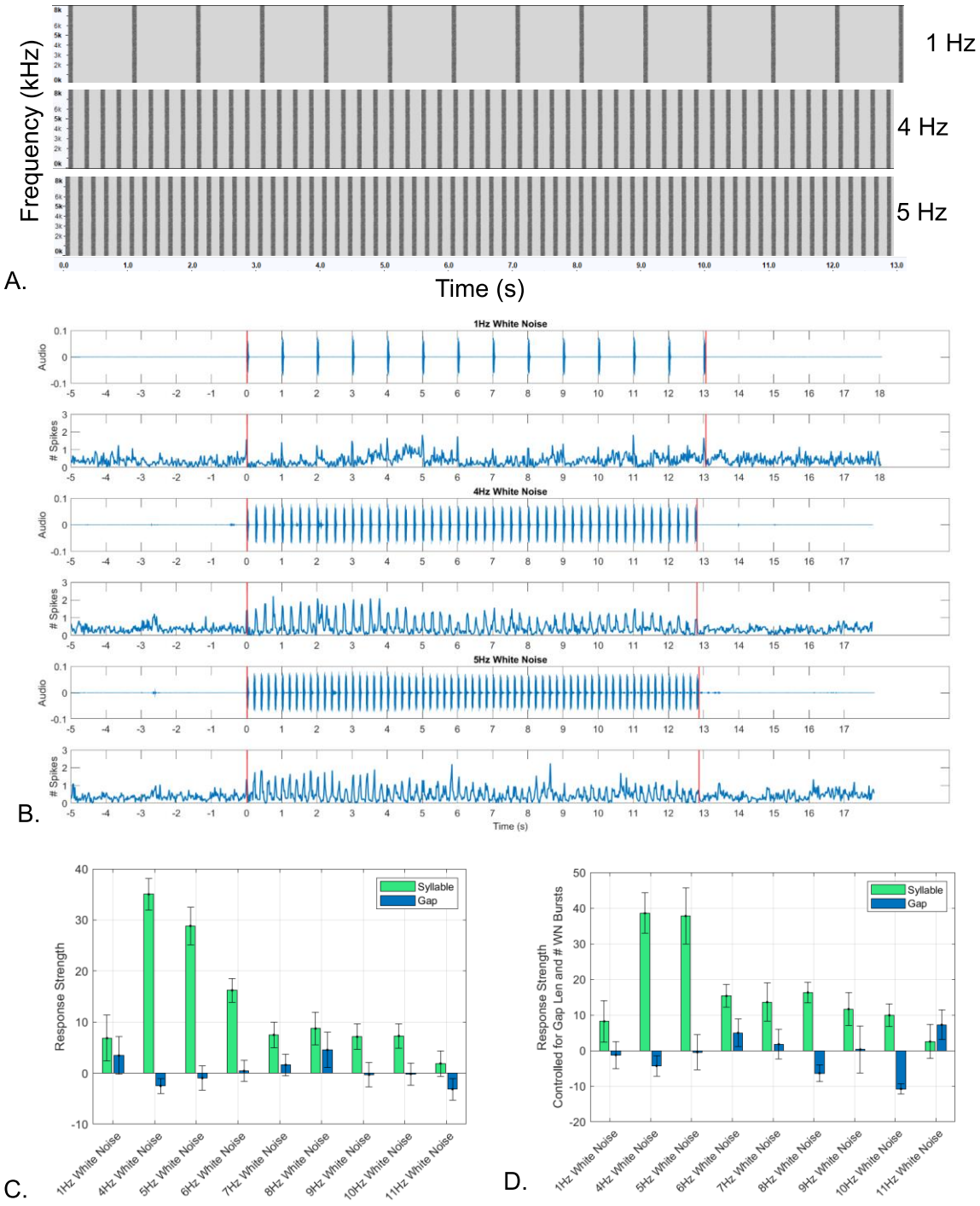
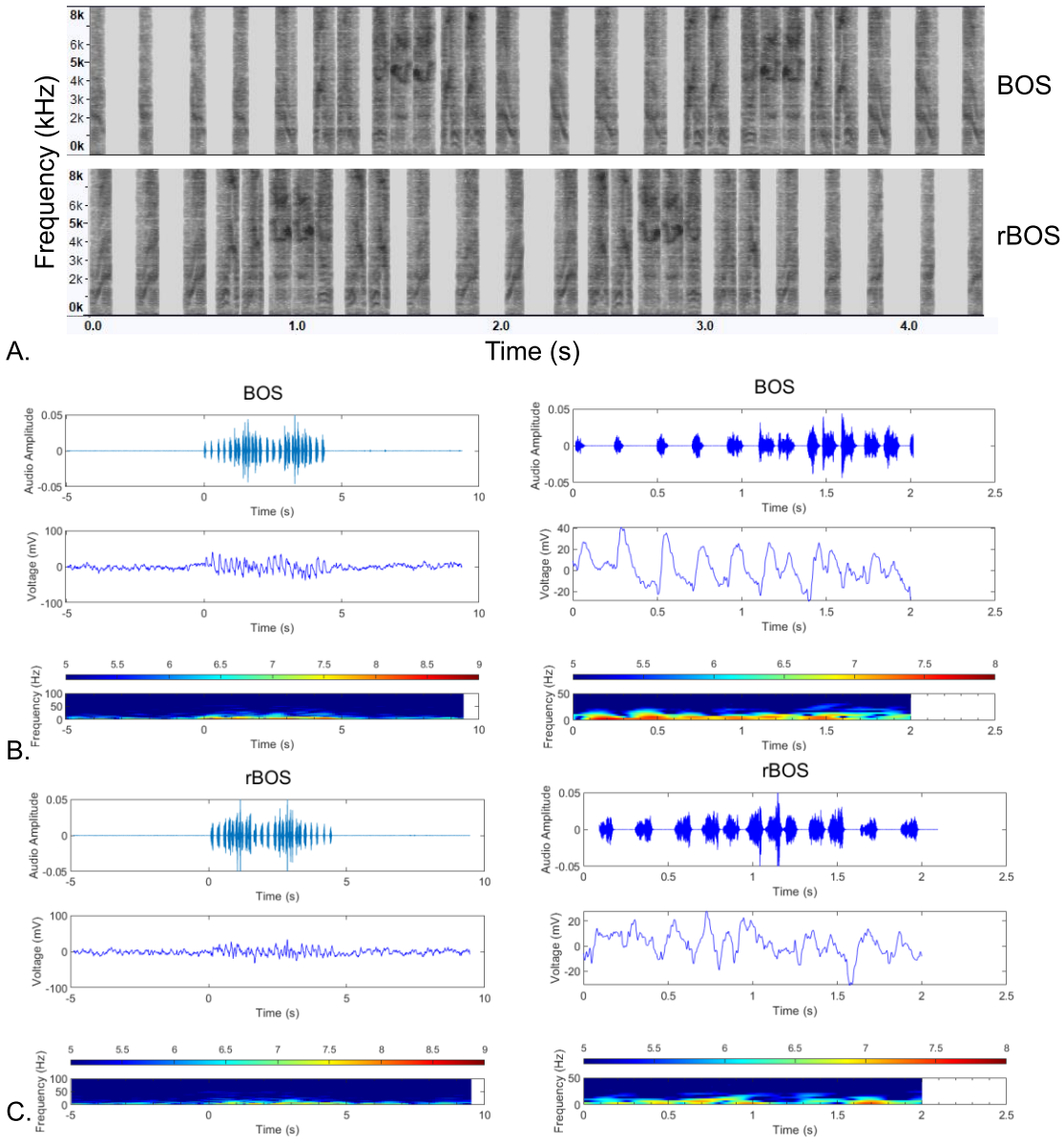
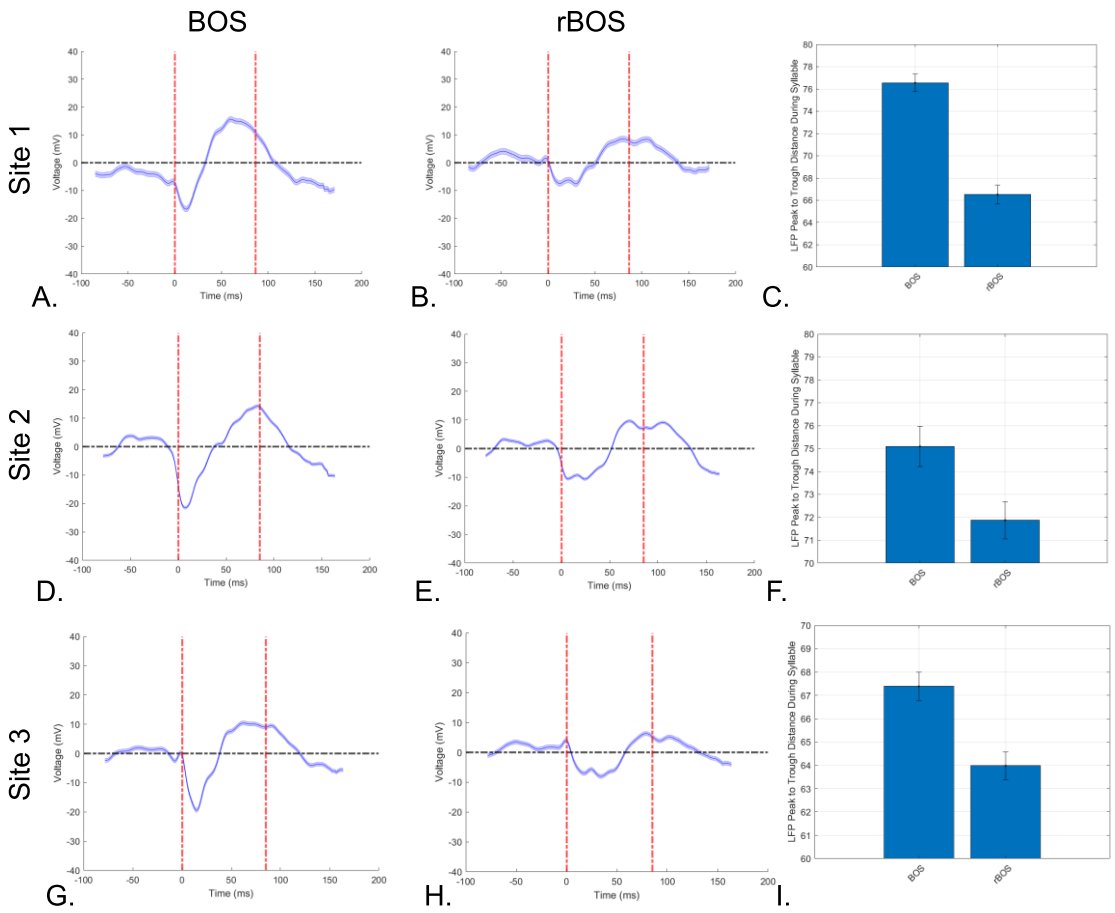


Figure 4.4. Tuning of HVC to 4-5 Hz tempo, multi-unit Site 2. A. Spectrogram of white noise bursts presented at 1 Hz (top panel), 4 Hz (middle panel), and 5 Hz (lower panel) with silent inter-syllable gaps (ISGs). B. Average audio oscillogram and average spike histogram (bin size = 20 ms) during playback of 1-Hz white noise (top panels), 4-Hz white noise (middle panels), and 5-Hz white noise (lower panels). C. Response strength (spike rate during playback minus spike

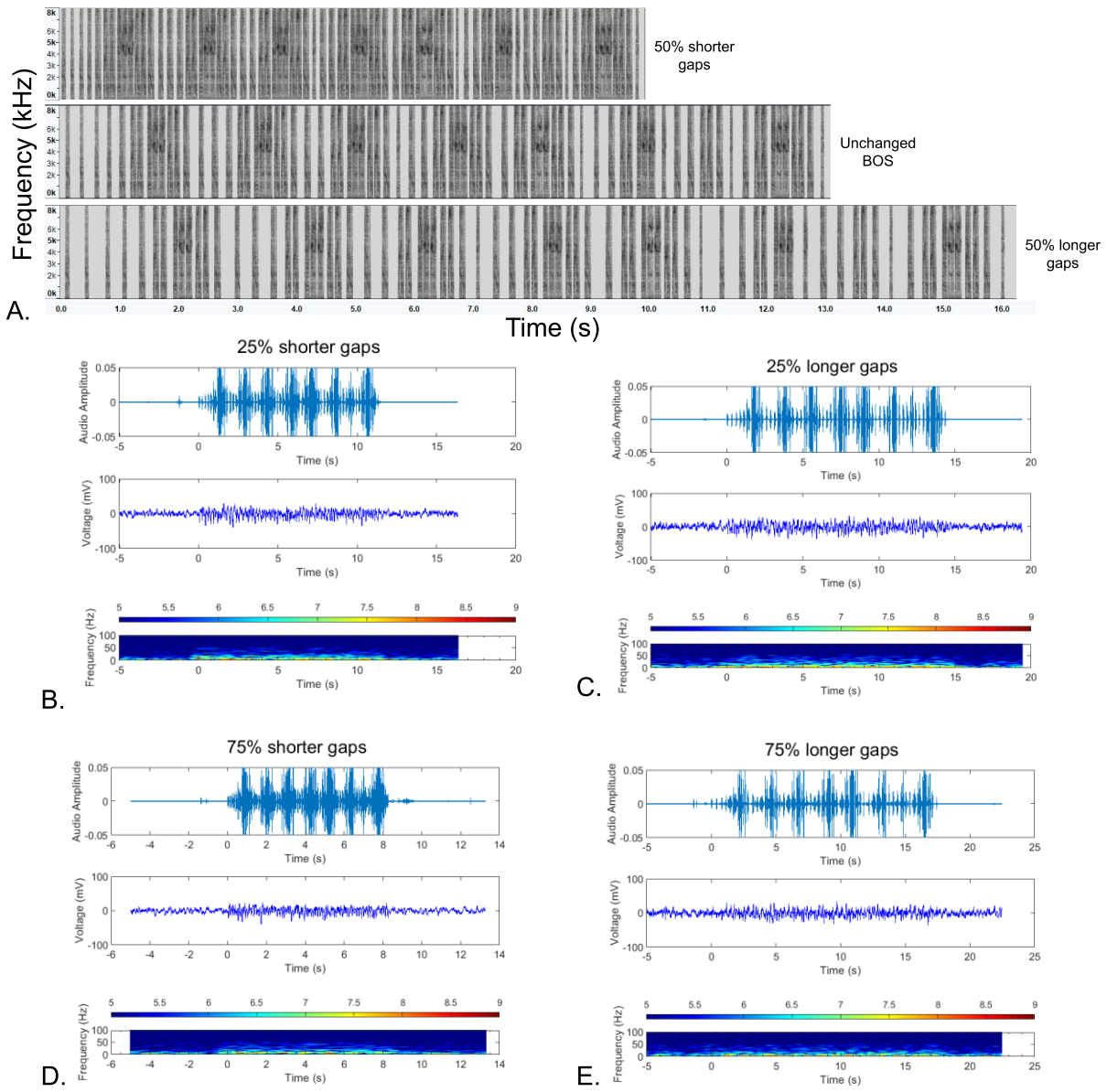
rate before playback) during syllables (green) and ISGs (blue) in white noise playback trials. Response strength was significantly different across stimulus playback condition ($p = 8.0E-17$, $df = 8$, F-statistic = 14.89, one-way ANOVA comparing syllable response strength; See Supplementary Table 4.4 for multiple comparisons). D. Response strength during syllables (green) and ISGs (blue) when controlling for the number of white noise bursts and ISG length across playback conditions. Response strength was significantly different across stimulus playback condition ($p = 3.51E-9$, $df = 8$, F-statistic = 7.92, one-way ANOVA comparing syllable response strength; See Supplementary Table 4.4 for multiple comparisons). All bar plots display mean ± SEM.



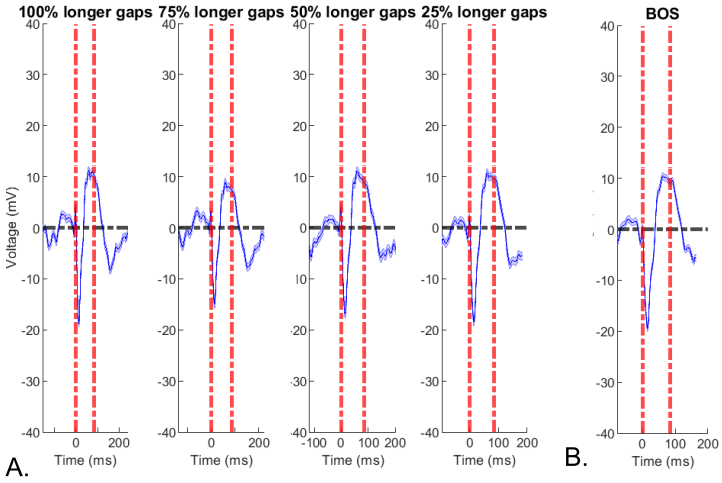
Supplementary Figure 4.1. LFP during bird's own song (BOS) and reversed BOS (rBOS) playback. A. *Upper panel:* Spectrogram of BOS with silent inter-syllable gaps (ISGs). *Lower panel:* Spectrogram of reversed BOS (rBOS) with silent ISGs. B. LEFT. *Upper panel:* Average audio oscillogram of BOS playback. *Middle panel:* Average LFP voltage trace. *Lower panel:* Spectrogram of average LFP voltage trace. RIGHT. Same as left panel, but zoomed in from 0 to 2 seconds. C. Same as B, but for rBOS playback.

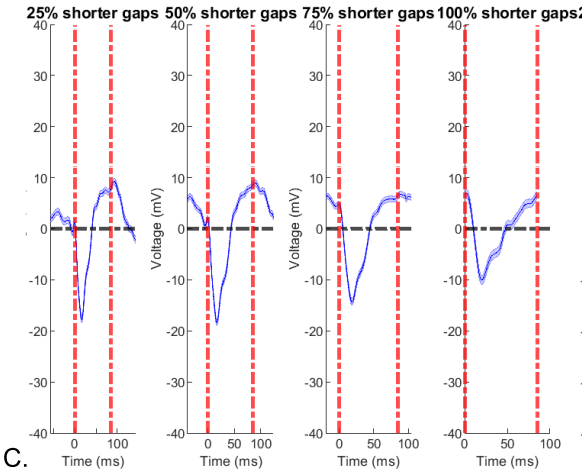


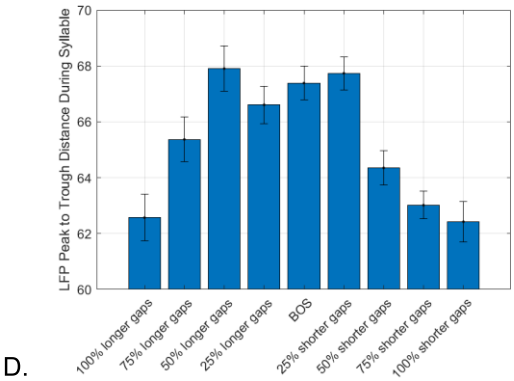
Supplementary Figure 4.2. HVC LFP activity reflects selective bird's own song (BOS) response. A. Stimulus-triggered average LFP trace aligned to onsets of BOS syllables, after accounting for the 15-ms response delay, for Site 1. The first red line indicates the start of the syllable and the second red line indicates the average duration of BOS syllables (again, after accounting for the 15-ms response delay). B. Same as A, but for reversed BOS (rBOS). C. Peak-to-trough distance in millivolts of the LFP during BOS or rBOS syllable presentation (i.e., in the sections enclosed by red lines in A-B) at Site 1. Peak-to-trough distance during BOS syllables was higher than during rBOS syllables ($p = 6.68$ E-18, $n = 1403$ BOS syllables, $n = 1380$ rBOS syllables, two-tailed t-test). D-E. Same as A-B, but for Site 2. F. Peak-to-trough distance of LFP during BOS vs. rBOS syllables at Site 2. Peak-to-trough distance during BOS syllables was higher than that of rBOS ($p = 0.0076$, $n = 3696$ BOS syllables, $n = 3696$ rBOS syllables, two-tailed t-test). G-H. Same as A-B, but for a third site. F. Peak-to-trough distance of LFP during BOS vs. rBOS syllables at a third site. Peak-to-trough distance during BOS syllables was higher than that of rBOS ($p = 5.95$ E-5, $n = 1794$ BOS syllables, $n = 1794$ rBOS syllables, two-tailed t-test). All bar plots display mean ± SEM.



Supplementary Figure 4.3. LFP during playback of tempo-manipulated bird's own song (BOS). A. Spectrogram of BOS with 50% shortened gap lengths (top panel), unchanged gap lengths (middle panel), and 50% lengthened gaps lengths (lower panel). B. *Upper panel:* Average audio oscillogram of BOS with 25% shorter gaps. *Middle panel:* Average LFP voltage trace. *Lower panel:* Spectrogram of average LFP voltage trace. C. Same as B, but for BOS with 25% longer gaps. D. Same as B, but for BOS with 75% shorter gaps. E. Same as B, but for BOS with 75% longer gaps.

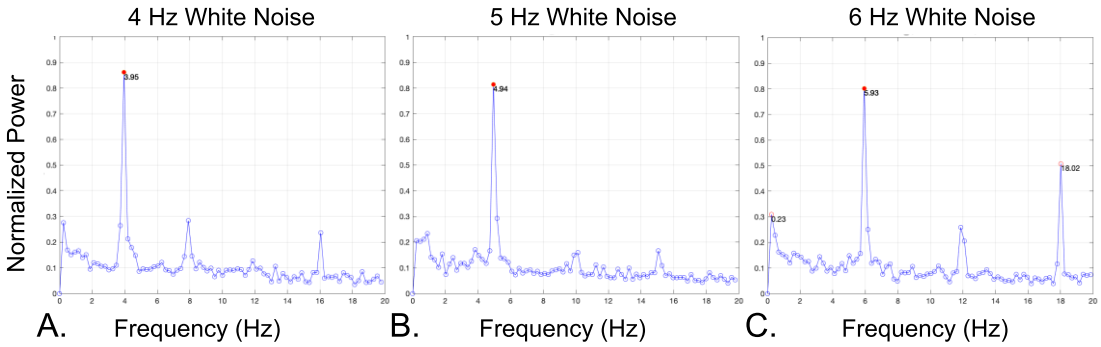




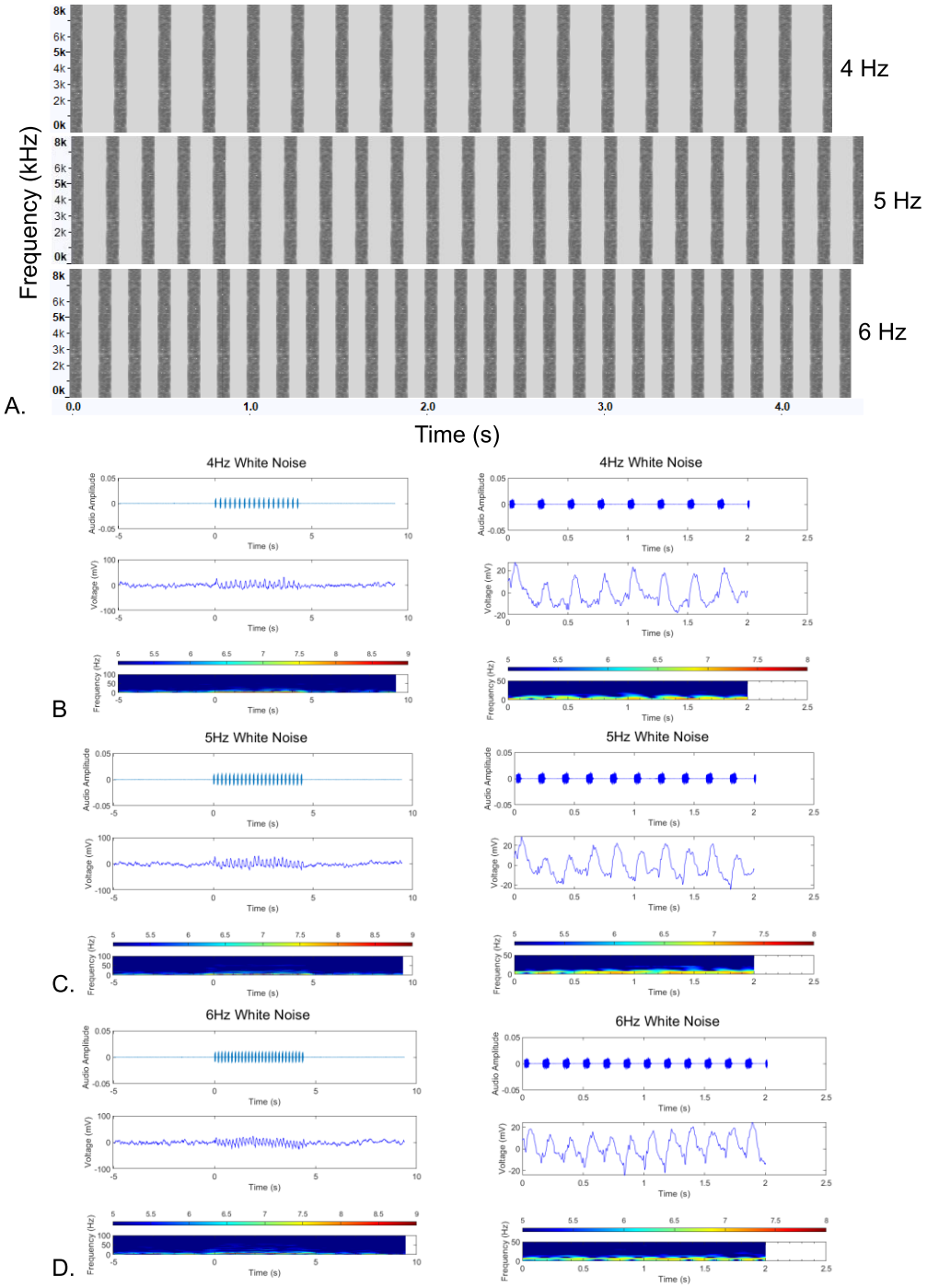


Supplementary Figure 4.4. HVC LFP activity reflects tuning to the bird's own song (BOS) tempo. A. Stimulus-triggered average (STA) LFP traces aligned to the onset of BOS syllables (after accounting for the 15-ms response delay) in stimuli with lengthened inter-syllable gaps (ISGs). The first red line indicates the start of the syllable and the second red line indicates the average duration of BOS syllables. From leftmost to rightmost panel: STA for BOS with 100%

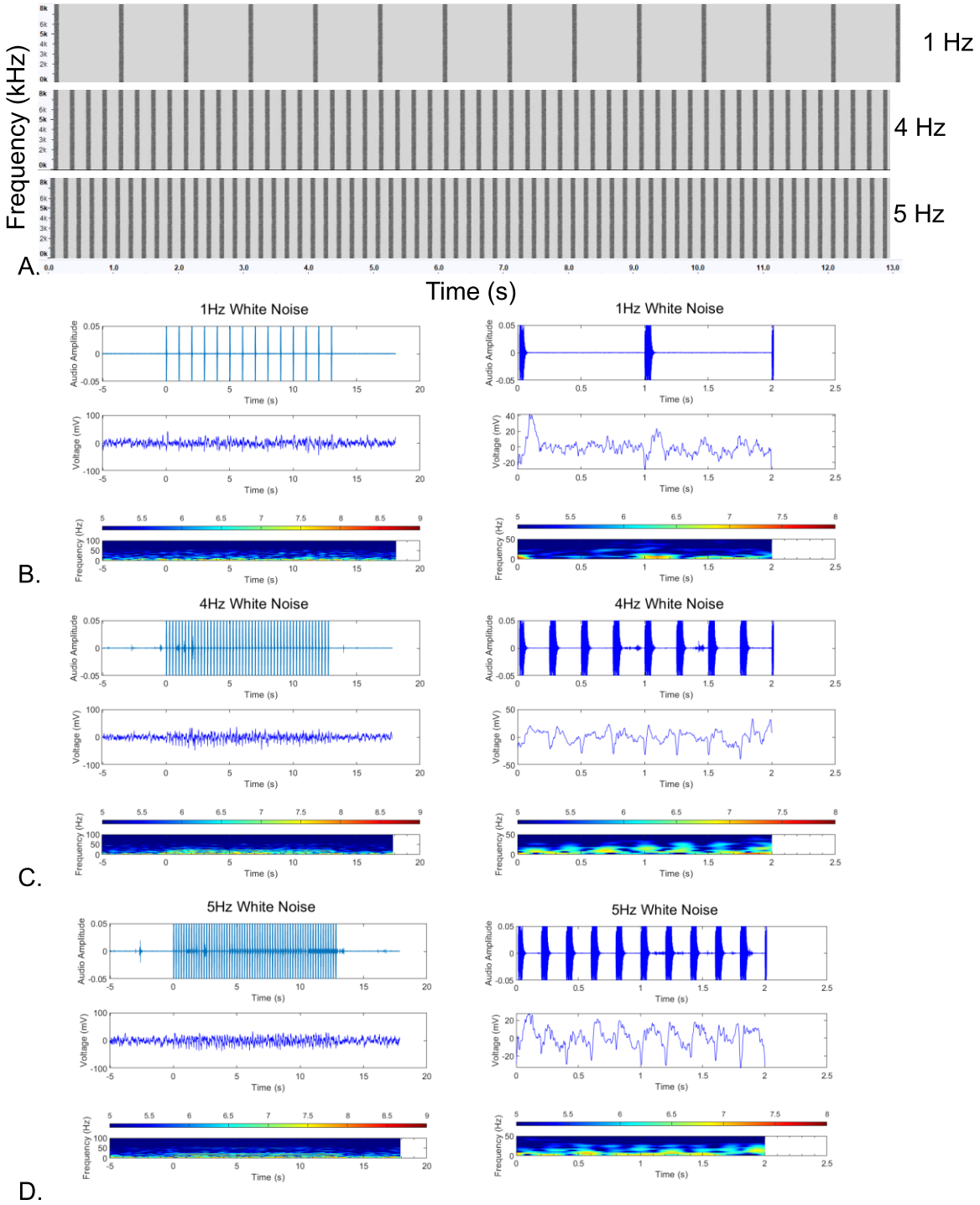
longer ISGs, STA for BOS with 75% longer ISGs, STA for BOS with 50% longer ISGs, STA for BOS with 25% longer ISGs. B. Same as A, but for unmanipulated BOS. C. Same as A, but for stimuli with shortened ISGs. From leftmost to rightmost panel: STA for BOS with 25% shorter ISGs, STA for BOS with 50% shorter ISGs, STA for BOS with 75% shorter ISGs, STA for BOS with 100% shorter ISGs. D. Peak-to-trough distance in millivolts of the LFP traces in A-C during syllable presentation (i.e., in the sections enclosed by red lines in A-C). Peak-to-trough distance was significantly different across stimuli ($p = 1.52$ E-15 $df = 8$, F-statistic = 10.97, one-way ANOVA; See Supplementary Table 4.2 for multiple comparisons). All bar plots display mean ± SEM.



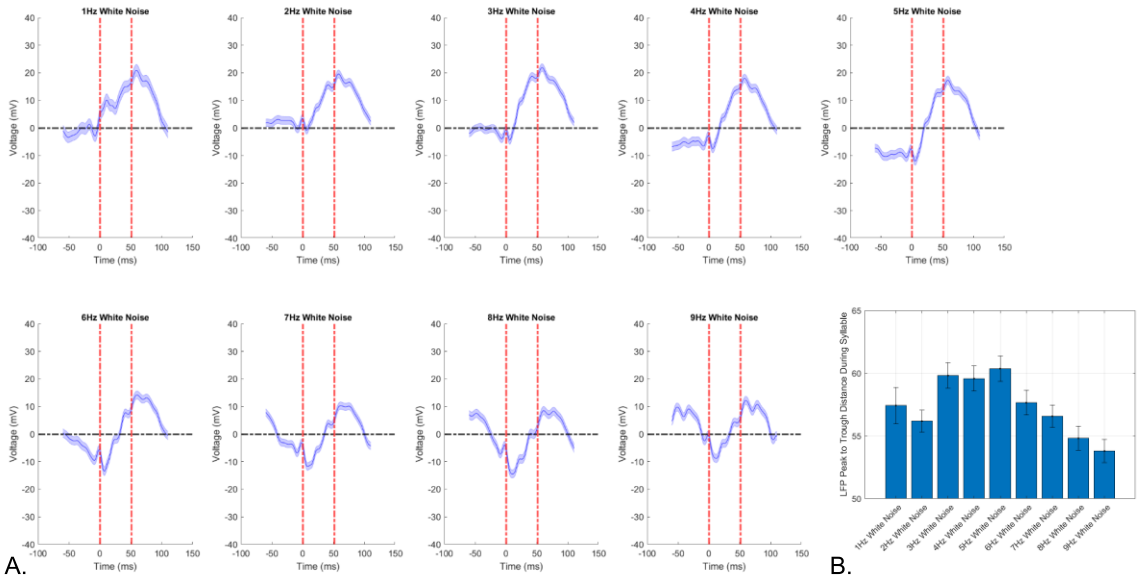
Supplementary Figure 4.5. Spikes are time-locked to white noise "syllables." Average power spectrum densities of spike times during playback of white noise bursts presented at (A) 4 Hz, (B) 5 Hz, and (C) 6 Hz with 0.08 Hz resolution.



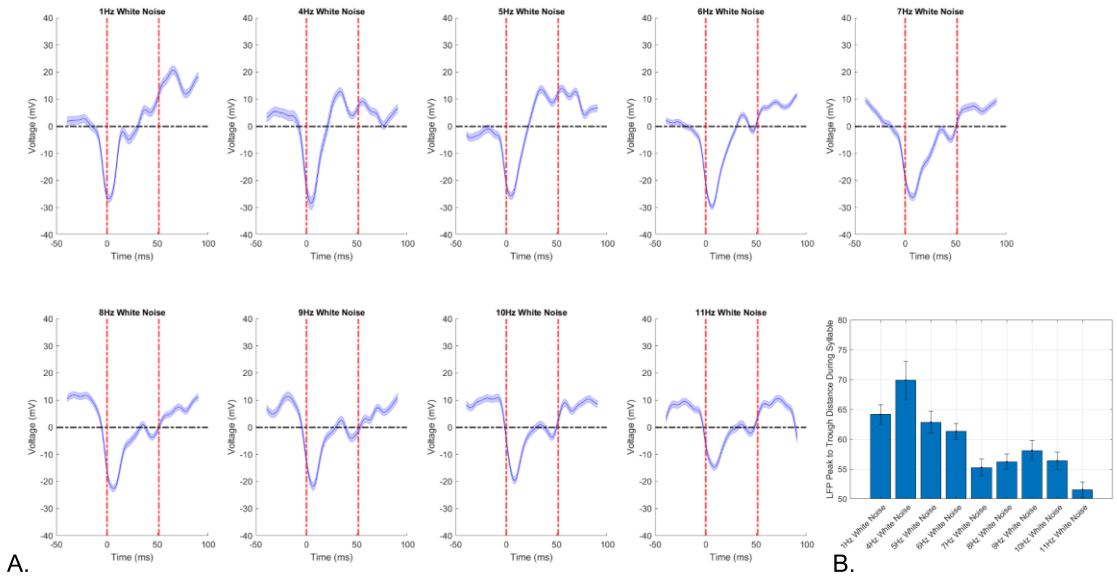
Supplementary Figure 4.6. LFP during white noise playback, Site 1. A. Spectrogram of white noise bursts presented at 4 Hz (top panel), 5 Hz (middle panel), and 6 Hz (lower panel) with silent inter-syllable gaps (ISGs). B. LEFT. *Upper panel:* Average audio oscillogram of 4-Hz white noise playback. *Middle panel:* Average LFP voltage trace. *Lower panel:* Spectrogram of average LFP voltage trace. RIGHT. Same as left panel, but zoomed in from 0 to 2 seconds. C. Same as B, but for 5-Hz white noise playback. D. Same as B, but for 6-Hz white noise playback.



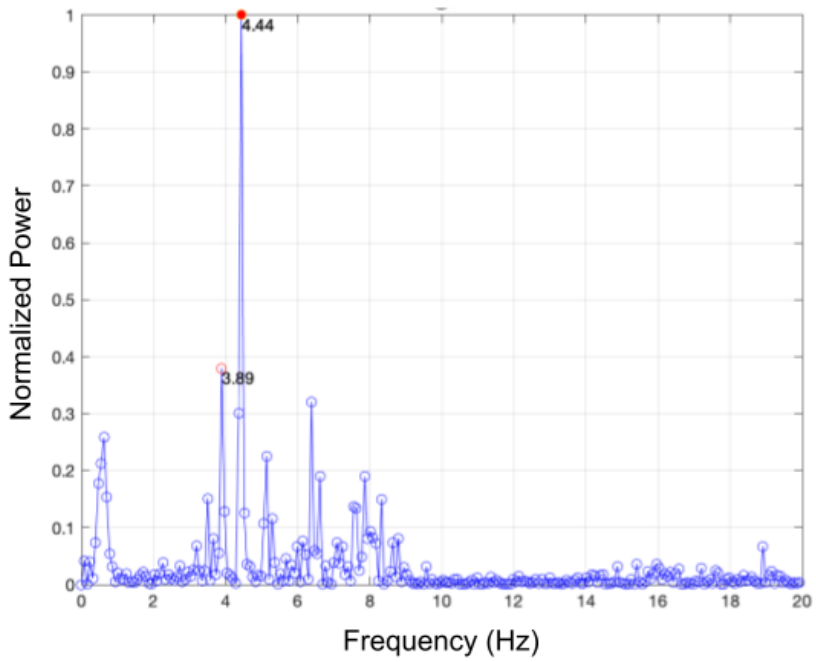
Supplementary Figure 4.7. LFP during white noise playback, Site 2. A. Spectrogram of white noise bursts presented at 1 Hz (top panel), 4 Hz (middle panel), and 5 Hz (lower panel) with silent inter-syllable gaps (ISGs). B. LEFT. *Upper panel:* Average audio oscillogram of 1-Hz white noise playback. *Middle panel:* Average LFP trace. *Lower panel:* Spectrogram of average LFP voltage trace. RIGHT. Same as left panel, but zoomed in from 0 to 2 seconds. C. Same as B, but for 4-Hz white noise playback. D. Same as B, but for 5-Hz white noise playback.



Supplementary Figure 4.8. HVC LFP activity at Site 1 mirrors HVC spiking activity and is suggestive of tempo tuning to 5 Hz. A. Stimulus-triggered average (STA) LFP voltage trace aligned to the onset of white noise syllables (after accounting for the 15-ms response delay) presented at (from left to right and top to bottom) 1 Hz, 2 Hz, 3 Hz, 4 Hz, 5 Hz, 6 Hz, 7 Hz, 8 Hz, or 9 Hz. The first red line indicates the start of the syllable and the second red line indicates the end of the white noise syllable. B. Peak-to-trough distance in millivolts of the LFP during syllables (i.e., in the sections enclosed by red lines in A) for each white noise stimulus. Peak-to-trough distance was significantly different across stimuli ($p = 7.45$ E-7 $df = 8$, F-statistic = 5.44, one-way ANOVA; See Supplementary Table 4.5 for multiple comparisons). All bar plots display mean ± SEM.



Supplementary Figure 4.9. HVC LFP activity at Site 2 mirrors HVC spiking activity and is suggestive of tempo tuning to 4 Hz. A. Stimulus-triggered average (STA) LFP voltage trace aligned to the onset of white noise syllables (after accounting for the 15-ms response delay) presented at (from left to right and top to bottom) 1 Hz, 4 Hz, 5 Hz, 6 Hz, 7 Hz, 8 Hz, 9 Hz, 10 Hz, or 11 Hz. The first red line indicates the start of the syllable and the second red line indicates the end of the white noise syllable. B. Peak-to-trough distance of the LFP voltage traces in A during white noise syllable presentation (i.e., in the sections enclosed by red lines in A). Peak-to-trough distance was significantly different across stimuli ($p = 1.48$ E-13 $df = 8$, F-statistic = 9.9, one-way ANOVA; See Supplementary Table 4.6 for multiple comparisons). All bar plots display mean ± SEM.



Supplementary Figure 4.10. Power spectrum density of the song's syllable-gap temporal structure with 0.08 Hz resolution. Briefly, a binary vector was created to describe the temporal patterning of song syllables. For each data point in the song, a 1 was added to the vector if the point was part of a song syllable and a 0 was added if the point was part of an inter-syllable gap (adapted from Kozhevnikov and Fee, 2007). The power spectrum density of this vector was then found (see Methods for further details).

Supplementary Table 4.1. Statistics from Figure 4.2, using multiple comparisons with Bonferroni correction. The first four columns display the identity of each stimulus being compared and the number of playbacks of each stimulus. The fifth and sixth columns display the p-value from the multiple-comparisons test from data displayed in Figure 4.2 C and the conclusion from the test if the p-value is significant (e.g., "BOS > 75% shorter gaps" indicates that the response strength is higher for BOS syllables when tempo is unchanged vs. when inter-syllable gap lengths are reduced by 75%). The seventh and eighth columns display the p-value from the multiple-comparisons test from data displayed in Figure 4.2 D and the conclusion from the test if the p-value is significant.

Tempo, Stim. 1	Number Trials, Stim. 1	Tempo, Stim. 2	Number Trials, Stim. 2	Response Strength During Full Playback		Response Strength Controlling for Stimulus Length and Gap Length	
				p-value	Conclusions (if significant difference)	p-value	Conclusions (if significant difference)
BOS	26	100% shorter gaps	15	**1.3E-06**	**BOS > 100% shorter gaps**	**7.1E-07**	**BOS > 100% shorter gaps**
BOS	26	25% shorter gaps	30	1.0E+00		1.0E+00	
BOS	26	50% shorter gaps	26	1.0E+00		4.2E-01	
BOS	26	75% shorter gaps	26	**7.6E-04**	**BOS > 75% shorter gaps**	**3.8E-04**	**BOS > 75% shorter gaps**
BOS	26	100% longer gaps	19	1.0E+00		1.0E+00	
BOS	26	25% longer gaps	16	1.0E+00		1.0E+00	
BOS	26	50% longer gaps	16	1.0E+00		1.0E+00	
BOS	26	75% longer gaps	18	1.0E+00		7.3E-01	
100% shorter gaps	15	25% shorter gaps	30	**1.1E-05**	**100% shorter gaps < 25% shorter gaps**	**1.3E-05**	**100% shorter gaps < 25% shorter gaps**
100% shorter gaps	15	50% shorter gaps	26	**4.0E-03**	**100% shorter gaps < 50% shorter gaps**	**1.0E-02**	**100% shorter gaps < 50% shorter gaps**
100% shorter gaps	15	75% shorter gaps	26	1.0E+00		1.0E+00	
100% shorter gaps	15	100% longer gaps	19	**4.9E-04**	**100% shorter gaps < 100% longer gaps**	**2.6E-03**	**100% shorter gaps < 100% longer gaps**
100% shorter gaps	15	25% longer gaps	16	**7.1E-06**	**100% shorter gaps < 25% shorter gaps**	**3.4E-05**	**100% shorter gaps < 25% shorter gaps**
100% shorter gaps	15	50% longer gaps	16	**1.1E-05**	**100% shorter gaps < 50% longer gaps**	**9.0E-05**	**100% shorter gaps < 50% longer gaps**
100% shorter gaps	15	75% longer gaps	18	**7.5E-03**	**100% shorter gaps < 75% longer gaps**	**3.0E-02**	**100% shorter gaps < 75% longer gaps**
25% shorter gaps	30	50% shorter gaps	26	1.0E+00		1.0E+00	
25% shorter gaps	30	75% shorter gaps	26	**6.3E-03**	**25% shorter gaps > 75% shorter gaps**	**6.5E-03**	**25% shorter gaps > 75% shorter gaps**
25% shorter gaps	30	100% longer gaps	19	1.0E+00		1.0E+00	
25% shorter gaps	30	25% longer gaps	16	1.0E+00		1.0E+00	
25% shorter gaps	30	50% longer gaps	16	1.0E+00		1.0E+00	
25% shorter gaps	30	75% longer gaps	18	1.0E+00		1.0E+00	
50% shorter gaps	26	75% shorter gaps	26	9.2E-01		1.0E+00	
50% shorter gaps	26	100% longer gaps	19	1.0E+00		1.0E+00	
50% shorter gaps	26	25% longer gaps	16	1.0E+00		1.0E+00	
50% shorter gaps	26	50% longer gaps	16	1.0E+00		1.0E+00	
50% shorter gaps	26	75% longer gaps	18	1.0E+00		1.0E+00	
75% shorter gaps	26	100% longer gaps	19	1.3E-01		4.8E-01	
75% shorter gaps	26	25% longer gaps	16	**2.7E-03**	**75% shorter gaps < 25% longer gaps**	**1.0E-02**	**75% shorter gaps < 25% longer gaps**
75% shorter gaps	26	50% longer gaps	16	**3.9E-03**	**75% shorter gaps < 50% longer gaps**	**2.4E-02**	**75% shorter gaps < 50% longer gaps**
75% shorter gaps	26	75% longer gaps	18	1.0E+00		1.0E+00	
100% longer gaps	19	25% longer gaps	16	1.0E+00		1.0E+00	
100% longer gaps	19	50% longer gaps	16	1.0E+00		1.0E+00	
100% longer gaps	19	75% longer gaps	18	1.0E+00		1.0E+00	
25% longer gaps	16	50% longer gaps	16	1.0E+00		1.0E+00	
25% longer gaps	16	75% longer gaps	18	1.0E+00		1.0E+00	
50% longer gaps	16	75% longer gaps	18	1.0E+00		1.0E+00	

Supplementary Table 4.2. Statistics from Supplementary Figure 4.4, using multiple comparisons with Bonferroni correction. The first four columns display the identity of each stimulus being compared and the number of syllable playbacks for each stimulus. The fifth and sixth columns display the p-value from the multiple-comparisons test from data displayed in Supplementary Figure 4.4 D and the conclusion from the test if the p-value is significant (e.g., "BOS > 50% shorter gaps" indicates that the LFP peak-to-trough distance is higher for BOS syllables when tempo is unchanged vs. when inter-syllable gap lengths are reduced by 50%).

Tempo, Stim. 1	Number Stimulus Presentations, Stim. 1	Tempo, Stim. 2	Number Stimulus Presentations, Stim. 2	Peak to Trough Distance	
				p-value	Conclusions (if significant difference)
BOS	1794	100% shorter gaps	1014	**4.3E-05**	**BOS > 100% shorter gaps**
BOS	1794	25% shorter gaps	2106	1.0E+00	
BOS	1794	50% shorter gaps	1872	**1.5E-02**	**BOS > 50% shorter gaps**
BOS	1794	75% shorter gaps	1872	**1.3E-05**	**BOS > 75% shorter gaps**
BOS	1794	100% longer gaps	1326	**1.2E-05**	**BOS > 100% longer gaps**
BOS	1794	25% longer gaps	1092	1.0E+00	
BOS	1794	50% longer gaps	1092	1.0E+00	
BOS	1794	75% longer gaps	1326	1.0E+00	
100% shorter gaps	1014	25% shorter gaps	2106	**3.3E-06**	**100% shorter gaps < 25% shorter gaps**
100% shorter gaps	1014	50% shorter gaps	1872	1.0E+00	
100% shorter gaps	1014	75% shorter gaps	1872	1.0E+00	
100% shorter gaps	1014	100% longer gaps	1326	1.0E+00	
100% shorter gaps	1014	25% longer gaps	1092	**8.0E-03**	**100% shorter gaps < 25% longer gaps**
100% shorter gaps	1014	50% longer gaps	1092	**4.8E-05**	**100% shorter gaps < 50% longer gaps**
100% shorter gaps	1014	75% longer gaps	1326	2.4E-01	
25% shorter gaps	2106	50% shorter gaps	1872	**1.5E-03**	**25% shorter gaps > 50% shorter gaps**
25% shorter gaps	2106	75% shorter gaps	1872	**4.0E-07**	**25% shorter gaps > 75% shorter gaps**
25% shorter gaps	2106	100% longer gaps	1326	**5.3E-07**	**25% shorter gaps > 100% longer gaps**
25% shorter gaps	2106	25% longer gaps	1092	1.0E+00	
25% shorter gaps	2106	50% longer gaps	1092	1.0E+00	
25% shorter gaps	2106	75% longer gaps	1326	3.4E-01	
50% shorter gaps	1872	75% shorter gaps	1872	1.0E+00	
50% shorter gaps	1872	100% longer gaps	1326	1.0E+00	
50% shorter gaps	1872	25% longer gaps	1092	8.1E-01	
50% shorter gaps	1872	50% longer gaps	1092	**1.2E-02**	**50% shorter gaps < 50% longer gaps**
50% shorter gaps	1872	75% longer gaps	1326	1.0E+00	
75% shorter gaps	1872	100% longer gaps	1326	1.0E+00	
75% shorter gaps	1872	25% longer gaps	1092	**1.0E-02**	**75% shorter gaps < 25% longer gaps**
75% shorter gaps	1872	50% longer gaps	1092	**2.9E-05**	**75% shorter gaps < 50% longer gaps**
75% shorter gaps	1872	75% longer gaps	1326	4.2E-01	
100% longer gaps	1326	25% longer gaps	1092	**5.2E-03**	**100% longer gaps < 25% longer gaps**
100% longer gaps	1326	50% longer gaps	1092	**1.9E-05**	**100% longer gaps < 50% longer gaps**
100% longer gaps	1326	75% longer gaps	1326	2.0E-01	
25% longer gaps	1092	50% longer gaps	1092	1.0E+00	
25% longer gaps	1092	75% longer gaps	1326	1.0E+00	
50% longer gaps	1092	75% longer gaps	1326	6.1E-01	

Supplementary Table 4.3. Statistics from Figure 4.3, using multiple comparisons with Bonferroni correction. The first four columns display the tempo of each stimulus being compared and the number of playbacks of each stimulus. The fifth and sixth columns display the p-value from the multiple-comparisons test from data displayed in Figure 4.3 C and the conclusion from the test if the p-value is significant. The seventh and eighth columns display the p-value from the multiple-comparisons test from data displayed in Figure 4.3 D and the conclusion from the test if the p-value is significant.

Tempo, Stim. 1	Number Trials, Stim. 1	Tempo, Stim. 2	Number Trials, Stim. 2	Response Strength During Full Playback		Response Strength Controlling For # WN Bursts and Gap Length	
				p-value	Conclusions (if significant difference)	p-value	Conclusions (if significant difference)
1 Hz	67	2 Hz	82	1.0E+00		1.0E+00	
1 Hz	67	3 Hz	81	1.0E+00		1.0E+00	
1 Hz	67	4 Hz	71	**6.6E-05**	**1 Hz < 4 Hz**	1.1E-01	
1 Hz	67	5 Hz	79	**5.9E-11**	**1 Hz < 5 Hz**	**1.6E-03**	**1 Hz < 5 Hz**
1 Hz	67	6 Hz	67	8.2E-02		1.0E+00	
1 Hz	67	7 Hz	80	4.0E-01		1.0E+00	
1 Hz	67	8 Hz	67	1.0E+00		1.0E+00	
1 Hz	67	9 Hz	65	1.0E+00		1.0E+00	
2 Hz	82	3 Hz	81	1.0E+00		1.0E+00	
2 Hz	82	4 Hz	71	**8.3E-05**	**2 Hz < 4 Hz**	**4.6E-03**	**2 Hz < 4 Hz**
2 Hz	82	5 Hz	79	**3.3E-11**	**2 Hz < 5 Hz**	**1.8E-05**	**2 Hz < 5 Hz**
2 Hz	82	6 Hz	67	1.3E-01		1.0E+00	
2 Hz	82	7 Hz	80	6.3E-01		1.0E+00	
2 Hz	82	8 Hz	67	1.0E+00		1.0E+00	
2 Hz	82	9 Hz	65	1.0E+00		1.0E+00	
3 Hz	81	4 Hz	71	1.3E-01		1.0E+00	
3 Hz	81	5 Hz	79	**3.6E-06**	**3 Hz < 5 Hz**	**2.7E-02**	**3 Hz < 5 Hz**
3 Hz	81	6 Hz	67	1.0E+00		1.0E+00	
3 Hz	81	7 Hz	80	1.0E+00		1.0E+00	
3 Hz	81	8 Hz	67	2.9E-01		1.0E+00	
3 Hz	81	9 Hz	65	**9.7E-03**	**3 Hz > 9 Hz**	1.0E+00	
4 Hz	71	5 Hz	79	7.8E-01		1.0E+00	
4 Hz	71	6 Hz	67	1.0E+00		2.0E-01	
4 Hz	71	7 Hz	80	5.4E-01		2.4E-01	
4 Hz	71	8 Hz	67	**4.1E-06**	**4 Hz > 8 Hz**	**1.5E-02**	**4 Hz > 8 Hz**
4 Hz	71	9 Hz	65	**1.7E-08**	**4 Hz > 9 Hz**	**6.1E-03**	**4 Hz > 9 Hz**
5 Hz	79	6 Hz	67	**2.4E-03**	**5 Hz > 6 Hz**	**3.5E-03**	**5 Hz > 6 Hz**
5 Hz	79	7 Hz	80	**4.7E-05**	**5 Hz > 7 Hz**	**3.6E-03**	**5 Hz > 7 Hz**
5 Hz	79	8 Hz	67	**1.1E-12**	**5 Hz > 8 Hz**	**1.1E-04**	**5 Hz > 8 Hz**
5 Hz	79	9 Hz	65	**7.8E-16**	**5 Hz > 9 Hz**	**3.7E-05**	**5 Hz > 9 Hz**
6 Hz	67	7 Hz	80	1.0E+00		1.0E+00	
6 Hz	67	8 Hz	67	**1.2E-02**	**6 Hz > 8 Hz**	1.0E+00	
6 Hz	67	9 Hz	65	**2.1E-04**	**6 Hz > 9 Hz**	1.0E+00	
7 Hz	80	8 Hz	65	6.9E-02		1.0E+00	
7 Hz	80	9 Hz	65	**1.6E-03**	**7 Hz > 9 Hz**	1.0E+00	
8 Hz	67	9 Hz	65	1.0E+00		1.0E+00	

Supplementary Table 4.4. Statistics from Figure 4.4, using multiple comparisons with Bonferroni correction. The first four columns display the tempo of each stimulus being compared and the number of trials for each stimulus. The fifth and sixth columns display the p-value from the multiple-comparisons test from data displayed in Figure 4.4 C and the conclusion from the test if the p-value is significant. The seventh and eighth columns display the p-value from the multiple-comparisons test from data displayed in Figure 4.4 D and the conclusion from the test if the p-value is significant.

Tempo, Stim. 1	Number Trials, Stim. 1	Tempo, Stim. 2	Number Trials, Stim. 2	Response Strength During Full Playback		Response Strength Controlling For # WN Bursts and Gap Length	
				p-value	Conclusions (if significant difference)	p-value	Conclusions (if significant difference)
1 Hz	12	4 Hz	23	**9.9E-07**	**4 Hz > 1 Hz**	**2.5E-03**	**4 Hz > 1 Hz**
1 Hz	12	5 Hz	20	**6.1E-04**	**5 Hz >1 Hz**	**5.9E-03**	**5 Hz >1 Hz**
1 Hz	12	6 Hz	34	1.0E+00		1.0E+00	
1 Hz	12	7 Hz	15	1.0E+00		1.0E+00	
1 Hz	12	8 Hz	25	1.0E+00		1.0E+00	
1 Hz	12	9 Hz	22	1.0E+00		1.0E+00	
4 Hz	23	5 Hz	20	1.0E+00		1.0E+00	
4 Hz	23	6 Hz	34	**2.9E-05**	**4 Hz > 6 Hz**	**3.4E-03**	**4 Hz > 6 Hz**
4 Hz	23	7 Hz	15	**2.3E-07**	**4 Hz > 7 Hz**	**1.5E-02**	**4 Hz > 7 Hz**
4 Hz	23	8 Hz	25	**9.9E-09**	**4 Hz > 8 Hz**	**8.9E-03**	**4 Hz > 8 Hz**
4 Hz	23	9 Hz	22	**3.4E-09**	**4 Hz > 9 Hz**	**1.1E-03**	**4 Hz > 9 Hz**
5 Hz	20	6 Hz	34	**4.4E-02**	**5 Hz > 6 HZ**	**1.2E-02**	**5 Hz > 6 HZ**
5 Hz	20	7 Hz	15	**3.0E-04**	**5 Hz > 7 Hz**	**3.3E-02**	**5 Hz > 7 Hz**
5 Hz	20	8 Hz	25	**7.1E-05**	**5 Hz > 8 Hz**	**2.5E-02**	**5 Hz > 8 Hz**
5 Hz	20	9 Hz	22	**2.3E-05**	**5 Hz > 9 Hz**	**3.6E-03**	**5 Hz > 9 Hz**
6 Hz	34	7 Hz	15	1.0E+00		1.0E+00	
6 Hz	34	8 Hz	25	1.0E+00		1.0E+00	
6 Hz	34	9 Hz	22	5.6E-01		1.0E+00	
7 Hz	15	8 Hz	25	1.0E+00		1.0E+00	
7 Hz	15	9 Hz	22	1.0E+00		1.0E+00	
8 Hz	25	9 Hz	22	1.0E+00		1.0E+00	
10 Hz	24	11 Hz	23	1.0E+00		1.0E+00	
10 Hz	24	1 Hz	12	1.0E+00		1.0E+00	
10 Hz	24	4 Hz	23	**1.9E-09**	**10 Hz < 4 Hz**	**2.5E-04**	**10 Hz < 4 Hz**
10 Hz	24	5 Hz	20	**1.7E-05**	**10 Hz < 5 Hz**	**9.5E-04**	**10 Hz < 5 Hz**
10 Hz	24	6 Hz	34	5.4E-01		1.0E+00	
10 Hz	24	7 Hz	15	1.0E+00		1.0E+00	
10 Hz	24	8 Hz	25	1.0E+00		1.0E+00	
10 Hz	24	9 Hz	22	1.0E+00		1.0E+00	
11 Hz	23	1 Hz	12	1.0E+00		1.0E+00	
11 Hz	23	4 Hz	23	**1.0E-12**	**11 Hz < 4 Hz**	**2.6E-07**	**11 Hz < 4 Hz**
11 Hz	23	5 Hz	20	**3.1E-08**	**11 Hz < 5 Hz**	**1.6E-06**	**11 Hz < 5 Hz**
11 Hz	23	6 Hz	34	**4.9E-03**	**11 Hz < 6 Hz**	3.3E-01	
11 Hz	23	7 Hz	15	1.0E+00		1.0E+00	
11 Hz	23	8 Hz	25	1.0E+00		5.5E-01	
11 Hz	23	9 Hz	22	1.0E+00		1.0E+00	

Supplementary Figure 4.5. Statistics from Supplementary Figure 4.8, using multiple comparisons with Bonferroni correction. The first four columns display the tempo of each stimulus being compared and the number of syllable playbacks for each stimulus. The fifth and sixth columns display the p-value from the multiple-comparisons test from data displayed in Supplementary Figure 4.8 B and the conclusion from the test if the p-value is significant.

Tempo, Stim. 1	Number Stimulus Presentations, Stim. 1	Tempo, Stim. 2	Number Stimulus Presentations, Stim. 2	Peak to Trough Distance	
				p-value	Conclusions (if significant difference)
1 Hz	198	2 Hz	539	1.0E+00	
1 Hz	198	3 Hz	790	1.0E+00	
1 Hz	198	4 Hz	710	1.0E+00	
1 Hz	198	5 Hz	790	1.0E+00	
1 Hz	198	6 Hz	640	1.0E+00	
1 Hz	198	7 Hz	760	1.0E+00	
1 Hz	198	8 Hz	650	1.0E+00	
1 Hz	198	9 Hz	630	1.0E+00	
2 Hz	539	3 Hz	790	2.4E-01	
2 Hz	539	4 Hz	710	4.9E-01	
2 Hz	539	5 Hz	790	6.6E-02	
2 Hz	539	6 Hz	640	1.0E+00	
2 Hz	539	7 Hz	760	1.0E+00	
2 Hz	539	8 Hz	650	1.0E+00	
2 Hz	539	9 Hz	630	1.0E+00	
3 Hz	790	4 Hz	710	1.0E+00	
3 Hz	790	5 Hz	790	1.0E+00	
3 Hz	790	6 Hz	640	1.0E+00	
3 Hz	790	7 Hz	760	4.8E-01	
3 Hz	790	8 Hz	650	**9.2E-03**	**3 Hz > 8 Hz**
3 Hz	790	9 Hz	630	**4.7E-04**	**3 Hz > 9 Hz**
4 Hz	710	5 Hz	790	1.0E+00	
4 Hz	710	6 Hz	640	1.0E+00	
4 Hz	710	7 Hz	760	9.2E-01	
4 Hz	710	8 Hz	650	**2.5E-02**	**4 Hz > 8 Hz**
4 Hz	710	9 Hz	630	**1.6E-03**	**4 Hz > 9 Hz**
5 Hz	790	6 Hz	640	1.0E+00	
5 Hz	790	7 Hz	760	1.4E-01	
5 Hz	790	8 Hz	650	**1.8E-03**	**5 Hz > 8 Hz**
5 Hz	790	9 Hz	630	**7.1E-05**	**5 Hz > 9 Hz**
6 Hz	640	7 Hz	760	1.0E+00	
6 Hz	640	8 Hz	650	1.0E+00	
6 Hz	640	9 Hz	630	2.8E-01	
7 Hz	760	8 Hz	650	1.0E+00	
7 Hz	760	9 Hz	630	1.0E+00	
8 Hz	650	9 Hz	630	1.0E+00	

Supplementary Figure 4.6. Statistics from Supplementary Figure 4.9, using multiple comparisons with Bonferroni correction. The first four columns display the tempo of each stimulus being compared and the number of syllable playbacks for each stimulus. The fifth and sixth columns display the p-value from the multiple-comparisons test from data displayed in Supplementary Figure 4.9 B and the conclusion from the test if the p-value is significant.

Tempo, Stim. 1	Number Stimulus Presentations, Stim. 1	Tempo, Stim. 2	Number Stimulus Presentations, Stim. 2	Peak to Trough Distance	
				p-value	Conclusions (if significant difference)
1 Hz	810	4 Hz	420	1.0E+00	
1 Hz	810	5 Hz	570	1.0E+00	
1 Hz	810	6 Hz	450	1.0E+00	
1 Hz	420	7 Hz	570	**3.3E-02**	**1 Hz > 7 Hz**
1 Hz	420	8 Hz	450	6.8E-02	
1 Hz	570	9 Hz	450	1.0E+00	
4 Hz	540	5 Hz	480	2.9E-01	
4 Hz	540	6 Hz	360	**5.6E-03**	**4 Hz > 6 Hz**
4 Hz	540	7 Hz	570	**3.2E-07**	**4 Hz > 7 Hz**
4 Hz	540	8 Hz	390	**5.3E-07**	**4 Hz > 8 Hz**
4 Hz	540	9 Hz	810	**2.4E-04**	**4 Hz > 9 Hz**
5 Hz	540	6 Hz	420	1.0E+00	
5 Hz	540	7 Hz	570	2.0E-01	
5 Hz	540	8 Hz	450	3.9E-01	
5 Hz	480	9 Hz	360	1.0E+00	
6 Hz	480	7 Hz	570	3.5E-01	
6 Hz	480	8 Hz	390	7.3E-01	
6 Hz	480	9 Hz	810	1.0E+00	
7 Hz	480	8 Hz	420	1.0E+00	
7 Hz	480	9 Hz	570	1.0E+00	
8 Hz	480	9 Hz	450	1.0E+00	
10 Hz	360	11 Hz	570	1.0E+00	
10 Hz	360	1 Hz	390	1.2E-01	
10 Hz	360	4 Hz	810	**2.6E-06**	**10 Hz < 4 Hz**
10 Hz	360	5 Hz	420	6.0E-01	
10 Hz	360	6 Hz	570	1.0E+00	
10 Hz	360	7 Hz	450	1.0E+00	
10 Hz	570	8 Hz	390	1.0E+00	
10 Hz	570	9 Hz	810	1.0E+00	
11 Hz	570	1 Hz	420	**5.6E-05**	**11 Hz < 1 Hz**
11 Hz	570	4 Hz	570	**5.6E-12**	**11 Hz < 4 Hz**
11 Hz	570	5 Hz	450	**8.5E-04**	**11 Hz < 5 Hz**
11 Hz	390	6 Hz	810	**6.4E-04**	**11 Hz < 6 Hz**
11 Hz	390	7 Hz	420	1.0E+00	
11 Hz	390	8 Hz	570	1.0E+00	
11 Hz	390	9 Hz	450	4.5E-01	

CHAPTER 5: Discussion

These findings suggest that a male's song tempo communicates valuable information about other behaviors to the listener, and that song tempo may be subserved by individual-specific HVC dynamics. The results described in Chapter 2 suggest that song tempo is related to how quickly and often a bird responds to a call. Synchronization of call timing is important for pair bonding and reproductive success in zebra finches (Gill et al., 2015), and since - based on the findings described here - females also show consistent inter-individual differences in call timing, the female may be motivated to choose a mate with compatible calling behavior. These results suggest that she may make this decision by listening to either song (i.e., by assessing song tempo) or to call exchanges. Future work may examine whether song tempo may reflect other behavioral traits, such as neophobia, which guide assortative mate choice (Schuett et al., 2011).

Results from Chapter 2 also demonstrate that the song tempo of the female's sire does not predict her calling behavior. Any correlation between a call metric and the sire's song tempo in females was expected to be weaker than the relationship between that call metric and the bird's song tempo in males. This is because the song tempo of the male reflects all the genetic and experience-dependent effects on song and related neural circuitry, while the sire's song tempo does not reflect any maternal effects or many experience-dependent factors that may impact calling behavior. However, if genetics or nest-wide experiences (e.g., exposure to the sire's song) did dictate calling behavior in females, within-nest clustering of call metrics would be expected. This was not observed in the data described in Chapter 2, and even when many spontaneous calls derived from a large sample of birds from a variety of nests were analyzed in Chapter 3, there was no statistically significant within-clustering of call features. However, call features in females were not random; results from Chapter 2 indicate that call response latency, call

response probability, and the number of call responses per trial all have meaningful relationships with one another. For instance, females that responded with low latency tended to respond in many trials and with many calls per trial. Given these findings, calls produced by naturally-reared females may provide highly individual-specific information to the listener, such as the female's motivational state or her preference for a given male. Intriguingly, a small cohort of non-tutored females showed strong within-nest clustering of calling behavior and a strong correlation between the sire's song tempo and call response latency. These data raise the possibility that call features are highly heritable in females in the absence of rich social experience, but that unique experiences may override this genetic predisposition in naturally-reared birds. Future work should expand the sample size of non-tutored females to determine whether this effect holds.

Across many species, auditory systems are tuned to encode biologically-relevant stimuli (Bowling and Purves, 2015). For instance, humans distinguish fine differences in speech sounds even in infancy (Streeter, 1976). In birds, the genetic predisposition to learn not just species-typical song (Marler, 1970; Marler and Peters, 1977), but family-typical song (Mets and Brainard, 2019), is likely supported by genetically-determined features of neural circuits that support the learning of some songs over others. Intriguingly, results in Chapter 4 from a single male Bengalese finch show that HVC is preferentially activated by stimuli that have 4 to 5 elements per second. This tempo is similar to the tempo of introductory notes produced by the bird. During learning, HVC is selective for the tutor song over any other natural stimulus including the bird's own song (Nick and Konishi, 2004). A prominent model of learning posits that organisms are born with an innate template that guides their attention toward relevant stimuli (Marler, 1997; Nelson and Marler, 1993). Notably, the degree of attention to the tutor stimulus is

a central predictor of song learning (Chen et al., 2016). Introductory gestures alert conspecifics to the onset of an important stimulus (Richards, 1981; Ord and Stamps, 2008). In white-crowned sparrows, the introductory whistle acts as a learning cue to guide learning of conspecific song over heterospecific song (Soha and Marler, 2000). If HVC is tuned to the tempo of introductory notes in Bengalese finches, this may direct the tutee's attention toward the father's song. In other words, HVC tempo tuning may act as an innate template that guides attention toward the father's song over other songs, thereby maximizing learning of the father's song. Expanding the sample size in the HVC tempo tuning experiment is therefore critical in elucidating the individual-specific neural dynamics that give rise to genetic predispositions to learn family-specific song. Future work may also examine the mechanisms underlying HVC tempo tuning. A primary candidate should be the role of the zinc transporter protein ZIP11, which partially underlies the genetic predisposition to sing at a particular tempo (Mets et al., 2021).

Individual-specific HVC tempo tuning may also impact the timing of call exchanges. Periodic neural discharges in HVC are triggered by the onset of a partner's vocal response in duetting white-browed sparrow weavers. This serves to synchronize the HVCs of the duetting birds and increase the coordination of their vocalizations (Hoffmann et al., 2019). Although Bengalese finches and related species do not perform song duets, they do exchange precisely-timed antiphonal calls (Benichov et al., 2016; Benichov and Vallentin, 2020; Elie et al., 2010). Future work should examine whether HVC tempo tuning impacts which birds a given Bengalese finch can synchronize antiphonal calls with. In particular, do males with tuning to faster tempos exchange calls more readily when the inter-call interval of the partner is lower?

Taken together, the results presented here suggest that a number of behavioral features that are relevant to inter-individual coordination may be communicated via song tempo or call

features. The identified relationship between song tempo and call timing in males may be subserved by shared HVC circuitry, which may be tuned to encode and produce vocalizations of particular tempos in learning and/or adulthood.

REFERENCES

Aronov, D., & Fee, M. S. (2012). Natural Changes in Brain Temperature Underlie Variations in Song Tempo during a Mating Behavior. *PLoS ONE*, *7*(10), 1–10. https://doi.org/10.1371/journal.pone.0047856

Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., & Tchernichovski, O. (2016). The forebrain song system mediates predictive call timing in female and male zebra finches. *Current Biology*, *26*(3), 309–318. https://doi.org/10.1016/j.cub.2015.12.037

Benichov, J. I., & Vallentin, D. (2020). Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-019-13938-0

Bolund, E., Schielzeth, H., & Forstmeier, W. (2012) Singing activity stimulates partner reproductive investment rather than increasing paternity success in zebra finches. *Behavior, Ecology, and Sociobiology 143,* 549-567.

Both, C., Dingemanse, N. J., Drent, P. J., & Tinbergen, J. M. (2005). Pairs of extreme avian personalities have highest reproductive success. *Journal of Animal Ecology*, *74*(4), 667–674. https://doi.org/10.1111/j.1365-2656.2005.00962.x

Bouchard, K. E., & Brainard, M. S. (2016). Auditory-induced neural dynamics in sensory-motor circuitry predict learned temporal and sequential statistics of birdsong. *Proceedings of the National Academy of Sciences*, *113*(34), 9641–9646. https://doi.org/10.1073/pnas.1606725113

Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(36), 11155–11160. https://doi.org/10.1073/pnas.1505768112

Brumm, H., Zollinger, S. A., & Slater, P. J. B. (2009). Developmental stress affects song learning but not song complexity and vocal amplitude in zebra finches. *Behavioral Ecology and Sociobiology*, *63*(9), 1387–1395. https://doi.org/10.1007/s00265-009-0749-y

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6641–6646. https://doi.org/10.1073/pnas.1522306113

Cooper, B. G., & Goller, F. (2006). Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *Journal of Neurophysiology*, *95*(6), 3798–3809. https://doi.org/10.1152/jn.01123.2005

Cooper, B. G., Mendez, J. M., Saar, S., Whetstone, A. G., Meyers, R., & Goller, F. (2012). Age-related changes in the Bengalese finch song motor program. *Neurobiology of Aging*, *33*(3), 564-568.

d'Amelio, P.B., Trost, L., ter Maat, A. (2017A). Vocal exchanges during pair bond formation and maintenance in the zebra finch. *Frontiers in Zoology 14*(13). https://doi.org/10.1186/s12983-017-0197-x

d'Amelio, P. B., Klumb, M., Adreani, M. N., Gahr, M. L., & ter Maat, A. (2017B). Individual recognition of opposite sex vocalizations in the zebra finch. *Scientific Reports 7*, 5579. https://doi.org/10.1038/s41598-017-05982-x

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1541), 847–852. https://doi.org/10.1098/rspb.2004.2680

Dougherty, L. R. (2021). Meta-analysis reveals that animal sexual signalling behaviour is honest and resource based. *Nature Ecology & Evolution 5,* 688 - 699. https://doi.org/10.1038/s41559-021-01409-z

Dunning, J. L., Pant, S., Bass, A., Coburn, Z., & Prather, J. F. (2014). Mate choice in adult female Bengalese finches: Females express consistent preferences for individual males and prefer female-directed song performances. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089438

Dunning, J. L., Pant, S., Murphy, K., & Prather, J. F. (2020). Female finches prefer courtship signals indicating male vigor and neuromuscular ability. *PLoS ONE*, *15*(1), 1–19. https://doi.org/10.1371/journal.pone.0226580

Elie, J. E., Mariette, M. M., Soula, H. A., Griffith, S. C., Mathevon, N., & Vignal, C. (2010). Vocal communication at the nest between mates in wild zebra finches: A private vocal duet? *Animal Behaviour*, *80*(4), 597–605. https://doi.org/10.1016/j.anbehav.2010.06.003

Elie, J.E., & Theunissen, F.E. (2018). Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications* 9(1): 4026. https://doi.org/10.1038/s41467-018-06394-9.

Geberzahn, N., & Deregnaucourt, S. (2020). Individual vocal recognition in zebra finches relies on syllable structure than song syllable order. *Journal of Experimental Biology 223*(9). https://doi.org/10.1242/jeb.220087

Gil, D., & Gahr, M. (2002). The honesty of bird song: Multiple constraints for multiple traits. *Trends in Ecology and Evolution*, *17*(3), 133–141. https://doi.org/10.1016/S0169-5347(02)02410-2

Gil, D., Naguib, M., Riebel, K., Rutstein, A., & Gahr, M. (2006). Early condition, song learning, and the volume of song brain nuclei in the zebra finch (Taeniopygia guttata). *Journal of Neurobiology*, *66*(14), 1602–1612. https://doi.org/10.1002/neu.20312

Gill, L. F., Goymann, W., Maat, A. Ter, & Gahr, M. (2015). Patterns of call communication between group-housed zebra finches change during the breeding cycle. *ELife*, *4*, 1–23. https://doi.org/10.7554/eLife.07770

Glaze, C., & Troyer, T. (2006). Temporal structure in zebra finch song: implications for motor coding. *Journal of Neuroscience, 23*(3), 991-1005. https://doi.org/10.1523/JNEUROSCI.3387-05.2006

Griffith, S. C. (2019). Cooperation and Coordination in Socially Monogamous Birds: Moving Away From a Focus on Sexual Conflict. *Frontiers in Ecology and Evolution*, *7*(455), 1–15. https://doi.org/10.3389/fevo.2019.00455

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, *421*(6920), 65-69. https://doi.org/10.1038/nature01221

Hall, M. L., & Magrath, R. D. (2007). Temporal coordination signals coalition quality. *Current Biology*, *17*(11), 406–407. https://doi.org/10.1016/j.cub.2007.04.022

Harris, M. R., & Siefferman, L. (2014). Interspecific competition influences fitness benefits of assortative mating for territorial aggression in Eastern Bluebirds (Sialia sialis). *PLoS ONE*, *9*(2), 1–6. https://doi.org/10.1371/journal.pone.0088668

Hoffmann, S., Trost, L., Voigt, C., Leitner, S., Lemazina, A., Sagunsky, H., Abels, M., Kollmansperger, S., Maat, A. Ter, & Gahr, M. (2019). Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10593-3

Jaffe, P. I., & Brainard, M. S. (2020). Acetylcholine acts on songbird premotor circuitry to invigorate vocal output. *ELife*, *9*, 1–26. https://doi.org/10.7554/eLife.53288

Janata, P., & Margoliash, D. (1999). Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. *The Journal of Neuroscience*, *19*(12), 5108-5118.

Kalra, S., Yawatkar, V., James, L., Sakata, J., & Rajan, R. (2021). Introductory gestures before songbird vocal displays are shaped by learning and biological predispositions. *Proceedings of the Royal Society B, 288*(1943). https://doi.org/10.1098/rspb.2020.2796

Kosche, G., Vallentin, D., & Long, M. A. (2015). Interplay of inhibition and excitation shapes a premotor neural sequence. *Journal of Neuroscience*, *35*(3), 1217–1227. https://doi.org/10.1523/JNEUROSCI.4346-14.2015

Kozhevnikov, A. A., & Fee, M. S. (2007). Singing-related activity of identified HVC neurons in the zebra finch. *Journal of Neurophysiology*, *97*(6), 4271–4283. https://doi.org/10.1152/jn.00952.2006

Lewicki, M. S., & Arthur, B. J. (1996) Hierarchical Organization of Auditory Temporal Context Sensitivity. *The Journal of Neuroscience* 16(21), 6987–98. https://doi.org/10.1523/JNEUROSCI.16-21-06987.1996.

Long, M. A., & Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, *456*(7219), 189–194. https://doi.org/10.1038/nature07448

Ma, S., Ter Maat, A., & Gahr, M. (2020). Neurotelemetry reveals putative predictive activity in HVC during call-based vocal communications in zebra finches. *Journal of Neuroscience*, *40*(32), 6219–6227. https://doi.org/10.1523/JNEUROSCI.2664-19.2020

Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *Journal of Neuroscience*, *3*(5), 1039–1057. https://doi.org/10.1523/jneurosci.03-05-01039.1983

Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *Journal of Neuroscience*, *6*(6), 1643–1661. https://doi.org/10.1523/jneurosci.06-06-01643.1986

Margoliash, D., & Fortune, E. S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *Journal of Neuroscience*, *12*(11), 4309–4326. https://doi.org/10.1523/jneurosci.12-11-04309.1992

Markowitz, J. E., Liberti, W. A., Guitchounts, G., Velho, T., Lois, C., & Gardner, T. J. (2015). Mesoscopic patterns of neural activity support songbird cortical sequences. *PLoS Biology*, *13*(6), 1–20. https://doi.org/10.1371/journal.pbio.1002158

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, *71*(2 PART 2), 1–25. https://doi.org/10.1037/h0029144

Marler, P. (1977). Selective Vocal Learning in a Sparrow. *Science*, *198*(4316), 519–521. https://doi.org/10.1126/science.198.4316.519

Marler, P. (1997). Three models of song learning: evidence from behavior. *Journal of Neurobiology 33*(5): 501-516.

McCasland, J. S., & Konishi, M. (1981). Interaction between auditory and motor activities in an avian song control nucleus. *Proceedings of the National Academy of Sciences*, *78*(12), 7815–7819. https://doi.org/10.1073/pnas.78.12.7815

Mets, D. G., & Brainard, M. S. (2018). Genetic variation interacts with experience to determine interindividual differences in learned song. *Proceedings of the National Academy of Sciences*, *115*(2), 421–426. https://doi.org/10.1073/pnas.1713031115

Mets, D.G., Mehaffey, H.M., Colquitt, B.M., Brainard, M.S. (2021). Heritable differences in synaptic zinc-transporter levels drive variation in learned birdsong. *BioRxiv* preprint. https://doi.org/10.1101/2021.05.01.442260

Mets, D. G., & Brainard, M. S. (2019). Learning is enhanced by tailoring instruction to individual genetic differences. *ELife*, *8*, 1–15. https://doi.org/10.7554/eLife.47216

Miller, D. B. (1979). Long-term recognition of father's song by female zebra finches. *Nature*, *280*, 389–391.

Nelson, D. A., & Marler, P. (1993). Innate recognition of song in white-crowned sparrows: a role in selective vocal learning? *Animal Behavior, 46*, 806–808.

Nick, T. A., Konishi, M. (2004). Neural song preference during vocal learning in the zebra finch depends on age and state. *Wiley Periodicals*, 231-242. https://doi.org/10.1002/neu.20087

Nowicki, S., Hasselquist, D., Bensch, S., & Peters, S. (2000). Nestling growth and song repertoire size in great reed warblers: Evidence for song learning as an indicator mechanism in mate choice. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1460), 2419–2424. https://doi.org/10.1098/rspb.2000.1300

Nowicki, S., & Searcy, W. A. (2004). Song function and the evolution of female preferences: Why birds sing, why brains matter. *Annals of the New York Academy of Sciences*, *1016*, 704–723. https://doi.org/10.1196/annals.1298.012

Ord, T. J., & Stamps, J. A. (2008). Alert signals enhance animal communication in “noisy” environments. *Proceedings of the National Academy of Sciences*, *105*(48), 18830–18835. https://doi.org/10.1073/pnas.0807657105

Perez, E. C., Fernandez, M. S. A., Griffith, S. C., Vignal, C., & Soula, H. A. (2015). Impact of visual contact on vocal interaction dynamics of pair-bonded birds. *Animal Behaviour*, *107*, 125–137. https://doi.org/10.1016/j.anbehav.2015.05.019

Prior, N. H., Smith, E., Dooling, R. J., & Ball, G. F. (2020). Familiarity Enhances Moment-to-Moment Behavioral Coordination in Zebra Finch (Taeniopygia guttata) Dyads. *Journal of Comparative Psychology*, *134*(2), 135–148. https://doi.org/10.1037/com0000201

Rajan, R. (2018). Pre-bout neural activity changes in premotor nucleus HVC correlate with successful initiation of learned song sequence. *Journal of Neuroscience*, *38*(26), 5925–5938. https://doi.org/10.1523/JNEUROSCI.3003-17.2018

Richards, D. (1981). Alerting and message components in songs of rufous-sided towhees. *Behavior*, *76*(3–4), 223–249.

Riebel, K. (2009). Song and female mate choice in zebra finches: a review. (2009). *Advances in the Study of Behavior,* 40, 197-238.

Roberts, T. F., Gobes, S. M. H., Murugan, M., Ölveczky, B. P., & Mooney, R. (2012). Motor circuits are required to encode a sensory model for imitative learning. *Nature Neuroscience*, *15*(10), 1454–1459. https://doi.org/10.1038/nn.3206

Roth, T.S., Samara, I., Tan, J., Prrochazkova, E., & Kret, M.E. (2021). A comparative framework of inter-individual coordination and pair-bonding. *Current Opinion in Behavioral Sciences, 39*:98-105. https://doi.org/10.1016/j.cobeha.2021.03.005

Sakata, J. T., Hampton, C. M., & Brainard, M. S. (2008). Social modulation of sequence and syllable variability in adult birdsong. *Journal of Neurophysiology*, *99*(4), 1700–1711. https://doi.org/10.1152/jn.01296.2007

Schuett, W., Dall, S. R. X., & Royle, N. J. (2011). Pairs of zebra finches with similar "personalities" make better parents. *Animal Behaviour*, *81*(3), 609–618. https://doi.org/10.1016/j.anbehav.2010.12.006

Soha, J.A., & Marler, P. (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Animal Behaviour 60*(3), 297–306. https://doi.org/10.1006/anbe.2000.1499

Sossinka, R., & Böhner, J. (1980). Song Types in the Zebra Finch Poephila guttata castanotis. *Zeitschrift Für Tierpsychologie*, *53*(2), 123–132. https://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, *259*(5538), 39–41. https://doi.org/10.1038/259039a0

Vallentin, D., Kosche, G., Lipkind, D., & Long, M. A. (2016). Inhibition protects acquired song segments during vocal learning in zebra finches. *Science*, *351*(6270), 267–271. https://doi.org/10.1126/science.aad3023

Volman, S. F. (1993). Development of neural selectivity for birdsong during vocal learning. *Journal of Neuroscience*, *13*(11), 4737–4747. https://doi.org/10.1523/jneurosci.13-11-04737.1993

Yamahachi, H., Zai, A. T., Tachibana, R. O., Stepien, A. E., Rodrigues, D. I., Cavé-Lopez, S., Lorenz, C., Arneodo, E. M., Giret, N., & Hahnloser, R. H. R. (2020). Undirected singing rate as a non-invasive tool for welfare monitoring in isolated male zebra finches. *PLoS ONE*, *15*, 1–18. https://doi.org/10.1371/journal.pone.0236333

Yoneda, T., & Okanoya, K. (1991). Ontogeny of sexually dimorphic distance calls in Bengalese finches (*Lonchura domestica*). *Journal of Ethology 9*(2), 41–46. https://doi.org/10.1007/BF02350206.

Zann, R., & Cash, E. (2008). Developmental stress impairs song complexity but not learning accuracy in non-domesticated zebra finches (Taeniopygia guttata). *Behavioral Ecology and Sociobiology, 62*(3), 391–400. https://doi.org/10.1007/s00265-007-0467-2

Zhao, W., Garcia-Oscos, F., Dinh, D., & Roberts, T. F. (2019). Inception of memories that guide vocal learning in the songbird. *Science*, *366*, 83–89.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Emily Merfeld
7B0F659992CE4B0...
Author Signature

8/30/2021
Date