



Published in final edited form as:

J Am Stat Assoc. 2013 June 1; 108(502): 514–526. doi:10.1080/01621459.2013.793118.

A Study of Mexican Free-Tailed Bat Chirp Syllables: Bayesian Functional Mixed Models for Nonstationary Acoustic Time Series

Josue G. MARTINEZ,

(Deceased) was recently at the Department of Radiation Oncology, The University of Texas M D Anderson Cancer Center, PO Box 301402, Houston, TX 77230-1402, USA

Kirsten M. BOHN [Assistant Professor],

School of Integrated Science, Florida International University, Miami, FL 33199

Raymond J. CARROLL [Distinguished Professor], and

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143

Jeffrey S. MORRIS [Professor]

The University of Texas M D Anderson Cancer Center, Unit 1411, PO Box 301402, Houston, TX 77230-1402, USA

Kirsten M. BOHN: kbohn@fiu.edu; Raymond J. CARROLL: carroll@stat.tamu.edu; Jeffrey S. MORRIS: jefmorris@mdanderson.org

Abstract

We describe a new approach to analyze chirp syllables of free-tailed bats from two regions of Texas in which they are predominant: Austin and College Station. Our goal is to characterize any systematic regional differences in the mating chirps and assess whether individual bats have signature chirps. The data are analyzed by modeling spectrograms of the chirps as responses in a Bayesian functional mixed model. Given the variable chirp lengths, we compute the spectrograms on a relative time scale interpretable as the relative chirp position, using a variable window overlap based on chirp length. We use 2D wavelet transforms to capture correlation within the spectrogram in our modeling and obtain adaptive regularization of the estimates and inference for the regions-specific spectrograms. Our model includes random effect spectrograms at the bat level to account for correlation among chirps from the same bat, and to assess relative variability in chirp spectrograms within and between bats. The modeling of spectrograms using functional mixed models is a general approach for the analysis of replicated nonstationary time series, such as our acoustical signals, to relate aspects of the signals to various predictors, while accounting for between-signal structure. This can be done on raw spectrograms when all signals are of the same length, and can be done using spectrograms defined on a relative time scale for signals of variable length in settings where the idea of defining correspondence across signals based on relative position is sensible.

Keywords

Bat Syllable; Bayesian Analysis; Chirp; Functional Data Analysis; Functional Mixed Model; Isomorphic Transformation; Nonstationary Time Series; Software; Spectrogram; Variable Overlap

1 Introduction

In animal communications, a common question is whether vocalizations differ among regions or individuals. Regional differences in vocalizations can be indicative of culturally transmitted dialects and vocal learning, as seen in birds (Catchpole and Slater, 1995; Slater 1986) and cetaceans (Payne and Payne, 1985; Deecke et al., 2000; Garland et al., 2011); whereas individual differences in vocalizations indicate the information capacity of vocalizations as “signature” signals (Beecher et al., 1981; Beecher, 1989).

The most common statistical approach to test regional and individual hypotheses using acoustic signals is to use a feature extraction approach by computing a number of predetermined summary measures from the signal, such as duration, beginning frequency, ending frequency, etc., and then perform separate analyses of variances on those original variables (e.g., Bohn et al., 2007; Cerchio et al., 2001; Nelson and Poesel, 2007; Beecher et al., 1981; Davidson and Wilkinson, 2002; Slabbekoorn et al., 2003) or their principal components (Bohn et al., 2007). If these summary measures captured all relevant chirp information, this approach would be reasonable, but given the rich complexity of these data there may be information in the chirp that would be missed by these summaries.

For auditory information, a common and intuitive way to represent sounds is the spectrogram (Bradbury and Vehrencamp, 2011), which captures changes in frequency over time by performing a series of fast Fourier transforms of small, fixed time intervals. Thus, the spectrogram can be considered an “image” of the nonstationary acoustic time series that captures the frequency variation over time. Holan, et al. (2010) introduced methods to analyze bioacoustic signals through statistical modeling of the spectrogram as an image. They present Bayesian regression models of the spectrogram as a predictor of mating success with regularization of the regression surface accomplished by stochastic search variable selection on empirical orthogonal functions (EOFs) of the spectrograms, and construct estimates and inference for the class mean spectrograms from these results. To our knowledge, their paper is the first to use the spectrogram as an image in a statistical modeling framework. Their work provides an excellent explanation for the use of spectrograms as images in statistical models.

In this paper, we present an alternative strategy for using the spectrogram as an image, using a Bayesian statistical model for bioacoustic signals that complements the approach of Holan, et al. (2010). Rather than modeling the spectrogram as a predictor, we model the spectrogram as an image response in a functional mixed model (Morris and Carroll, 2006; Morris, et al., 2011). This allows us to handle aspects of our data not encountered by Holan, et al., including correlation among multiple chirps from the same bat and incorporation of other covariates affecting the spectrograms. It also allows us to perform a within-and between-bat relative variability analysis. We account for within-spectrogram correlation using 2D wavelets rather than EOFs. Also, unlike the data in Holan, et al. (2010), our data have the additional difficulty of different length time series per acoustic signal. We introduce a strategy for dealing with varying signal lengths that involves specifying variable time window overlaps across signals, one that preserves the frequencies of the original signals and produces spectrograms defined on the same relative time grid, which can be interpreted as the relative signal position. We apply our modeling strategy and method to the bat chirp data to characterize regional differences in bat chirp spectrograms while investigating whether individual bats have signature chirps through a relative variability analysis. Our bioacoustical data are from bats that sing like birds, the free-tailed bats (*Tadarida brasiliensis*).

Free-tailed bats range freely in the North American south and southwest as well as in Mexico, Central America and the Caribbean. They form some of the largest mammalian aggregations in the world. Some 20 million gather in Bracken Cave, north of San Antonio, Texas and devour moths that produce the corn earworm, a pest (McCracken 1996). Free-tailed bats are of particular interest because they produce vocalizations (courtship songs) with a degree of complexity that is rarely seen in non-human mammals, making them excellent candidates for research into the production and evolution of human speech. Free-tailed bat songs have a hierarchical structure in which syllables are combined in specific ways to construct phrases and phrases are combined to form songs. There are three types of phrases: chirps, trills and buzzes (Bohn et al. 2008, 2009). We focus on one of the two types of syllables that form chirp phrases, the chirp type B syllable (Bohn et al., 2009), which we simply refer to as chirps. These syllables are of interest because they are the most acoustically complex syllables of the courtship song. We compare multiple chirp syllables from collections of individual bats from two central Texas localities. We assess whether systematic differences exist between bat chirps in different regions, and whether individual bats have characteristic “chirps” such that replicate chirps from the same bat are more similar than chirps from different bats.

We address these questions by modeling the spectrograms as image responses using the functional mixed model framework, first described by Guo (2002), further developed by Morris and Carroll (2006), and extended to image data by Morris, et al. (2011). Morris, et al. (2011) describe a general approach to fitting the functional mixed model which they call isomorphic functional mixed modeling (ISO-FMM). This model relates a sample of functional/image responses to a vector of scalar predictors, each with functional/image coefficients called fixed effect functions. The model also includes random effect functions/images to accommodate correlations among functions, e.g. coming from the same subject. The framework can also incorporate other types of correlation structures across curves such as serial or spatially dependent correlations. In our application, we use an ANOVA structure for the fixed effects, with location indicators for the two regions, as we seek to characterize systematic differences between chirps from the two populations, with one random effect function per individual bat used to accommodate within-bat correlations.

This paper is organized as follows. In Section 2 we give an overview of the application at hand, the data, and their importance and relevance. In Section 3 we describe the model used to fit the data and show how to perform inference in the bat chirp application using the ISO-FMM on the spectrograms. In Section 4, we describe the spectrogram calculation, and explain our strategy for dealing with variable chirp length by computing spectrograms on a relative time axis. We discuss the results of our modeling approach in Section 5 and provide a sensitivity analysis on the smoothing parameters used to construct the spectrograms. Lastly, we give concluding remarks in Section 6.

2 Application Overview

2.1 Chirps of Free-tailed Bats

Chirps are syllables embedded in male songs that are used to communicate with conspecifics. The sequences and arrangements of these syllables follow specific syntactical rules while simultaneously varying from one rendition to the next (Bohn et al., 2009).

Scientists at Texas A&M University collected bat songs from 27 distinct bats, 14 from Austin and 13 from College Station. Each bat is represented in the data by repeated chirp syllables from mating songs, with 25 randomly selected chirp syllables per bat per year, and some bats having multiple sets of 25 chirps. There is no natural ordering to the chirps collected from a bat. After the removal of twelve chirp syllables from the data because of

low quality, the total data set we analyzed includes 788 chirps. A sample chirp for an individual bat is displayed in Figure 1, and sample chirps from each bat are displayed in Figure 2.

Chirps vary in length and shape, not only between bats but also for a single bat. Figure 3 shows boxplots of the length of each chirp in each of the 32 sets of 25 chirps. Each boxplot consists of the chirp length (in milliseconds) for a set of 25 chirps for a single bat, with the boxplots are ordered from shortest to longest mean length. Although there is heterogeneity in the chirp lengths, a vast majority of chirps are between 12 and 20 milliseconds long.

Our goal is to test the hypotheses that chirp syllables vary among individuals and between regions, i.e., that there are regional dialects. The detection and characterization of variation in vocalizations is complicated by the rich, complex nature of the signal (see Figure 1A). We address these issues by delineating systematic spatial discrepancies in the “sound image” or spectrogram through modeling the spectrogram image as a response in a statistical model. The structure of a chirp is much more evident and easily modeled in the spectrogram than in the raw time series (see Figure 1B vs. Figure 1A). Spectrograms are constructed by taking overlapping, windowed Fourier transforms of the raw signal. Once the Fourier coefficients are obtained, the norm of each is calculated and the squared values of these norms are stored as pixel values on the image. These squared values of the norm of Fourier coefficients are usually referred to as the power spectrum, where a large value at a particular frequency indicates the strong presence of that frequency in the data. The chirp in Figure 1B is mainly composed of frequencies between 20–60 kilohertz (kHz), that start at about 40 kilohertz (kHz) and slowly decrease to 20 kHz from 0 to 8 milliseconds into the chirp. The bat then transitions to predominant frequencies at 60 kHz that slowly decrease back down to 40 kHz and then rise up to 60 kHz towards the end of the chirp. Frequencies above ~ 80 kHz are harmonics of the fundamental signal.

2.2 Literature Review

Besides Holan, et al. (2010), who we believe were the first to rigorously model the spectrogram as an image in a statistical model, others have published work that uses spectrograms to capture animal vocalizations of all types and in many species. Examples of these methods include that of Brandes et al. (2006), who classified the calls derived from crickets and frogs among surrounding sounds in the Costa Rican rainforest using image processing techniques that employ blur and other filters to identify the calls of these animals. Bardeli, et al. (2010), on the other hand, propose the use of a novelty measure derived directly from the spectrogram to detect the calls of birds, the Eurasian bittern and Savi's warbler. Moreover, Bardeli, et al. (2010) provided an extensive review of the literature that deals with classification via feature selection in spectrograms. Russo and Jones (2002) and Obrist, et al. (2004) provided analytic studies aimed at detecting different bat species via their echolocation calls. Russo and Jones (2002) studied the calls of 22 bat species in Italy, while Obrist, et al. (2004) attempted to identify 24 species of bats in Switzerland. Both groups approached the problem of identification using call measures, i.e., duration, start frequency, end frequency, etc. They built classifiers using these measures and tried to distinguish between species using these classifiers.

Additionally, nonstationary time series approaches have been used to model this type of data, for example echolocation calls from *Nyctalus noctula* bats (Gray et al., 2005; Jiang et al., 2006). These two works have extended methodology to not only model but also forecast chirp behavior. Gray et al., (2005) used discrete Euler(p) models to model and forecast the observations and introduced instantaneous frequency and spectrum for M-stationary processes. Jiang et al. (2006) introduced a general framework for modeling nonstationary time series. They used a time deformation in the form of the Box-Cox transformation so that

the resulting time series become stationary processes which they term $G(\lambda)$ -stationary processes. Our bat vocalization data, by contrast, present greater challenges in that mating calls are far more complex than echolocation calls, which generally exhibit linear behavior. Our goal is not to forecast chirp behavior, the focus of many time series methods, but rather to characterize individual and regional differences within the domain of the chirps, a focus more typical of functional data analysis methods. Thus, we use a functional data analysis approach to analyze these data.

3 Model

3.1 Isomorphic Functional Mixed Models on Spectrograms

Here, we briefly review the functional mixed model and describe how we use a Bayesian, isomorphic modeling approach to apply it to image data, and how we use the model to obtain the desired inference in our bat chirp application.

To characterize individual and regional variability in the chirp spectrograms, we choose models that treat the spectrogram image as a response. Various models have been proposed to model functional data as a response that is regressed on a set of predictors. Functional mixed models incorporate general functional fixed and random effects that can accommodate many different types of modeling and experimental designs. Functional mixed models were introduced by Guo (2002) for independent and identically distributed residual errors, and then further extended by Morris and Carroll (2006) to include correlation among random effect functions and correlation within and among residual error functions, thus accommodating multi-level functional data.

Consider a set of functional data $Y_i(t)$, $i = 1, \dots, n$. The functional mixed model (FMM) described by Morris and Carroll (2006) is

$$Y_i(t) = \sum_{a=1}^p X_{ia} B_a(t) + \sum_{b=1}^q Z_{ib} U_b(t) + E_i(t), \quad (1)$$

where $X = \{X_{ia}, i = 1, \dots, n; a = 1, \dots, p\}$ and $Z = \{Z_{ib}, i = 1, \dots, n; b = 1, \dots, q\}$ are the $n \times p$ and $n \times q$ design matrices associated with fixed and random effects function sets $\mathbf{B}(t) = \{B_1(t), \dots, B_p(t)\}$ and $\mathbf{U}(t) = \{U_1(t), \dots, U_q(t)\}$, respectively. Here $\mathbf{E}(t) = \{E_1(t), \dots, E_n(t)\}$ is the set of residual error functions. Morris and Carroll (2006) assume that $U_b(t) \sim \mathcal{MGPP}(P, Q)$, which is a mean-zero multivariate Gaussian process with $q \times q$ between-function covariance matrix P characterizing the between-function covariance and covariance surface $Q(t, t')$ characterizing the within-function autocovariance of the random effect functions. The residual error functions $E_i(t) \sim \mathcal{MGPP}(R, S)$ with $n \times n$ between-function covariance matrix R and within-function covariance surface $S(t, t')$. These general assumptions make this model capable of capturing complex behaviors and interactions between and within the functions. The model also accommodates multiple levels of random effect functions or residuals varying by strata, but these are not needed here. The surfaces Q and S become $T \times T$ covariance matrices when all functions are sampled on the same grid t of length T and inference is only desired at these grid points.

This model was extended to higher-dimensional functions such as image data by Morris, et al. (2011). We use this approach to model our spectrograms, which are treated as image or two-dimensional functional data. Section 4 contains details of the spectrogram calculation. Our functional responses, $Y_i(f, t)$, $i = 1, \dots, n = 788$ are the power spectra of the chirps that depend on frequency, f , and time, t , and are represented by the FMM

$$Y_i(f, t) = \sum_{a=1}^p X_{ia} B_a(f, t) + \sum_{b=1}^q Z_{ib} U_b(f, t) + E_i(f, t). \quad (2)$$

The random effects and residual functions in this model follow the same assumptions as in (1). In our data, we have two fixed effect images, $B_1(f, t)$ and $B_2(f, t)$, corresponding to the mean spectrogram for bats from Austin and College Station, respectively, so that $p = 2$. The random effect images $U_b(f, t)$ are defined for each bat and represent the deviation of bat b 's mean spectrogram from its location (Austin or College Station) mean. Since we have 27 bats, $q = 27$. We model the bats as independent ($P = I_q$) and the chirps as independent ($R = I_n$) since the chirps are not taken consecutively. If the chirps were taken consecutively or the individual bats were known to be related by some structure, then this could be accommodated by appropriate specification of R and P , respectively. The inclusion of independent random effect images for each bat accounts for the covariance among replicate chirp spectrogram images from the same bats.

As described in Section 4, we obtain the spectrograms from all chirps on the same discrete lattice of frequency by time, with frequency grid $f = (f_1, \dots, f_F)$ of size F and relative time grid $t = \{1/(T+1), \dots, T/(T+1)\}$ of length T . We can simplify the notations by vectorizing the spectrogram images, as well as the fixed effect, random effect and residual error images of the functional mixed model. Let Y'_i denote the vectorized version of the i^{th} observation, $Y'_i(f, t)$, and concatenate the vectorized images as rows to form a $n \times D$ matrix Y' , where $D = T * F$ represents the size of the time by frequency grid, i.e., the number of pixels in the spectrogram. We can write the discrete space functional mixed model as

$$Y' = XB' + ZU' + E', \quad (3)$$

where B' , U' and E' are the vectorized versions of the corresponding quantities in model (2), of dimension $p \times D$, $q \times D$, and $n \times D$, respectively. The rows of U' follow a multivariate normal distribution $MVN(0, P \otimes Q)$ with $P = I_q$ and Q a $D \times D$ within-spectrogram covariance matrix. The rows of the residual matrix E'_i are $MVN(0, R \otimes S)$ with $R = I_n$ and S a $D \times D$ within-spectrogram covariance matrix.

We do not fit model (3) directly, but instead use the so-called *isomorphic modeling approach* described by Morris, et al. (2011). This involves first mapping the data to an alternative domain using an isomorphic transformation, $Y_i^* = \mathcal{I}(Y'_i)$, i.e., a lossless, invertible transform that preserves all information in the original functions, and then fitting the functional mixed model in the alternative domain, before transforming back to the original domain for inference. This general strategy could be used with various transforms, although as in Morris, et al. (2003) and Morris and Carroll (2006), here we use wavelet transforms, which provide a multiresolution decomposition of the function. Let $\mathcal{I}(\cdot)$ represent the 2D wavelet transform, which is linear so the transformation can be written as $Y^* = Y' W^T$, where W^T is the wavelet transform matrix. For this application, we choose the square, non-separable 2D wavelet transform, which captures locality in both columns and rows of the spectrogram. Applying this transform to an image of grid size D yields a set of wavelet coefficients of size D^* with the wavelet coefficients triple-indexed by scale $v = 1, \dots, V$, orientation $l = 1, 2, 3$ (representing the row, column, and cross-product wavelet coefficients), and location $k = 1, \dots, K_{vl}$. The number of coefficients $D^* = \sum_{v=1}^V \sum_{l=1}^3 K_{vl}$ is on the same order as D , with the specific number subject to the choice of wavelet, boundary condition, and number of decomposition levels V .

The wavelet domain model corresponding to model (3) is given by

$$Y^* = XB^* + ZU^* + E^*, \quad (4)$$

where B^* , U^* , and E^* are the wavelet-space versions of the fixed effect, random effect and error images, respectively, with columns triple-indexed by wavelet scale v , orientation l , and location k . Based on the linear wavelet transform, we see that the wavelet domain within-function covariance matrices in the matrix normal distributions of the random effects matrix, U^* , and the residuals matrix E^* become $I_q \otimes Q^* = I_q \otimes \mathcal{W} Q \mathcal{W}^T$ and $I_n \otimes S^* = I_n \otimes \mathcal{W} S \mathcal{W}^T$, where \mathcal{W} is the inverse wavelet transform matrix such that $\mathcal{W}^T \mathcal{W} = I_D$. Based on the whitening property of the wavelet transform (e.g., Johnstone and Silverman, 1997), we make simplified wavelet domain assumptions on these covariance matrices, making them diagonal $D^* \times D^*$ matrices, where $Q^* = \text{diag}(q_{v,lk}^*)$ and $S^* = \text{diag}(s_{v,lk}^*)$. This independence assumption in the wavelet domain greatly reduces the dimensionality of the fitted covariance matrices, and allows model (4) to be fit by column, and yet does not in general induce independence within the spectrogram. The heteroscedasticity across v , l , and k in the wavelet domain leads to covariances among the pixels of the time-frequency domain that are spatial in nature and nonstationary, allowing for different variances and different autocovariances in different parts of the spectrogram (Morris and Carroll, 2006; Morris et al., 2011). This produces adaptive borrowing of strength across pixels, and thus adaptive smoothing of the image quantities of model (3), most notably of the fixed effects $B_a(f, t)$.

3.2 Bayesian Inference

The wavelet domain model specified in (4) is fit using a Bayesian approach. Although more general choices are possible (Zhu et al., 2011), here we use spike Gaussian-slab priors for the fixed effects in the wavelet domain, $B_{avlk}^* = \gamma_{avlk} \text{Normal}(0, \tau_{avl}) + (1 - \gamma_{avlk}) \delta_0$, where B_{avlk}^* is the wavelet coefficient at scale v , orientation l , and location k for fixed effect a , $\gamma_{avlk} \sim \text{Bernoulli}(\pi_{avl})$ is the spike-slab indicator variable, τ_{avl} and π_{avl} are regularization hyperparameters, and δ_0 is a point mass at zero. The use of this prior in the wavelet domain induces adaptive regularization of the fixed effect images in the time-frequency domain, $B_a(f, t)$. Vague proper priors are assumed for the variance components $q_{v,lk}^*$ and $s_{v,lk}^*$ that are centered on conditional maximum likelihood estimates of these quantities, with the prior having a small effective sample size, as outlined in Herrick and Morris (2006) and Morris and Carroll (2006). An empirical Bayes approach is used on the hyperparameters τ_{avl} and π_{avl} which are estimated by the conditional maximum likelihood as described in Morris and Carroll (2006). A random-walk Metropolis within Gibbs sampler is used to perform an MCMC analysis of this wavelet domain model as detailed in Morris, et al. (2011), which is automated using automatic random walk proposal variances for the variance components as described in Herrick and Morris (2006). The posterior samples of B^* can then be transformed back from the wavelet space $B = B^* \mathcal{W}$ to yield estimates and inference on the quantities $B_a(f, t)$ in model (2) on the spectrogram grid. This model can be fit using freely available software (Herrick and Morris, 2006) that has default values for all hyperparameters and only requires the data matrix Y and design matrices X and Z in order to run if one is satisfied with the defaults.

For our data, we have two fixed effect functions, the mean spectrograms for each location, $B_{Austin}(f, t)$ and $B_{CollegeStation}(f, t)$. To compare these, we compute posterior samples of their difference, $B_{Contrast}(f, t) = B_{Austin}(f, t) - B_{CollegeStation}(f, t)$, from which we can compute estimates, credible intervals, and posterior probabilities for Bayesian inference. Since the spectrograms are on a \log_{10} scale, a difference of δ between the mean spectrograms at (f, t) corresponds to a fold-change difference of 10^δ in the local power at frequency f at time t .

Based on this, we compute the posterior probability of a specified fold-change 10^δ as a function of (f, t) by $p^\delta(f, t) = \text{pr}(|B_{\text{Contrast}}(f, t)| > \delta)$. These probabilities can be plotted as a *probability discovery image*, where regions of (f, t) with large p have strong evidence of differences between locations. The quantities $1 - p^\delta(f, t)$ can be interpreted as Bayesian local false discovery rates (FDR), and thresholds on these quantities can be determined using Bayesian FDR considerations (Morris, et al. 2008).

A global threshold can be determined to declare a set of significantly different regions of (f, t) based on a prespecified average Bayesian false discovery rate (FDR) α , which controls the integrated relative measure of false discoveries in flagged regions to be no more than α .

This set can be written as $\mathcal{T}_\alpha^\delta = \{(f, t): p^\delta(f, t) > \varphi_\alpha^\delta\}$, where φ_α^δ is the threshold that controls the average Bayesian FDR at α . In our setting, the threshold φ_α^δ will correspond to the probability discovery spectrogram value for which the cumulative average of the sorted $\{1 - p^\delta(f, t)\}$ is at most α . This is the interpretation of $\varphi_\alpha^\delta = p_{(\xi)}^\delta$, where

$\xi = \max \left\{ z' : \left[(1/z') \sum_{z=1}^{z'} \{1 - p_{(z)}^\delta(f, t)\} \right] \leq \alpha \right\}$ and $p_{(z)}^\delta(f, t)$ are the vectorized and descending-in-order values of the original probability discovery spectrogram on the $F \times T$ grid. By considering effect size and computing probabilities from a statistical model, this approach adjusts for multiple testing while simultaneously considering both practical and statistical significance.

4 Data Preparation

4.1 Data Pre-processing and Spectrogram Calculation

Syllables were cut from recordings of bat mating songs and then prepared using SIGNAL 4.04 (Engineering Design, Berkeley, CA). Recordings had background noise at low frequencies and not all samples were recorded using the same sample rate. Therefore, all signals were high-pass filtered at 5 kHz below the lowest energy of the fundamental frequency and re-sampled so that all files had a sample rate of $S = 250\text{kHz}$.

Here we discuss how to compute the spectrogram for a given bat chirp. Let $C_i(s), j = 1, \dots, N_i$ represent bat chirp i of length N_i , with the sampling rate given by S . The local discrete Fourier transform for this chirp is

$$F_{i,r}(l) = \sum_{s=1}^N w(s) C_i\{s+(r-1)M\} \exp\{-j(s/S-1)l\}, \quad (5)$$

where $j = \sqrt{-1}$, $w(s)$ is a chosen window function of length N , with windows spaced M units apart. This is computed on the series of $F = N/2 + 1$ frequencies given by $f_l = 2\pi l/N, l = 0, \dots, N/2$, and for a series of $T_i = \lfloor (N_i - N)/M + 1 \rfloor$ window centers given by $t_r = (r-1)M/S, r = 1, \dots, T_i$. Oppenheim (1970) first characterized this method of obtaining spectrograms. The spectrograms used in our model are the result of converting the spectrum obtained from the discrete Fourier transform (DFT) into decibels. For the i^{th} chirp, the resulting spectrogram is then $Y_i(f_l, t_r) = 20 \log_{10} |1 + F_{i,r}(l)|$, where $|\cdot|$ is the complex norm, defined on the grid of F frequencies $f_l^* S/(2\pi)$.

In order to compute the spectrogram, one must specify three user parameters: window function $w(s)$, window size N , and window spacings M . Rectangular windows may seem like a natural choice, but induce undesirable high frequency artifacts in the spectrogram (Oppenheim, 1970). It is recommended to use a function that emphasizes the middle section

of the window and de-emphasizes the edges (Ziemer et al., 1983; Chu, 2008). As done by others (Oppenheim, 1970; Holan et al., 2010), we use the Hamming window

$$w(l) = \begin{cases} 0.575 + 0.425 \cos\left(\frac{2\pi l}{N}\right) & \text{for } l \in (-N/2, N/2); \\ 0 & \text{for } |l| \geq N/2. \end{cases} \quad (6)$$

The window size N must be chosen carefully, since if too small, it leads to a spectrogram with a very coarse time grid and if too large leads to a coarse frequency grid. We choose $N = 256$, which we found yields a reasonable grid on both the frequency ($F = 129$) and time ($T_j \approx 80$) scales. For the window spacings M , a larger choice will result in a smaller time grid and less smoothing in time, but have quicker calculation time; whereas a smaller M will result in a larger time grid, more smoothing, and slower calculation times. One must choose $M < N$, otherwise the windows would not overlap and the spectrogram would skip some time points. The visual effects of choice of N and M on the spectrogram can be seen in Supplementary Figures 9–12 in the Supplementary Material. In our application, we used window sizes of $M = 46$, which is $0.18N$.

For bioacoustical data, if all n signals are of the same length $\mathcal{N}_i \equiv \mathcal{N}$, this would yield a series of n spectrograms, each of dimension $F \times T$, where $F = N/2 + 1$ and $T = \lfloor (\mathcal{N} - N)/M + 1 \rfloor$, and to which the methods described in Section 3.1 could be directly applied. However, as noted in Section 2, the chirp signals in our application are of variable length, leading to spectrograms of a variable time range. In this case, it is not immediately clear how to combine information across these spectrograms for any kind of joint population-level inference such as is the goal of the ISO-FMM here and the methods described in Holan, et al. (2010). Thus, we introduce a strategy involving variable window spacings that yields spectrograms on a common relative time grid, and discuss its implications and suitability for this and other applications.

4.2 Variable Window Overlaps for Signals of Varying Length

When the signals are of variable length \mathcal{N}_i but have a common sampling rate S , we propose to keep the window width N constant across signals, but to vary the window spacings $M_j = (\mathcal{N}_i - N)/(T - 1)$ across signals of varying length. This results in all spectrograms being defined for a common grid of frequencies (defined on the raw time scale) $f = \{f_j\}$; $f_j = S * I / N$; $I = 0, \dots, N/2$ of length $F = N/2 + 1$, and for a series of T time grid points. The use of a constant window size N across signals ensures each spectrogram is computed on the same set of frequencies, and the varying window spacings ensure the spectrograms for all signals are computed on a time grid of constant size. This time grid should not be interpreted as absolute time, but relative signal position, defined on $(0,1)$ as $t = \{t_r\}$; $t_r = r/(T + 1)$; $r = 1, \dots, T$, where 0 corresponds to the beginning of a chirp and 1 corresponds to the end of a chirp. The use of this relative time grid effectively compresses/expands the signals on the time axis of the spectrogram, but does not distort the frequencies since the frequencies are still computed on the absolute time scale and have the same range for all signals.

This strategy can also be called a *variable overlap procedure* since the amount of overlap between neighboring windows is given by $O_j = N - M_j$, which varies across signals based on their lengths. The amount of overlap determines the degree of smoothing in the time direction in computing the spectrograms. Our procedure results in variable time smoothing based on chirp length, with longer signals having more smoothing than shorter signals. We also considered other approaches to produce a common grid, including cubic spline interpolation and up-sampling or down-sampling of the chirp signals, but these methods resulted in considerable artifacts and a simulation study demonstrated that our variable overlap procedure results in smaller mean squared error than these alternatives (see the

Supplementary Material, Section 1). In our chirp analysis, we chose this strategy because our primary interest is the internal pitch sequence, not the chirp durations, and our supplemental linear mixed model analysis on duration described below shows that chirp duration does not systematically vary across regions.

To use this variable overlap procedure, one must choose the desired relative time grid size T . Given the requirement that $M_j < N$ for all $j = 1, \dots, n$, one should ensure that $T > N/\min(M_j)$. We chose $T = 80$ for our analysis, which results in average window spacings on the order of $M_j \approx 46$.

Since the duration information is lost in these adjusted spectrograms on the relative signal time scale, this procedure is most natural to use when there is not too much variability in the duration of the signal and when signal duration is not expected to be meaningful in the given application. One can separately model signal duration as a response in an alternative linear mixed model to see if it is related to any outcomes of interest; we performed this analysis for our data and found that duration was not significantly related to region. Additionally, another diagnostic to check whether signal duration has any systematic effect on the spectrogram features is to include duration in the functional mixed model as a continuous predictor. We also performed this analysis, and found that the results of our regional analyses were retained even after the inclusion of signal duration in the model, see the Supplementary Material, Section 4.

5 Data Analysis and Interpretation of Results

5.1 Fitting the Model

As described above, our data consisted of $n = 788$ bat chirps $C_i(s)$, $i = 1, \dots, n$ from a total of $q = 27$ bats, from which we computed spectrograms $Y_i(f, t)$, and defined covariates $X_{i1} = 1$ if chirp i was from a College Station bat, 0 otherwise; $X_{i2} = 1$ if chirp i was from an Austin bat, 0 otherwise; and $Z_{ib} = 1$ if chirp i was from bat b , $b = 1, \dots, q$, 0 otherwise. As in model (2), the functional mixed model is

$$Y_i(f, t) = \sum_{a=1}^2 X_{ia} B_a(f, t) + \sum_{b=1}^{27} Z_{ib} U_b(f, t) + E_i(f, t). \quad (7)$$

As described in Section 4, the spectrogram image matrices $Y_i(f, t)$ are of size $F \times T$, where F is the frequency resolution with f ranging from 0 to 125kHz and T is the time resolution, with t indicating relative times ranging from 0 to 1. Here $F = (N/2) + 1 = 129$, since the size of the window $N = 256$, and $T = 80$. After vectorizing the matrices, $Y_i(f, t)$, model (7) is

$$Y' = XB' + ZU' + E', \quad (8)$$

where Y' is the vectorized data matrix of size 788×10320 . Of course, X is the fixed effect matrix of size 788×2 , with corresponding fixed effect vectorized image matrix, B' , of size 2×10320 . In addition, Z is the random effect matrix of size 788×27 with vectorized random effect image matrix, U' , of size 27×10320 . Finally, the vectorized residual image matrix E' is of size 788×10320 . We then transformed model (8) into the wavelet domain using the square, non-separable, two-dimensional wavelet transform, and fit this model using the Bayesian approach and freely available software described in Section 3.

After a burn-in of 1000, we obtained 20,000 MCMC samples of all model parameters, keeping every 10. Trace plots of the fixed effects and variance components for selected

wavelet coefficients suggest the MCMC converged and mixed well, which is not surprising since given that the modeling amounts to a series of parallel Gaussian linear models, whose posteriors are well-behaved and stable. Next, we applied the inverse 2D discrete wavelet transforms to the wavelet domain posterior samples to obtain posterior samples of model (8), where our Bayesian inference was done on the frequency-relative time grid of the spectrograms. Specifically, we computed the posterior samples of the difference between the Austin and College Station mean spectrograms, $B_{\text{contras}}(f, t) = B_{\text{Austin}}(f, t) - B_{\text{CollegeStation}}(f, t)$, and then computed the corresponding posterior means and probability discovery images.

5.2 Interpretation of Results

Figure 4 shows the posterior mean of $B_{\text{contras}}(f, t)$, the contrast between the mean Austin and College Station chirp profiles on the frequency-relative time grid of the spectrograms. Positive regions favor bats from Austin; while negative regions favor those from College Station. As can be seen, the dominant frequencies for Austin bats are concentrated in the lower bands than those from the College Station bats. Austin bats have pronounced frequencies in the range of 20–30 kHz, while College Station bats show dominance in the frequencies between 30–50 kHz. Both groups show a decreasing emphasis in the frequencies of preference as the sound wave progresses. College Station bats start with a predominant frequency of about 50 kHz that slowly shifts towards 35 kHz by the middle of the chirp. A parallel frequency shift is observed for the Austin bats; however, their chirps start at frequencies around 40 kHz and slowly change to 20 kHz about halfway through the chirp. The chirp frequencies of both Austin and College Station bats show other dominant regions. In the middle of the chirp syllable, College Station bats switch from a dominant frequency at around 35 kHz to one at 60 kHz. They maintain both frequencies from the midpoint to the end of the chirp, at which time they suddenly switch to the 60-kHz range.

Overall, we see that College Station bat chirps tend to have higher frequencies than those of the Austin bats. Figure 5 shows the probability discovery image for a 1.5-fold change, $p^{1.5}(f, t) = \text{pr}(|B_{\text{Contras}}(f, t)| > \log_{10}(1.5))$, computed as described in Section 3.2. The areas in the contrast plot that were shown to have high probability of 1.5-fold difference in favor of either the Austin or College Station bats are the ones highlighted by the probability plot. The largest of these regions is in the frequency range of 20–30 kHz in the last three quarters of the chirp. The flagged regions of (f, t) corresponding to a global Bayesian FDR threshold of 0.15 are displayed by the white space in Figure 6. The frequency shift is also detected there, where frequency regions of preference for the College Station bats are between 30–50 kHz.

5.3 Sensitivity Analysis

In all the results presented above, we fixed the size of the temporal grid used in the spectrogram at $T = 80$, which is a key parameter in the variable window overlap strategy described in Section 4 that determines the size of the window spacings for each chirp, M_i . To explore the sensitivity of our results to a change in the time grid, we also ran our analyses for choices of $T = \{35, 43, 56, 141\}$, which correspond to average window spacings M_i of roughly $\{106, 86, 66, 26\}$, which as a proportion of N are $\{0.41, 0.34, 0.26, 0.10\}$. Figures showing the corresponding probability discovery images and their 15% Bayesian FDR counterparts are provided in the Supplementary Material, Section 2.

This analysis demonstrates that differing time resolutions do not affect the substantive results in any meaningful way. Regardless of the temporal resolution, we still observe a frequency shift between the chirps of the Austin and College Station bats, characterized by a preference for the lower frequency range of 20–30 kHz in the Austin bats and a preference for the range of 30–50 kHz by the College Station bats. The mating chirps from the two sets

of bats are parallel in that their frequency preference decreases from the start to the middle of the chirp.

5.4 Assessing Individual Bat Signatures

We performed a variance component analysis to assess whether individual bats had distinct signatures, which would be indicated if the bat-to-bat variability in the chirp spectrograms (i.e., random effects) was larger than the chirp-to-chirp within bat variability (i.e., residual error in our model). This type of variance component analysis can be done with the results of our unified Bayesian fitting of the functional mixed model. Given the spectrograms on a grid of size D , the bat-to-bat covariance matrix is given by $Q(D \times D)$, and the chirp-to-chirp covariance matrix is given by $S(D \times D)$. An overall intraclass correlation coefficient measuring the relative percent variation explained by the bat-to-bat variability from the sum of the variability at the bat-to-bat and chirp-to-chirp within bat levels can be computed by $\rho = \text{trace}(Q) / \{\text{trace}(Q) + \text{trace}(S)\}$. This measure $\rho \in (0, 1)$ has the extreme $\rho = 1$ corresponding to the case where all chirps within a bat are identical, and $\rho = 0$ suggesting that two chirps from the same bat are no more alike than two chirps from different bats. Given that $Q^* = WQW^T$ and $S^* = WS^*W^T$ from the wavelet space model (4) and that $W^T W = I_D$, we can easily compute posterior samples for ρ from our posterior samples of $Q^* = \text{diag}(q_{v\ell k}^*)$ and $S^* = \text{diag}(s_{v\ell k}^*)$ given that $\rho = q^* / (q^* + s^*)$ where $q^* = \sum_{v\ell k} q_{v\ell k}^*$ and $s^* = \sum_{v\ell k} s_{v\ell k}^*$. From this, we find that the posterior mean and 90% credible interval for ρ are 0.51 and (0.50, 0.52), respectively, demonstrating that the between-bat variability is slightly greater in magnitude than the chirp-to-chirp within bat variability and considerably bounded away from zero, meaning that two chirps from a given bat are significantly more alike than two chirps from different bats. This suggests that individual bats have a somewhat unique aspect to their chirps that might make them distinguishable from other bats.

6 Conclusion

We analyzed chirp syllables from the mating songs of free-tailed bats from two regions of Texas in which they are predominant: Austin and College Station. Our goal was to discern systematic differences between the mating chirps from bats in two physically distant regions. The data were analyzed by modeling spectrograms of the chirps as responses using the ISO-FMM (Morris et al., 2011). Given the variable chirp lengths, we computed the spectrograms on a relative time scale, interpretable as relative chirp position using a variable window overlap based on chirp length. We used 2D wavelet transforms to capture the correlation within the spectrogram in our modeling and obtain adaptive regularization of the estimates and inference for the region-specific spectrograms. Our model included random effect spectrograms at the bat level to account for correlation among chirps from the same bat, and to perform an analysis to assess relative variability in chirp spectrograms within and between bats. We also performed analyses of chirp duration in both linear mixed models and as a predictor in our functional mixed model, and found that chirp duration did not differ systematically across regions and its inclusion in the FMM did not appreciably change the region differences.

The analysis revealed both regional and individual differences in chirp syllables. We found evidence of highly stereotyped chirp syllables within individuals. This suggests that bats may be able to use chirp syllables to recognize different individuals. At the regional level there is a frequency shift between the two regions, where chirps from College Station bats tend to have higher frequencies, ranging from 30–50 kHz range, and those from Austin bats tend to range from 20–30 kHz. Both groups tended to show decreasing frequencies from the start to the midpoint of the chirp; however the direction and magnitude of the differences shifted between regions along the duration of the chirps and were not mirror images (Fig. 4).

Therefore, not only were frequencies shifted between the two regions but among the spectral pattern, or shape of the signal, as well. Generally, regional variation in songs can be the result of genetic differentiation or vocal plasticity (Catchpole and Slater, 1995; Slabbekoorn and Smith, 2002). Research has shown that free-tailed bats show no genetic differentiation across their entire range because they migrate over such long distances (Russell et al., 2005) and that although song syntax (e.g., the way elements are ordered and combined) varies immensely within individuals, it also does not vary across regions (Bohn et al., 2009). Our finding of regional variation in song syllables suggests that songs may not be entirely innate, like the majority of mammalian vocalizations, but instead can be influenced and modified by experience.

While applied to bat chirp data in these analyses, the modeling of spectrograms of non-stationary time series such as acoustical signals as responses in isomorphic functional mixed models is a general approach that can be used to relate aspects of a signal to various predictors while accounting for the structure between the signals. Although we used wavelets to capture the internal spectrogram structure, isomorphic transforms based on other basis functions could also be used with this method, including EOFs or splines. Although not of interest in this paper, incidentally it is also possible to perform classification using the ISO-FMM approach by fitting the model to training data and then computing posterior predictive probabilities of class given spectrogram for the test data (Zhu, Brown and Morris, 2012).

If all signals are of the same length, the ISO-FMM can be applied directly to the raw spectrograms. If the signals are of different lengths, some type of correspondence must be defined across the signals to perform analyses combining information across signals. Here, we introduce a strategy to define the correspondence based on relative signal position, which yields spectrograms on this relative time scale that are obtained by varying the window overlap across signals according to their lengths. Since these spectrograms lose any information on signal lengths, sensitivity analyses or alternative analyses should be done to assess this factor, as we have done.

In conclusion, we have developed an alternative strategy for statistical analysis of animal vocalizations. While most techniques take individual measurements that can not capture the frequency and amplitude variations over time, our method permits inclusion of the entire signal without losing any information. This is likely more representative of what an animal actually hears than a set of extracted features. Furthermore, our method allows us to develop sophisticated models that incorporate within individual variation and other covariates that are crucial components of animal communication systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Martinez was supported by a post-doctoral training grant from the National Cancer Institute (R25T-CA90301). Morris was supported by a grant from the National Cancer Institute (R01-CA107304). Carroll's research was supported by a grant from the National Cancer Institute (R37-CA05730) and by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST). We thank Richard C. Herrick for implementing the functional mixed model into the WFMM software used to obtain the results presented in this work. Additionally, this work was supported by the Statistical and Applied Mathematical Sciences Institute (SAMSI) Program on the Analysis of Object Data. We also thank the editor, associate editor and reviewers for their insightful comments that helped improve this paper.

References

- Bardeli R, Wolff D, Kurth F, Koch M, Tauchert KH, Frommolt KH. Detecting Bird Sounds in a Complex Acoustic Environment and Application to Bioacoustic Monitoring. *Pattern Recognition Letters*. 2010; 31:1524–1534.
- Beecher MD. Signalling Systems for Individual Recognition: an Information Theory Approach. *Animal Behaviour*. 1989; 38:248–261.
- Beecher MD, Beecher IM, Hahn S. Parent-offspring Recognition in Bank Swallows (*Riparia riparia*): II. Development and Acoustic Basis. *Animal Behaviour*. 1981; 29:95–101.
- Bohn KM, Schmidt-French B, Ma TS, Pollack GD. Syllable Acoustics, Temporal Patterns and Call Composition Vary with Behavioral Context in Mexican Free-Tailed Bats. *Journal of the Acoustical Society of America*. 2008; 124:1838–1848. [PubMed: 19045674]
- Bohn KM, Schmidt-French B, Schwartz C, Smotherman M, Pollak GD. Versatility and Stereotypy of Free-Tailed Bat Songs. *PLoS ONE*. 2009; 4(8):e6746. [PubMed: 19707550]
- Bohn KM, Wilkinson GS, Moss CF. Discrimination of Infant Isolation Calls by Female Greater Spear-Nosed Bats, *Phyllostomus hastatus*. *Animal Behaviour*. 2007; 73:423–432. [PubMed: 18311319]
- Bradbury, JW.; Vehrencamp, SL. *Principles of Animal Communication*. 2. Sunderland MA: Sinauer Associates; 2011.
- Brandes TS, Naskrecki P, Figueroa HK. Using Image Processing to Detect and Classify Narrow-Band Cricket and Frog Calls. *Journal of the Acoustical Society of America*. 2006; 120:2950–2957. [PubMed: 17139751]
- Catchpole, CK.; Slater, PJB. *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge University Press; 1995.
- Cerchio S, Jacobsen JK, Norris TF. Temporal and Geographical Variation in Songs of Humpback Whales, Megaptera Novaeangliae: Synchronous Change in Hawaiian and Mexican Breeding Assemblages. *Animal Behaviour*. 2001; 62:313–329.
- Chu, E. *Discrete and Continuous Fourier Transforms: Analysis, Applications and Fast Algorithms*. Boca Raton: Chapman & Hall/CRC; 2008.
- Davidson SM, Wilkinson GS. Geographic and Individual Variation in Vocalizations by Male *Saccolaryx Bilineata* (Chiroptera: Emballonuridae). *Journal of Mammalogy*. 2002; 83:526–535.
- Deecke VB, Ford JKB, Spong P. Dialect Change in Resident Killer Whales: Implications for Vocal Learning and Cultural Transmission. *Animal Behaviour*. 2000; 60:629–638. [PubMed: 11082233]
- Garland EC, Goldizen AW, Rekdahl ML, Constantine R, Garrigue C, Hauser ND, Poole MM, Robbins J, Noad MJ. Dynamic Horizontal Cultural Transmission of Humpback Whale Song at the Ocean Basin Scale. *Current Biology*. 2011; 21:687–691. [PubMed: 21497089]
- Gray HL, Vijverberga C-PC, Woodward WA. Nonstationary Data Analysis by Time Deformation. *Communications in Statistics*. 2005; 34:163–192.
- Guo W. Functional Mixed Effects Models. *Biometrics*. 2002; 58:121–128. [PubMed: 11890306]
- Herrick, RC.; Morris, JS. Wavelet-Based Functional Mixed Model Analysis: Computational Considerations. *Proceedings, Joint Statistical Meetings, ASA Section on Statistical Computing*; 2006. p. 2051–2053.
- Holan SH, Wikle CK, Sullivan-Beckers LE, Cocroft RB. Modeling Complex Phenotypes: Generalized Linear Models Using Spectrogram Predictors of Animal Communication Signals. *Biometrics*. 2010; 66:914–924. [PubMed: 19764952]
- Jiang H, Gray HL, Woodward WA. Time-Frequency Analysis-G(λ)-Stationary Processes. *Computational Statistics & Data Analysis*. 2006; 51:1997–2028.
- Johnstone IM, Silverman BW. Wavelet Threshold Estimators for Data With Correlated Noise. *Journal of the Royal Statistical Society, Series B*. 1997; 59:319–351.
- McCracken GF. Bats Aloft: A Study of High-Altitude Feeding. *BATS Magazine*. 1996; 14:7–10.
- Morris JS, Baladandayuthapani V, Herrick RC, Sanna PP, Gutstein H. Automated Analysis of Quantitative Image Data using Isomorphic Functional Mixed Models, with Application to Proteomics Data. *Annals of Applied Statistics*. 2011; 5:894–923. [PubMed: 22408711]

- Morris JS, Brown PJ, Herrick RC, Baggerly KA, Coombes KR. Bayesian Analysis of Mass Spectrometry Proteomics Data Using Wavelet Based Functional Mixed Models. *Biometrics*. 2008; 64:479–489. [PubMed: 17888041]
- Morris JS, Carroll RJ. Wavelet-Based Functional Mixed Models. *Journal of the Royal Statistical Society, Series B*. 2006; 68:179–199.
- Morris JS, Vannucci M, Brown PJ, Carroll RJ. Wavelet-Mased Non-parametric Modeling of Hierarchical Functions in Colon Carcinogenesis. *Journal of the American Statistical Association*. 2003; 98:573–583.
- Nelson DA, Poesel A. Segregation of Information in a Complex Acoustic Signal: Individual and Dialect Identity in White-Crowned Sparrow Song. *Animal Behaviour*. 2007; 74:1073–1084.
- Obrist MK, Boesch R, Flückiger PF. Variability in Echolocation Call Design of 26 Swiss Bat Species: Consequences, Limits and Options for Automated Field Identification with a Synergetic Pattern Recognition Approach. *Mammalia*. 2004; 68:307–322.
- Oppenheim AV. Speech Spectrograms Using the Fast Fourier Transform. *IEEE Spectrum*. 1970; 17:57–62.
- Payne K, Payne R. Large Scale Changes Over 19 Years in Songs of Humpback Whales in Bermuda. *Zeitschrift fur Tierpsychologie*. 1985; 68:89–114.
- Rabiner LR, Schafer RW. Introduction to Digital Speech Processing. *Foundations and Trends in Signal Processing*. 2007; 1:1–194.
- Russell AL, Medellin A, McCracken GF. Genetic Variation and Migration in the Mexican Free-Tailed Bat (*Tadarida Brasiliensis Mexicana*). *Molecular Ecology*. 2005; 14:2207–2222.
- Russo D, Jones G. Identification of Twenty-Two Bat Species (Mammalia: Chiroptera) from Italy by Analysis of Time-Expanded Recordings of Echolocation Calls. *Journal of Zoology*. 2002; 258:91–103.
- Slabbekoorn H, Jesse A, Bell DA. Microgeographic Song Variation in Island Populations of the White-Crowned Sparrow (*Zonotrichia Leucophrys Nutalli*): Innovation Through Recombination. *Behaviour*. 2003; 140:947–963.
- Slabbekoorn H, Smith TB. Bird Song, Ecology and Speciation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*. 2002; 140:947–963.
- Slater PJ. The Cultural Transmission of Bird Song. *Trends in Ecology & Evolution*. 1986; 1:94–97. [PubMed: 21227788]
- Zhu H, Brown PJ, Morris JS. Robust, Adaptive Functional Regression in Functional Mixed Model Framework. *Journal of the American Statistical Association*. 2011; 106:1167–1179. [PubMed: 22308015]
- Zhu H, Brown PJ, Morris JS. Robust Classification of Functional and Quantitative Image Data Using Functional Mixed Models. *Biometrics*. 2012; 68:1260–1268. [PubMed: 22670567]
- Ziemer, RE.; Tranter, WH.; Fannin, DR. *Signals and Systems*. New York: Macmillan; 1983.

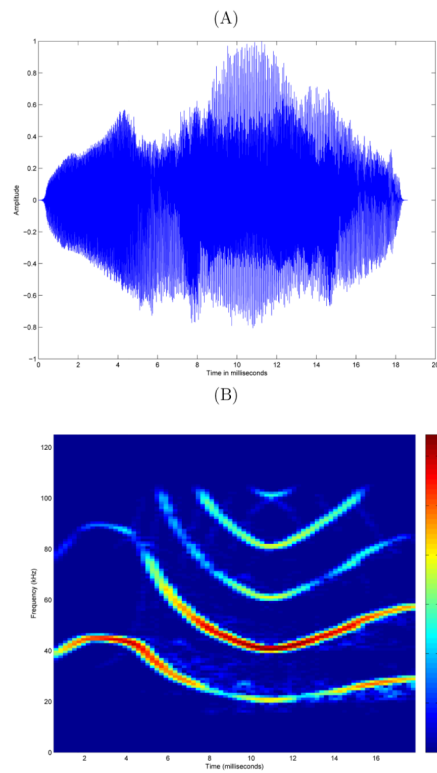


Figure 1.

(A) Shows an example bat chirp and (B) shows its corresponding spectrogram. Time on the x-axis is measured in milliseconds. In plot (A) the y-axis displays the normalized amplitude of the bat chirp so that the absolute value of the amplitude is no larger than 1. In plot (B) the units in the y-axis are in kilohertz (10^3Hz) and those in the z-axis are the dimensionless power derived from the Fourier transform with a frequency resolution of $l = 0, 1, \dots, N/2$, where $N = 256$.

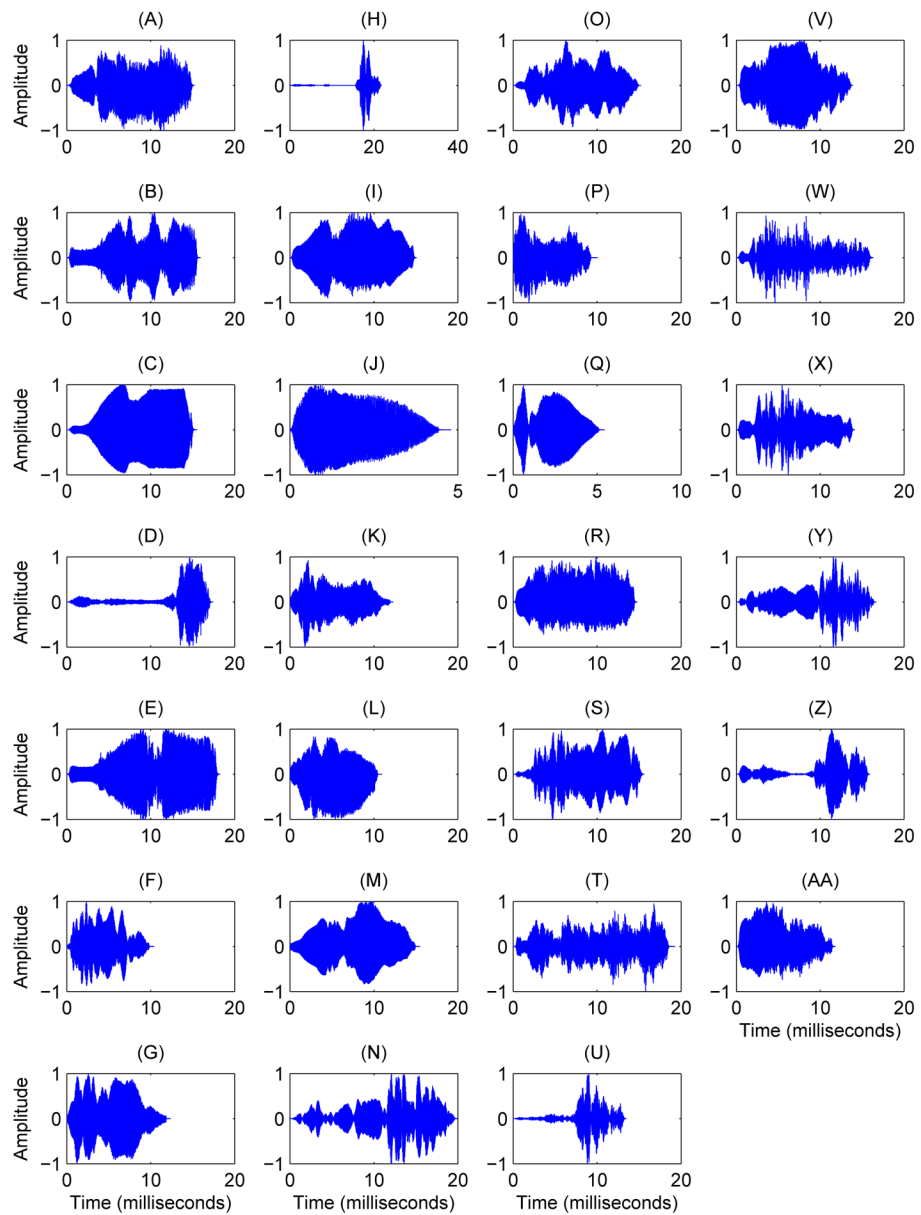


Figure 2. Representative Chirps of the 27 bats in the study. Plots (A) – (N) show sample chirps for the 14 bats from Austin and plots (O)-(AA) show sample chirps for the 13 bats from College Station.

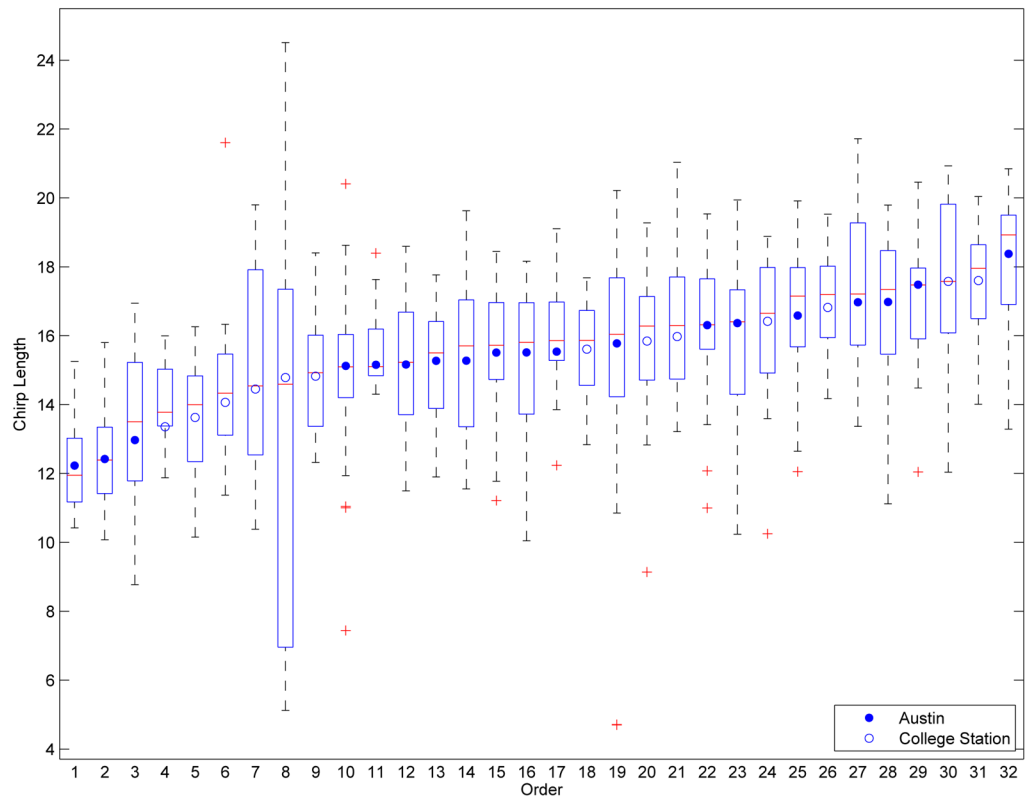


Figure 3. Box plots of the chirp lengths ordered by the average length for each the 32 recording sets. Each box plot shows the distribution of the length of the 25 chirps collected per year. All 27 bats are represented at least once.

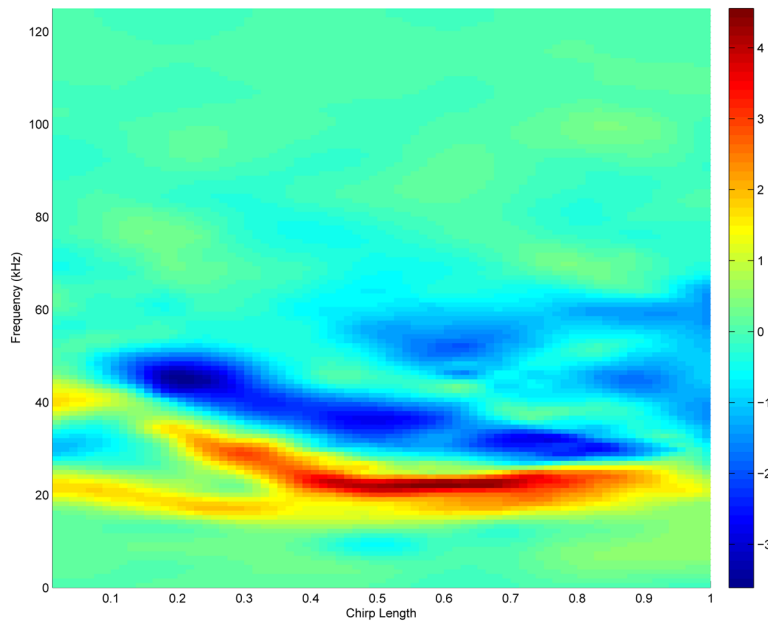


Figure 4. Difference in mean spectrogram inferred from our model. Positive values favor bats from Austin while negative values favor those from College Station. The highly positive (red) regions suggest that, on average, bats from Austin have more pronounced frequencies in the 20–30 KHz range through the length of their chirp. In contrast, bats from College Station have more pronounced frequencies in the 60 KHz range at the end, and between 30–50 KHz in the first two thirds of their chirps.

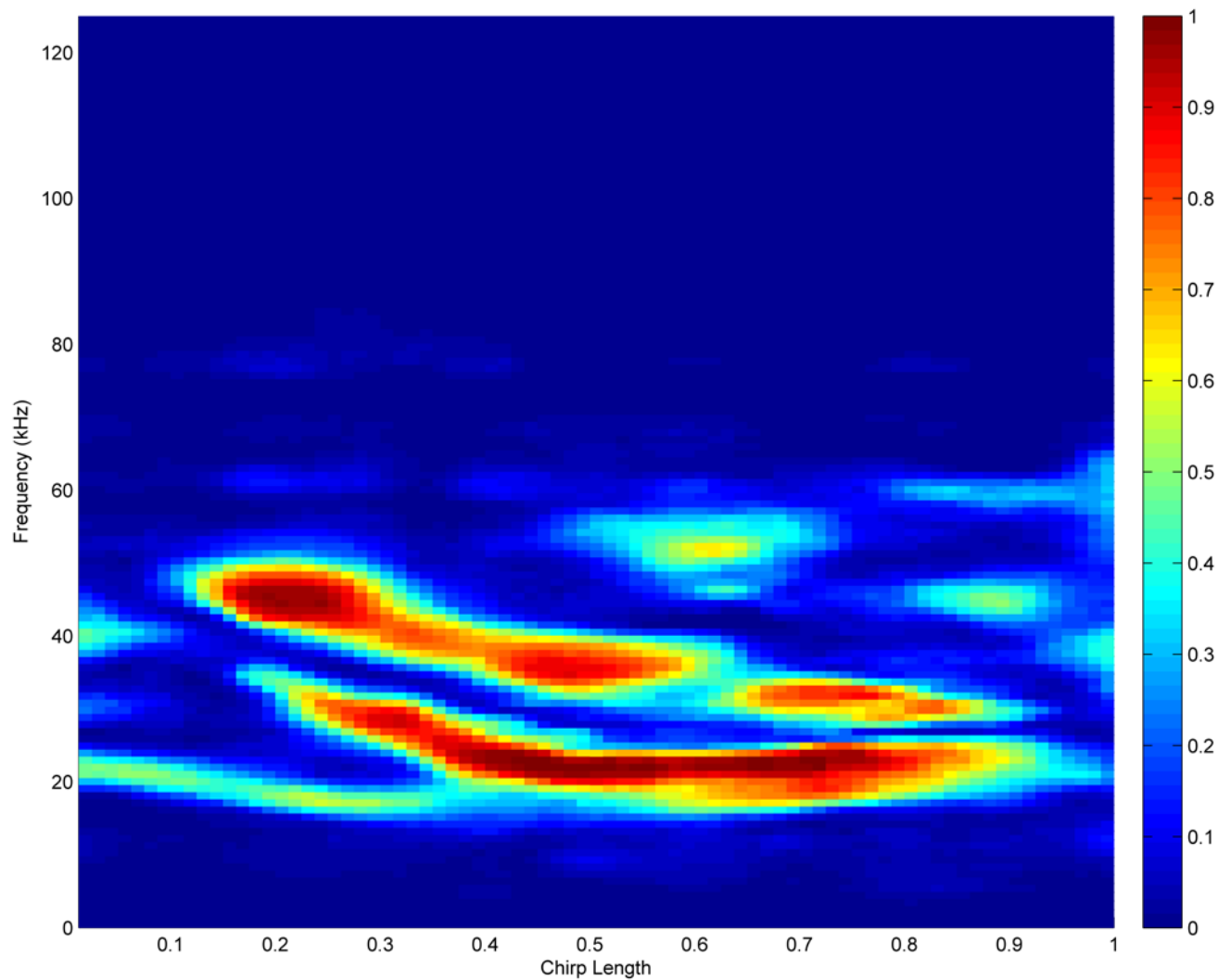


Figure 5.

Posterior probability plot indicating the regions of most pronounced differences between Austin and College Station bats. The probability that there is a 1.5 fold difference between Austin and College Station bats is shown. High probability indicates a very likely chance that there is a difference between Austin and College Station bats, and where that difference is located in terms of frequency (kHz) and location along the chirp.

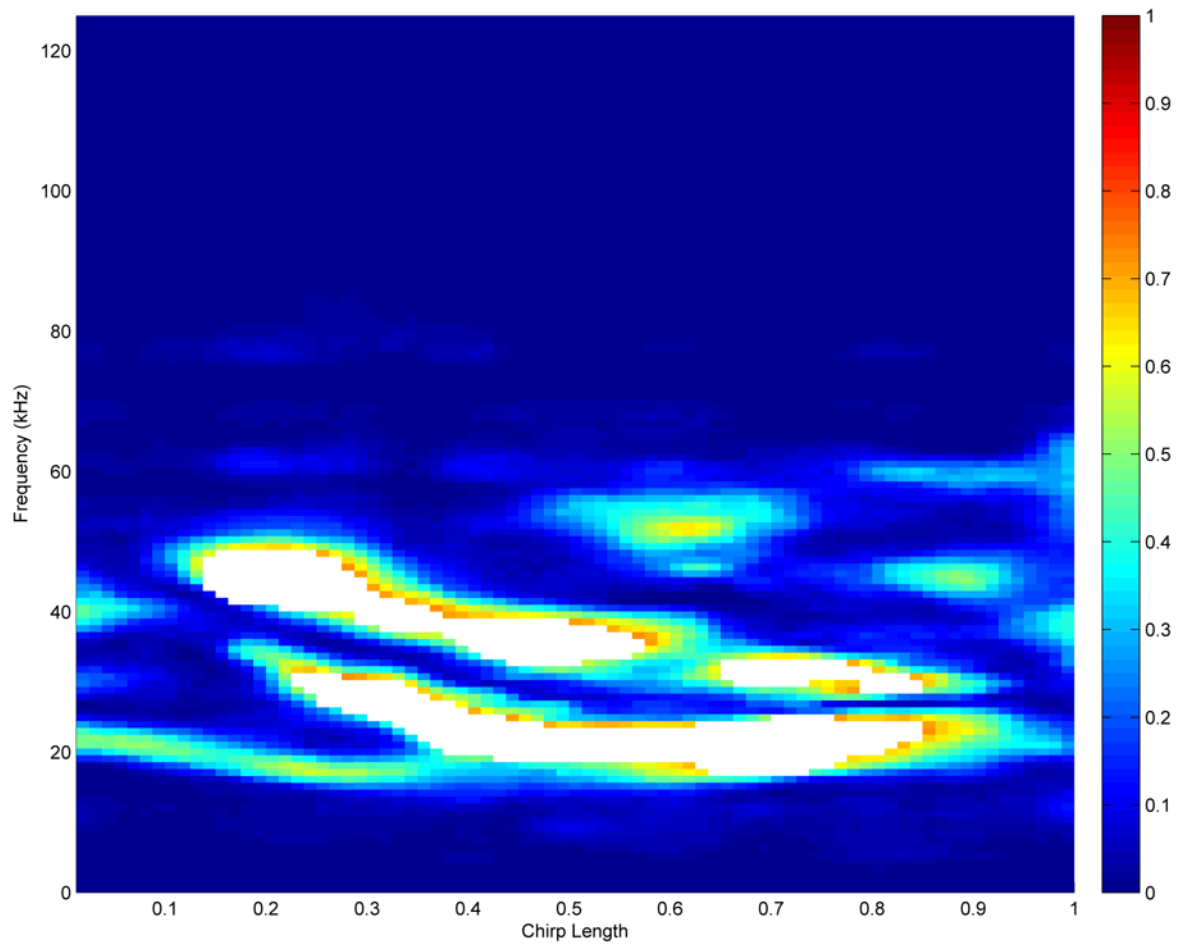


Figure 6. Spectrogram regions where highly significant differences between Austin and College Station bats exist. The white space shows the regions with highest probability of a 1.5 fold difference between Austin and College Station bats, that also have a controlled Bayesian false discovery rate of 15%.