Comparing methodologies for classification of zebra finch distance calls

Prateek K. Sahu,[1] Kimberley A. Campbell,[1] Alexandra Oprea,[2] Leslie S. Phillmore,[2] and Christopher B. Sturdy [1,a,b]

[1]*Department of Psychology, University of Alberta, Edmonton, T6G 2R3, Canada*

[2]*Department of Psychology and Neuroscience, Dalhousie University, Halifax, B3H 4R2, Canada*

Running title: Multiple methods for distance call classification

[a] Also at *Neuroscience and Mental Health Institute, University of Alberta, Edmonton, T6G 2R3, Canada*

[b] Electronic mail:csturdy@ualberta.ca

ABSTRACT

Bioacoustic analysis have been used for a variety of purposes including classifying vocalizations for biodiversity monitoring and understanding mechanisms of cognitive processes. A wide range of statistical methods, including various automated methods, have been used to successfully classify vocalizations based on species, sex, geography, and individual. A comprehensive approach focusing on identifying acoustic features putatively involved in classification is required for the prediction of features necessary for discrimination in the real world. Here, we used several classification techniques, namely Discriminant Function Analyses (DFAs), Support Vector Machines (SVMs), and Artificial Neural Networks (ANNs), for sex-based classification of zebra finch (*Taeniopygia guttata*) distance calls using acoustic features measured from spectrograms. We found that all three methods (DFAs, SVM, and ANNs) correctly classified the calls to respective sex-based categories with high accuracy between 92 and 96%. Frequency modulation of ascending frequency, total duration, and end frequency of the distance call were the most predictive features underlying this classification in all of our models. Our results corroborate evidence of the importance of total call duration and frequency modulation in the classification of male and female distance calls. Moreover, we provide a methodological approach for bioacoustic classification problems using multiple statistical analyses.

I. INTRODUCTION

Acoustic communication is used throughout the animal kingdom in the contexts of mate attraction, territorial defense, raising alarm, and recognition of species, group, and individuals (Bradbury and Vehrencamp, 2011). Understanding the context in which animal vocalizations are used plays a key role in understanding biological function and evolution of animal communication, as well as the underlying mechanisms of vocal communication in the animals producing the vocalizations under study (Bradbury and Vehrencamp, 2011). Research in bioacoustics focuses primarily on the mechanisms of production, transmission, and reception of acoustic signals (Erbe, 2016; Hopp et al., 1998). One approach to bioacoustics research involves describing and then classifying animal vocalizations into categories. This approach helps to reduce naturally-occurring complexity among signal classes by forming categories of signals based on acoustic similarity (Garcia and Favaro, 2017). The categories can be vocal repertoires of different species (Ficken et al., 1978; Salmi et al., 2013), based on the sex of the vocalizer (Campbell et al., 2016), based on geographical locations (Hahn et al., 2013a; Tuncer, 2013), based on ecological habitats (Anderson et al., 2008; Gómez et al., 2018) or based on the individuals (Elie and Theunissen, 2016; Hahn et al., 2013b; Laiolo et al., 2000; (Montenegro et al., 2021, Průchová et al., 2017). The application of this approach varies widely from biological scales (Gentry et al., 2020) to wildlife management and conservation (Laiolo et al., 2008; Teixeira et al., 2019) to animal welfare (Manteuffel et al., 2004; Röttgen et al., 2020; Schön et al., 2004) to life history, and evolutionary biology (Warwick et al., 2015; Xu and Shaw, 2019).

Bioacoustics methods, especially vocalization classification, play an important role in investigations of cognitive processes such as perception, memory, and decision making (Shettleworth, 2009). Thorough description and classification of vocalizations are an integral part of understanding the mechanisms involved in biologically relevant processes like mate selection (Delgado, 2006; Hernandez et al., 2016; Vignal et al., 2008), predator interaction (Bee et al., 2016; Congdon et al., 2020), territoriality (Walcott et al., 2006), social interaction (Slocombe and Zuberbühler, 2005), and individual recognition (D'Amelio et al., 2017; Elie and Theunissen, 2018). Classification of vocalizations into specific classes as a tool of bioacoustic analyses dates to the early history of bioacoustics in the 1950s and 60s where scientists used sound spectrograms to describe the prominent features of vocalization types in domestic fowl and weaverbird (Collias, 1963; Collias and Joos, 1953). Since then, the field of bioacoustics has come a long way introducing new concepts, powerful analysis techniques (Herbst et al., 2013; Kershenbaum et al., 2016; Tallet et al., 2013; Wadewitz et al., 2015), and moving towards data-driven and automated classification (Bravo Sanchez et al., 2021; Brooker et al., 2020; Elie and Theunissen, 2016; Mcloughlin et al., 2019; Priyadarshani et al., 2018; Salamon et al., 2016).

A multitude of statistical methods, including automated methods, have been used for classification of vocalizations for biodiversity monitoring (Caycedo-Rosales et al., 2013; Priyadarshani et al., 2018), constructing vocal repertoires (Elie and Theunissen, 2016; Wadewitz et al., 2015), and classifying based on sex (Campbell et al., 2016), geography (Hahn et al., 2013a; Tuncer, 2013), and individuals (Elie and Theunissen, 2018; Průchová et al., 2017). These methods mainly include random forest, decision trees, Hidden Markov models, spectrogram cross-correlation, support vector machines (SVMs), and artificial neural networks (ANNs) (Knight et al., 2017). These automated methods are useful for classification, especially for large data sets, although human, and possibly non-human animal involvement, is required to verify the reliability and validity of such analyses. Thus, semi-automated methods with human involvement work best. Uncovering the acoustic features primarily responsible for the classification into types based on species, sex, or individual can reveal the locus of biologically significant stimulus control involved in animal communication signals. An integrated approach is required for classification, from identifying acoustic units to choosing methods of analyses for identifying features responsible for classification (Kershenbaum et al., 2016).

Zebra finches are flocking songbirds native to Australia that are highly sexually dimorphic in a number of important aspects. Only male zebra finches produce songs, though both sexes produce a variety of calls (Elie and Theunissen, 2016; Zann, 1996). Distance calls or "long calls" are the most characteristic, species-typical calls produced and are used in a variety of contexts, especially when birds are visually isolated from their mates or conspecifics (Zann, 1996). Distance calls are sexually dimorphic: males produce shorter, more acoustically complex calls and females produce longer, relatively unmodulated calls (Zann, 1996). The male distance call is composed of a downsweep frequency modulation with a fundamental frequency of approximately 600-1000 Hz (Figure 1). The female distance call is composed of a harmonic series of unmodulated frequencies with fundamental frequency of 350-550 Hz (Vicario et al., 2001; Zann, 1996). Zebra finches are capable of discriminating mates from others (Vignal et al., 2008) and of recognizing conspecifics (Vignal et al., 2004) using distance calls. The differences in the acoustic structure between male and female versions of these calls allow this discrimination (Vignal and Mathevon, 2011). Call duration, fundamental frequency, and rapid frequency modulation seem to play an important role in eliciting differential behavioural response to male and female distance calls (Vicario et al., 2001; Vignal and Mathevon, 2011).

In the past, zebra finch distance calls have been investigated in a number of different manners using a variety of bioacoustically-based classification approaches. A previous study, where the primary aim was to classify vocalizations into types (e.g., song, distance call etc.)

based on sexually-dimorphic acoustic features, quantified the potential acoustic features in the distance calls in comparison to other vocalization classes (Elie and Theunissen, 2016). This analysis showed that females produced longer, and lower pitched distance calls compared to male distance calls (Elie and Theunissen, 2016). A subsequent study (Mouterde et al., 2014) used DFA with spectral envelopes, temporal envelopes, and spectrogram features of distance calls, to classify distance calls based on the distance of the emitter of the calls (i.e., from 2m, 16m, 64m, 128m, and 256m) from the microphone. In the Mouterde et al. (2014) study, several density functions (mean, standard deviation, skewness, kurtosis, and entropy) of spectral and temporal envelopes and spectrogram principal component parameters were used successfully to classify distance calls at various propagation distances via the individual acoustic signature of the birds. However, the complex acoustic features of distance calls used in the bioacoustic analyses of the two studies just discussed are not straightforward to either measure or manipulate in an experimental context. The acoustic features described in these two studies are problematic for use in an experimental context as they are complex to either measure or manipulate by an experimenter.

In this study, we used three statistical methods (DFA, SVM and, ANN) to classify Zebra finch distance calls by sex of the emitter based on bioacoustic features, some of which are known to differ between sexes. We used 10 acoustic measurements in our analyses including both temporal and spectral measures. We predict similar classification performance in all three classification methods given the past successes using these methods for similar tasks (Mouterde et al., 2014, Elie and Theunissen, 2016). Furthermore, we predict that total call duration and frequency modulation will be the predominant features used to classify the calls, as these features are visually distinct in the spectrograms of these calls and previous studies (Elie and

Theunissen, 2016; Vignal and Mathevon, 2011) suggest that these are the key features facilitating sex-based call discrimination.

II. METHODS AND RESULTS

A. Recordings

In total, 83 zebra finch distance calls were obtained from the data sets of D'Amelio et al. (2017), Elie and Theunissen (2016), and from adult zebra finches recorded by members of the Phillmore lab at Dalhousie University, Halifax, NS Canada. The set consisted of 38 female and 45 male distance calls produced by 21 females and 26 males, with 1-2 vocalizations per individual. There were 12 male calls and 12 female calls obtained from 12 male individuals and 12 female individuals respectively from D'Amelio et al. (2017), 18 male calls and 20 female calls from 9 male individuals and 10 female individuals respectively from Elie and Theunissen (2016), and 15 male calls and 6 female calls from 8 male individuals and 4 female individuals respectively from the Phillmore lab. All the recorded calls were recorded in the laboratory with digital recorders and microphones having frequency response ranges from 200 Hz to 10,000 Hz. The calls from D'Amelio et al. (2017) and Elie and Theunissen (2016) were recorded at a sampling rate of 44,100 Hz. The calls obtained from Phillmore lab were recorded at a sampling rate of 48,000 Hz. For all sources, calls were recorded at a distance between 0 and 100 cm from the birds. Thus, these recordings provided us with a diverse dataset of high-quality recordings of distance calls. Because calls were recorded with different sampling rates, all distance calls were resampled using SIGNAL software version 5.16.11 (Beeman, 2017) at 44,100 Hz before further analyses.

B. Acoustic measurements

The following acoustic analyses and measurements were conducted in SIGNAL and performed by the first author (PS). For each sound file, 5 ms of silence was added to the beginning and end of the vocalization and tapered to remove transients; peak RMS amplitude was equalized to 1. Spectrograms were created with a Hanning window and 256 pts transform, frequency resolution of 172.3 Hz and 5.8 ms time resolution. Power spectra were produced with an FFT window of 16,384 points and 88 Hz smoothing for amplitude measurement. The following acoustic features were measured manually from the spectrograms of individual calls: (a) Total duration (TD), measured from the start to the end of the highest amplitude harmonic band, (b) Start frequency (SF), measured at the start of the first clearly visible and continuous harmonic band, in this case, the second frequency band in the spectrogram, (c) End frequency (EF), measured at the end of the first clearly visible and continuous harmonic band, in this case, the second frequency band in the spectrogram (d) Peak frequency (PF), measured at the highest frequency observed of the highest amplitude harmonic band, (e) Ascending duration (AD), measured from the start to the peak frequency of the highest amplitude harmonic band, (f) Descending duration (DD), measured from the peak to the end of the highest amplitude harmonic band, (g) Frequency modulation of ascending frequency (Fmasc; Peak frequency-Start frequency/Ascending duration), (h) Frequency modulation of descending frequency (Fmdsc; End frequency-Peak frequency/Descending duration), (i) Frequency at highest amplitude (Fmax), measured at the peak frequency of the highest amplitude harmonic band from the power spectra and, (j) Fundamental frequency (F0) (Campbell et al., 2016; Nowicki and Nelson, 1990). The fundamental frequency was measured in Praat 6.1.38 (Boersma and van Heuven, 2001;

Goldstein, 2021) Figure 1 shows the measured acoustic features from the spectrograms of male and female zebra finch distance calls.

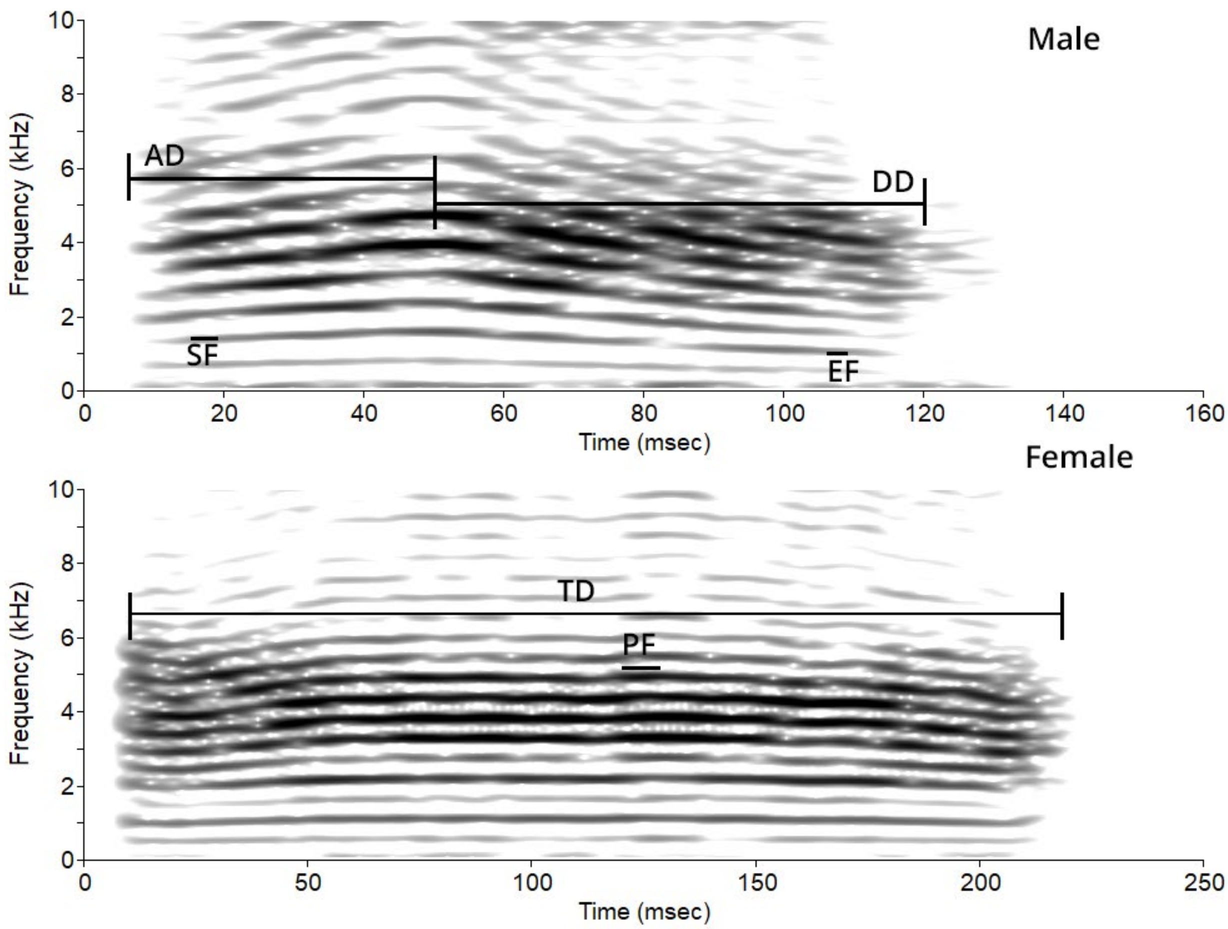

Figure 1. Measured acoustic features from the spectrogram showing Total duration (TD), Start frequency (SF), End frequency (EF), Peak frequency (PF), Ascending duration (AD), Descending duration (DD) with male distance call at top and female distance call at bottom. F0 was measured in Praat (not pictured here).

C. Statistical analyses

All analyses were conducted in R 3.6.2 (R Core Team, 2019). The linear discriminant analysis (LDA) and discriminant function analysis (DFA) were conducted using the *MASS*

(Venables and Ripley, 2002) and *klaR* (Weihs et al., 2005) packages, the SVM was conducted using the *e1071* package (Meyer et al., 2019), and the ANN was conducted using the *neuralnet* package (Günther and Fritsch, 2010). Mathews correlation coefficient (MCC) was calculated using *mltools* (Gorman, 2018). For LDA, standardized coefficients were obtained using canonical discriminant analysis from the *candisc* package (Friendly and Fox, 2021). The relative importance of variables or weights for SVM were calculated using the weight vectors (Meyer et al., 2019). The relative importance of input variables for ANN were calculated using the olden function of *NeuralNetTools* (Beck, 2018).

All measured acoustic features were scaled by z-standardization, using the *scale* function in R to account for and standardize across multiple units of measurement. This allowed us to compare between measures, even when those measures differed in units. The z-standardization of an individual acoustic feature involves subtracting the mean of the specific acoustic feature from the individual measurement and dividing by its standard deviation. We conducted correlation analyses to identify and omit redundant and highly correlated acoustic features. The Ascending duration (AD) and Descending duration (DD) were highly correlated with each other (*Pearson's* r =0.75, $p < 0.001$) and with Frequency modulation of ascending frequency, Fmasc (AD and Fmasc: *Pearson's* r = -0.85, $p < 0.001$) and Frequency modulation of descending frequency, Fmdsc (DD and Fmdsc: *Pearson's* r = - 0.83, $p < 0.001$). Thus, AD and DD were not included in further analyses. TABLE 1 shows correlation across the measured acoustic features. Below, we review how each technique operates, what kinds of results they return, what results we obtained using each technique and finally provide a comparison among the techniques.

TABLE I. Table showing Pearson's correlation coefficients across acoustic features. * represents significant correlation

	TD	SF	EF	PF	AD	DD	Fmasc	Fmdsc	Fmax	Fo
TD										
SF	-0.06									
EF	0.26*	0.48*								
PF	-0.01	-0.16	-0.22*							
AD	0.56*	0.04	0.43*	-0.26*						
DD	0.61*	0.03	0.30*	-0.1	0.75*					
Fmasc	-0.39	-0.19	-0.46*	0.52*	-0.85*	-0.54				
Fmdsc	0.43*	0.15	0.49*	-0.42*	0.72*	0.83*	-0.71*			
Fmax	0.05	-0.14	-0.08	0.51*	-0.03	0.02	0.27*	-0.2		
Fo	0.22	0.03	0.14	-0.17	0.42*	0.23*	-0.50*	0.31*	-0.16	

1. DFA, pDFA, and LDA.

Discriminant function analysis (DFA) is used for classification of exemplars into groups based on a linear combination of features which separate the groups. In bioacoustics analyses, DFA can be used to classify vocalizations into types (Jaiswara et al., 2013) or across individuals (Chen and Goldberg, 2020; Mundry and Sommer, 2007). For example, DFA has been used to classify vocalizations of mountain chickadees (*Poecile gambeli*) based on elevation gradient (Branch and Pravosudov, 2015, 2019). DFA has also been used to classify black-capped chickadee (*Poecile atricapillus*) vocalizations based on geography (British Columbia and Ontario; Hahn et al., 2013a) and sex (Campbell et al., 2016).

We used a stepwise DFA with the leave-one-out method of cross-validation for classifying distance calls based on sex. In this process, a single vocalization is withheld while the rest of the vocalizations are used to obtain the discriminant functions. The accuracy of the discrimination functions can then be obtained by comparing the predicted group, male or female, of the withheld vocalization to the original class of that vocalization (i.e., was the function able to classify a male call as male, and a female call as female.). This method was repeated until all the vocalizations were classified, thus giving us overall percent correct classification (Betz, 1987; Mundry and Sommer, 2007).

When multiple vocalizations from the same individuals are used for DFA, there is the possibility of pseudoreplication. Pseudoreplication occurs when non-independent data points from the same subject (e.g., multiple vocalizations from one individual) are analyzed as independent replicates (Mundry and Sommer, 2007). A permuted DFA (pDFA) can be used to account for pseudoreplication (Mundry and Sommer, 2007). With pDFA, we compared the percent correct classifications by DFA from the original distance call distribution to percent correct classifications obtained from null distributions. The null distributions of distance calls are constructed by randomly assigning individual calls as male or female. One thousand such null distributions were constructed, and percent correct classifications were obtained by leave-one out method of DFA as mentioned above. The proportion of times percent correct classification by pDFA were equal to or greater than correct classification by original DFA was obtained and was noted as p-value as described by (Mundry and Sommer, 2007).

The stepwise DFA accurately classified distance calls based on sex using all eight of the remaining measured acoustic features. In the forward stepwise DFA method for variable selection for classification, where each feature is entered individually, one by one (rather than all features entered all at once) with total duration (TD) as starting variable, Total duration (TD), End frequency (EF), Frequency modulation of ascending frequency (Fmasc), Frequency modulation of descending frequency (Fmdsc), and Frequency at highest amplitude (Fmax), were all used together for the sorting of distance calls into the respective sex category that produced the calls. The forward stepwise DFA classified the distance calls into the correct category with 96.3% accuracy. We then used a pDFA to test the validity of the stepwise DFA; conducting the pDFA involved constructing null distributions of distance calls, where sex identity of each call was randomized. The mean correct classification for all 1,000 null distributions was 50.2% ± 6.6 (mean±sd), meaning pDFA could only classify the null distributions with ~50% accuracy. None of the pDFAs produced correct percent classification greater than the stepwise DFA classification percentage, thus giving a p-value of 0 for the pDFA null model, indicating that the stepwise DFA accurately classified calls by sex of producer.

We also used a supervised linear discriminant analysis (LDA) with the hold out method of cross-validation to classify distance calls based on sex for a direct comparison with support vector machine (SVM) and artificial neural network analysis (ANN). In the hold out method of cross-validation, the data set is separated into two sets: training and testing. The function uses the training set to build a model to predict the output of the testing set. In supervised LDA, 75% of the vocalizations were chosen randomly for training and then the remaining 25% are used in a test to validate the accuracy of the same testing dataset. This procedure was repeated 1,000 times and mean percent accuracy was calculated (Engler et al., 2014; Ligout et al., 2016).

In the supervised LDA, all the eight acoustic features were used to calculate the discriminant functions and predict classification for testing datasets. This process was repeated and cross-validated 1,000 times to obtain the mean correct percent classification. Using all the eight features, the LDA classified distance calls with 96.3% ± 3.4 accuracy. The mean MCC for the LDA was 0.96 (range: 0.63-1.00) from 1,000 testing datasets which indicates high classification performance, meaning there was no significant effect of unbalanced datasets with unequal numbers of samples in the two groups. This can potentially pose a problem resulting in a larger dataset overestimating the classifier. Figure 2 shows the distribution of the individual distance calls according to the first discriminant function, LD1, male and female calls are well separated.

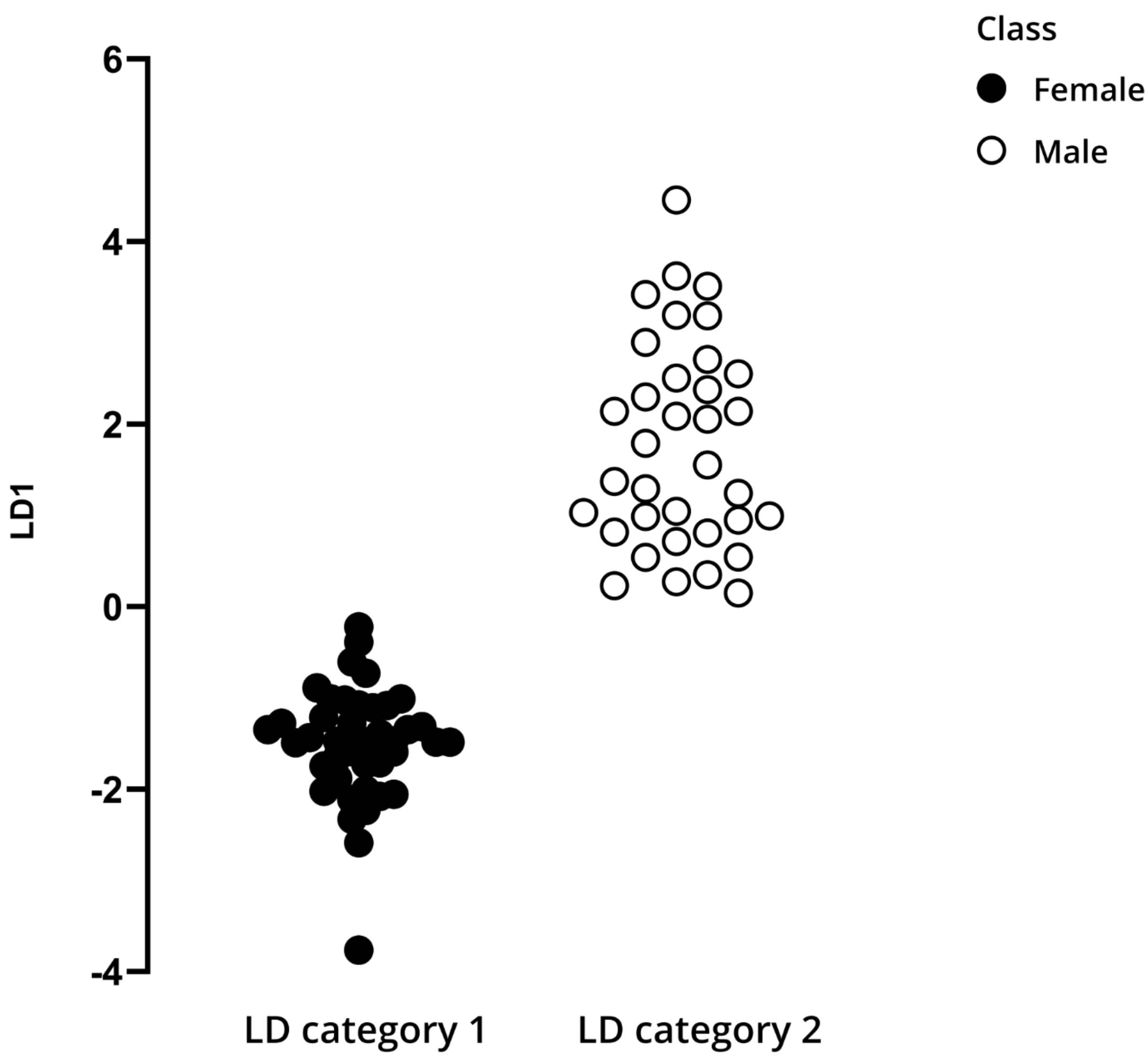

Figure 2. Distribution of the first discriminant function (LD1) for all male and female distance calls.

DFA is a robust and stable classification technique when the classes are well separated, thus giving accurate parameter estimates that separate the classes. However, when the dataset is limited with high dimensionality (has more features than samples), there is risk of over-fitting. This over-fitting might reduce the cross-validation performance of classifiers (James et al., 2013; Tachibana et al., 2014). Our overall dataset is not high dimensional (i.e., we have more samples, n=83 than features, n=8), but data are from various sources produced from different individuals. We used support vector machine (SVM) algorithms, as SVM avoids the problem of overfitting, a potential issue with DFA. SVM also helps reduce human effort involved in other classification methods, as SVM works with a relatively small instruction (training) dataset in comparison to other methods.

2. SVM

Support vector machines (SVMs) are supervised learning algorithms used for mainly two-group classification problems (Cortes and Vapnik, 1995). SVMs use all the eight measured acoustic features as input variables, similar to a DFA, and then build a prediction model. However, when data are not linear, there is a possibility for interaction among variables, which can happen when classifying using a DFA. SVMs solve this problem by using a kernel approach. Kernels are various functions (e.g., linear, polynomial, radial, and sigmoid) that can be applied to input data so that data are separated linearly in the feature space. Here, we used a linear kernel for both training and prediction. SVMs have been widely used for classification of songbirds to their species by their songs, for example, using song syllables of 7 bird species (Dufour et al., 2014) or using flight calls of 11 species of birds (Tung et al., 2003), and more recently, using additional automated methods such as hybrid model of deep convolutional neural networks and hidden Markov models for classification of birdsong using song notes and syllable elements

(Koumura and Okanoya, 2016; Tachibana et al., 2014). In this study, we used a linear SVM where classification boundaries are determined by maximizing margins between the nearest samples and boundary hyperplane for distance call classification. In this supervised semi-automated method, we randomly divided all 83 vocalizations using a 3:1 ratio to serve as training and testing datasets, respectively. The validity of the model based on the training dataset was measured against a testing dataset.

We cross validated performance SVM with the testing datasets. This process of cross validation was repeated 1,000 times with randomly chosen testing dataset for mean correct classification percentage. SVM classified distance calls with a mean of 94% ± 5.1 correct. The mean MCC for SVM was 0.88 (range: 0.42-1.00). Next, we built ANNs to compare ANN classification accuracy with the accuracy obtained from discriminant analyses and SVM.

3. ANN

Artificial neural networks (ANN) consist of connected input nodes and edges in multiple layers; acoustic features can be used as input to produce a predicted category as target output (Izenman, 2008). In bioacoustic analyses, neural networks have been used in the context of species classification using acoustic features, ranging from whole vocalizations to individual song and call notes, to sort vocalizations by species (Chou and Liu, 2009, 2009; Piczak, 2016) or to sort notes into note types (Dawson et al., 2006). ANNs for binary classification are very similar to SVM, apart from the training algorithms that are used for calculation of classification functions: ANNs use backpropagation whereas SVM uses hyperplane to make predictions. In backpropagation, the weights of a neural net are fine-tuned according to error rate or loss function of previous epochs or iterations in training while in hyperplane, observations are separated into two classes by a threshold hyperplane, calculated from linear combination of the dependent variables (Izenman, 2008; James et al., 2013). ANNs can account for potentially complex relationships among input features without compromising classification performance (Collobert and Bengio, 2004; Jakkula, 2011).

We used a supervised ANN, which used the eight measured acoustic features from distance calls as input, to classify the calls. We built an artificial neural network using the *neuralnet* package in R with the default logistic activation function (Günther and Fritsch, 2010). The neural network consisted of eight acoustic input features with a single hidden layer consisting of two neurons and one output unit to predict sex of the producer of the distance call. The input features were multiplied by a random set of weights prior to the training. The logistic activation function applied to the multiplied numbers and output as neurons in the hidden layer. The neurons in the hidden layer were again multiplied by a random set of weights, and the activation function was applied to these numbers to produce a single output. The prediction output (lies between 0 and 1) was compared with the true output. The loss or error was then calculated with a cross-entropy function to know how far off our prediction from true output (Izenman, 2008). We used resilient backpropagation algorithms to get the gradients for each weight from the initial random weights. During epochs of training, the error got smaller, and weights got optimized for best prediction of output (Günther and Fritsch, 2010). A schematic of the neural network is shown in Figure 3.

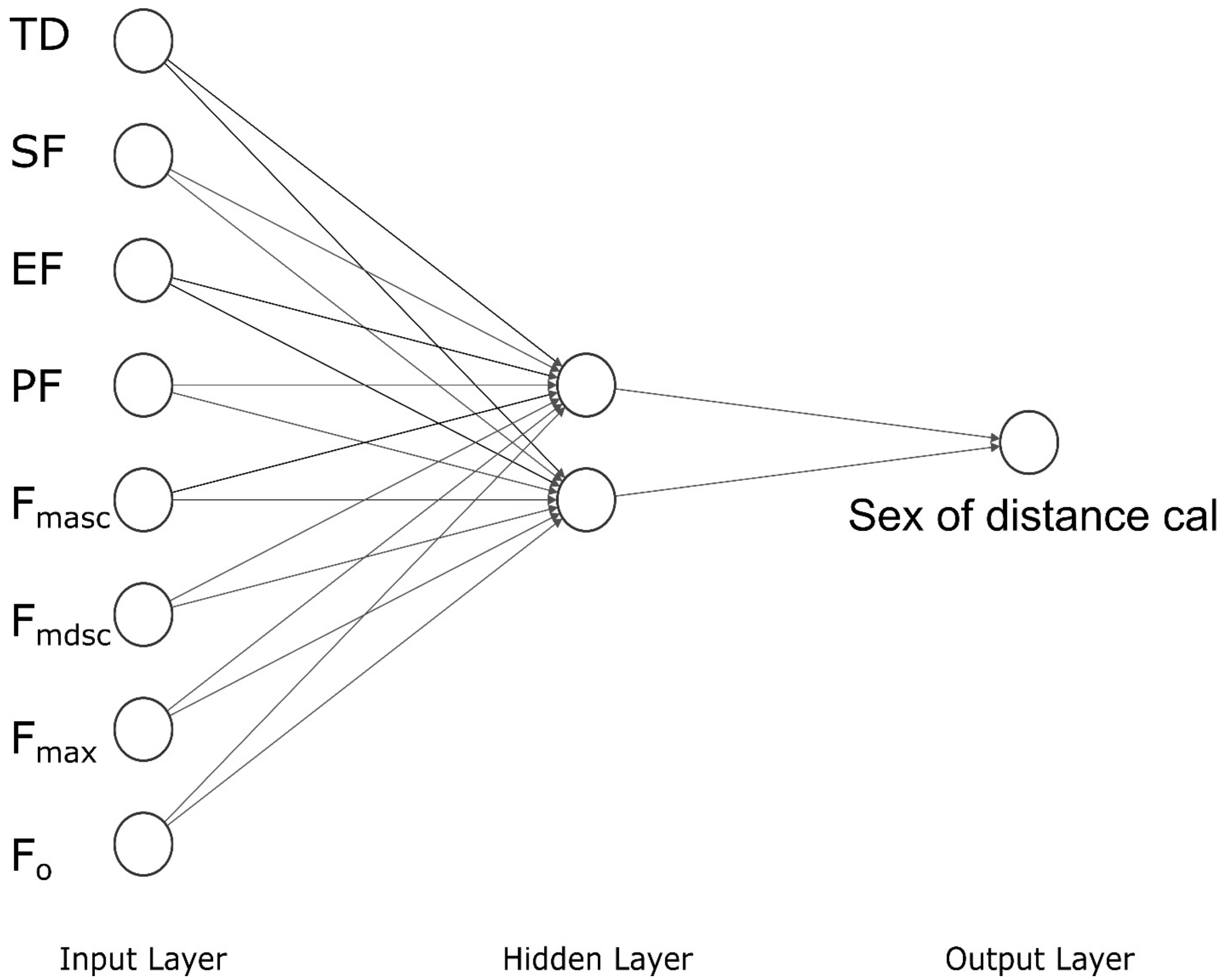

Figure 3. Schematics of the neural network showing acoustic features as input, two neurons in hidden layer and output layer where TD: Total duration, SF: Start frequency, EF: End frequency, PF: Peak frequency, Fmasc: Frequency modulation of ascending frequency, Fmdsc: Frequency modulation of descending frequency, Fmax: Frequency at highest amplitude, and Fo: Fundamental frequency.

Seventy-five percent of the total pool of vocalizations were chosen randomly to be used as a training set for supervised learning, while the remaining 25% of the vocalizations were withheld and used to validate the accuracy of the training model. This training and validation method was repeated 1,000 times. We trained the ANN until all absolute partial derivatives of the error function were smaller than 0.01 meaning we achieved asymptotic performance, a standard stopping point for confirming validity of ANNs (Günther and Fritsch, 2010).

The neural network classified the distance calls to the respective sex of producer with a mean accuracy of 92.5%±5.4 correct. The mean MCC for neural networks was 0.85 (range: 0.46-1.00), consistent with the MCC for both LDA and SVM.

4. Model comparison

The use of multiple methods of classification of distance calls will give us an overview of classification using a variety of methods while constructing a base for future classifications of similar problems. All methods used (DFA, LDA, SVM and ANN) classified calls into the correct sex of the produce with high accuracy (DFA and LDA: 96 %, SVM: 94 %, ANN: 92 %). For the stepwise DFA, pDFA validated the classification. To evaluate the relative classification performance for the rest of the methods (LDA, SVM and ANN), we calculated and compared MCC and the classification accuracy of each. MCC is a measure of quality of two-class classification used in various fields of research including songbird vocalization classification (Chicco and Jurman, 2020; Matthews, 1975; Wellock and Reeke, 2012). The MCC for all the methods (LDA: 96, SVM: 0.88, ANN: 0.85) were high and consistent with each other. Figure 4 shows a comparison of classification performance with accuracy and MCC. Further, we assessed the relative importance of specific acoustic features in classification across stepwise DFA, LDA, SVM and ANN methods. Comparing the relative importance across various methods will inform us as to whether the same acoustic features were used preferentially for each method for classification. Such methodological comparisons will further allow researchers to make informed decisions when selecting which methodological tools they will employ for their particular set of circumstances.

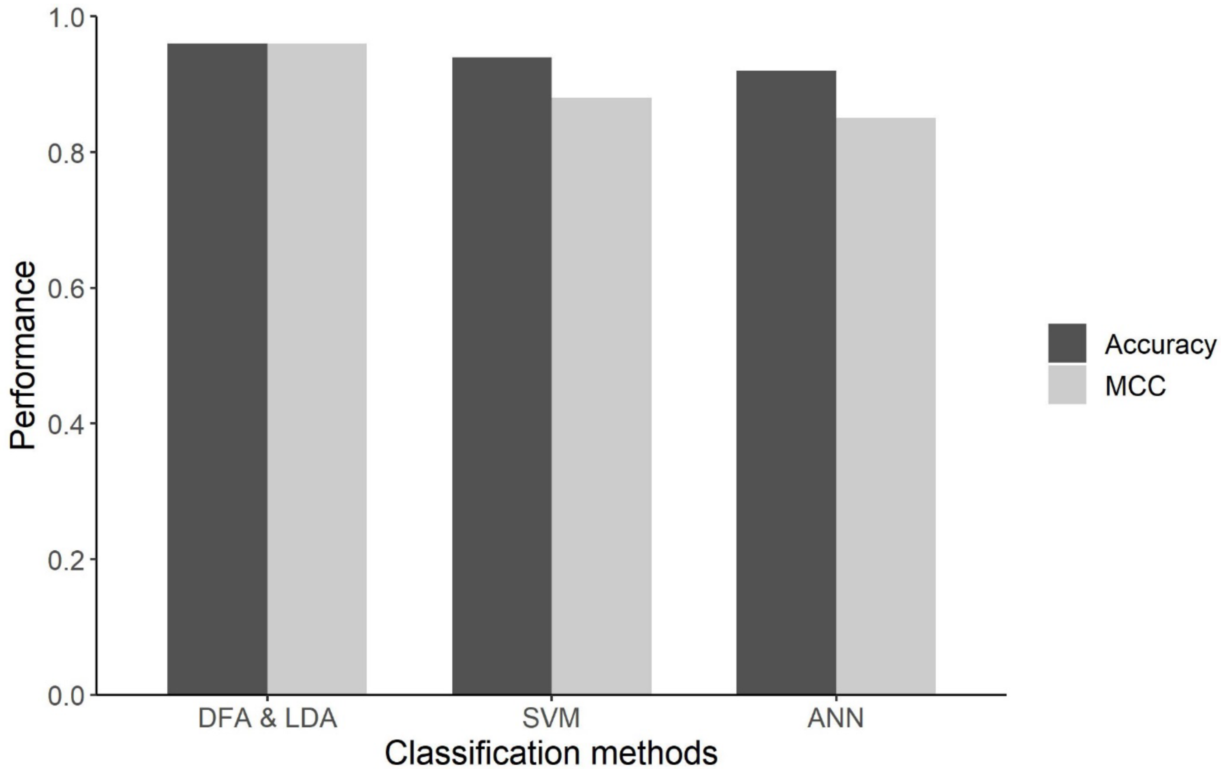

Figure 4. A bar graph showing classification accuracy (scaled to 1) and MCC (Matthews correlation coefficient) values for all classification methods; pDFA & LDA, SVM, and ANN. Higher value is better.

In all four methods (stepwise DFA, LDA, SVM, and ANN), Frequency modulation of ascending frequency (Fmasc), Total duration (TD), and End frequency (EF) were three top features used for classifying distance calls according to the sex of the producer (see TABLE II for full list). Although the methods used apply different algorithms for classification, they produced similar classification results with a similar relative importance for the input features. In sum, all methods tested successfully classified zebra finch calls by sex of the producer.

TABLE II. A table showing acoustic features used and their relative importance[a] (in descending order with relative proportion in the brackets) used for classification by Linear discriminant analysis (LDA), Support vector machines (SVM), and Artificial neural networks (ANN). For stepwise DFA, the DFA column shows features used for obtaining a classification accuracy of 96%.

Order of Importance	DFA	LDA	SVM	ANN
1	TD	Fmasc (0.43)	Fmasc (0.41)	Fmasc (0.42)
2	EF	TD (0.15)	EF (0.2)	EF (0.25)
3	Fmasc	EF (0.11)	TD (0.12)	TD (0.12)
4	Fmdsc	PF (0.1)	Fmdsc (0.08)	Fmdsc (0.07)
5	Fmax	Fo (0.08)	PF (0.07)	Fo (0.04)
6		Fmdsc (0.01)	SF (0.06)	PF (0.04)
7		SF (0.01)	Fo (0.04)	SF (0.04)
8		Fmax (0.004)	Fmax (0.0001)	Fmax (0.01)

[a] Refer to Statistical analysis section for calculation of relative importance of variables for each method.

III. DISCUSSION

All the methods (DFA, LDA, SVM, and ANN) were highly accurate at classifying distance calls into male and female; each had a classification accuracy greater than 92%. Three methods (LDA, SVM and ANN) had MCC values greater than 0.85, indicating highly correct predictions for both male and female calls independent of their potentially problematic unequal sample size in the dataset. The results from DFA, LDA, SVM, and ANN consistently and accurately classified male and female distance calls. Both the leave-one-out method and holdout method of cross-validation produced similarly excellent classification performance. This suggests that there are acoustic features that differ between male and female distance calls such that they can be used to effectively classify them with all four of these methods. Frequency modulation of ascending frequency (Fmasc), end frequency (EF), and total call duration (TD) of the distance call were the top ranked acoustic features used by stepwise DFA, LDA, SVM, and ANN. The LDA, SVM, and ANN all ranked Fmasc as the most important acoustic feature. SVM and ANN ranked EF and TD as the second and third most important features whereas LDA ranked TD and EF as second and third most important features. The order change may be due to different algorithms used for classification and for relative importance nevertheless frequency modulation was the most crucial feature used for classification. The stepwise DFA approach is useful and efficient for investigating and pruning variables when there are a large number of input variables involved; all variables can be entered in one step and the DFA outputs the variables used in the classification. SVM, on the other hand, works best for binary classifications with the use of maximum margin linear classifiers and for high dimension data, relatively large datasets, with the help of various available kernel functions. ANN is useful with multi-class classification with large datasets. Ideally, we recommend the use of a combination of these methods to account for stochasticity of real-world data. Pragmatically, if we were to choose one method, SVM would be our recommendation for the current question of sex-based call classification, due to its simplicity and ease of use for binary classification problems. Our study adds to the literature of methodological comparisons of vocalization classification (Bat echolocation: Armitage and Ober, 2010; Mouse ultrasonic: Ivanenko et al., 2020).

The distance calls used here were from several sources: birds were from the colonies in the USA, Germany, and Canada (D'Amelio et al., 2017; Elie and Theunissen, 2016). Thus, the study involved calls from a diverse sampling space which extends the external validity of the study. The acoustic features we measured and entered into the algorithms resulted in successful classification by DFA, SVM, and ANNs; all approaches were able to classify the distance calls with high accuracy of over 92%. It would be ideal to test vocalizations from other captive colonies and to wild birds to determine whether accuracy remains high with vocalizations from other groups of finches, including non-domesticated birds. Distance calls are sexually dimorphic, making the classification task relatively easy. It would be interesting to expand this study to test the performance of the classification methods with other zebra finch calls such as stacks and tets which contain individuals' sex identity.

Because all measurements for acoustic features were collected manually, there is a degree of subjectivity in the data that could have resulted in some potential for increased variability in the measurements collected. In the future, one refinement might be using an automated process to measure acoustic features with more consistency and less chance of bias. However, even automated or semi-automated measurement techniques require some level of human involvement, either for establishing the method or verifying the accuracy of the chosen method (Priyadarshani et al., 2018). We did not use additional acoustic measures that were difficult to obtain (e.g., Mel-Frequency Cepstral Coefficient or moments of spectral density functions like skewness, kurtosis, entropy etc.) for classification. Thus, variables used for the classification may have been oversimplified and as a result some important acoustic features potentially used by zebra finches for discrimination may not have been detected. It would be interesting to compare classification performance with the methods discussed here when using predefined acoustic features (e.g., intensity measures, pitch, frequency measures) vs a complete representation (e.g., Modulation power spectrum, full spectrogram, Mel frequency cepstral coefficients) of the acoustic stimuli (Elie and Theunissen, 2016). In future, larger samples from many individuals may be helpful to alleviate this issue by being able to assess interrater reliability of acoustic measurements in cases where more than one individual measured calls.

Distance calls contain information about individual identity of the caller. Studies could compare the classification performance of these methods for the classification of distance call based on individual identity which would require large number of calls from each individual. The features used here for distance call classification based on sex of caller can be used as a starting point to design future experiments to validate the acoustic measures used in the present study, such as an operant conditioning study to directly test the birds' ability to discriminate the manipulated on those features measured here. Such an operant study would add to the literature combining detailed bioacoustics analysis with perceptual studies by assisting in identifying and then manipulating simple spectrogram features to create experimental stimuli. That is to say; studies could test if only duration or frequency cues from the calls are discriminable. We would expect the acoustic features identified here would be relatively easily discriminable based on previous research works (Lohr et al., 2006, Lohr et al., 2003, Prior et al., 2018). Apart from distance calls, other zebra finch calls like stacks and tets also contain sex identity of the caller.

We predict that a single acoustic feature from distance call will not be enough to convey information about sex. Our study adds further evidence about the importance of these acoustic features and methodologies for classification and these methodologies can be used for classification of other call types.

The relative importance of variables in classification models provide information about what acoustic features animals may attend to preferentially when listening to and making decisions about responding to conspecific vocalizations. Previous studies focusing classification of vocalizations have primarily used Canonical loadings from DFA (Khan and Qureshi, 2017; Tooze et al., 1990), Gini index, or mean decrease accuracy for Random Forest algorithm (Armitage and Ober, 2010; Elie and Theunissen, 2016; Henderson et al., 2011; Robakis et al., 2018; Valletta et al., 2017) to determine relative importance of input variables due to their successful use in various contexts and ease of implementation in statistical software. We used similar measures for variable importance and expanded with the connection weight algorithm (Olden and Jackson, 2002) for variable importance in ANN. Future studies could use the above variables of importance and possibly improve with other methods for assessing the relative importance of input variables for ANN (Ibrahim, 2013).

In conclusion, we show that discriminant functions, support vectors, and neural networks were consistent with each other in accurately classifying zebra finch distance calls by sex of caller. Zebra finch distance calls can be accurately classified by sex using primarily three acoustic features: total duration, end frequency and frequency modulation ascending frequency. Highly similar patterns of acoustic feature rankings were observed for classification for all the methods. We believe our framework used in this bioacoustic analysis and subsequent classification of distance calls can be used as a starting point for researchers wanting to conduct similar bioacoustics studies in the future.

ACKNOWLEDGEMENTS

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (NSERC RGPIN 249884) and Discovery Accelerator Supplement (NSERC RGPAS 412311), an Alberta Ingenuity Fund (AIF) New Faculty Grant, and the Canada Foundation for Innovation (CFI) New Opportunities Fund (NOF).

REFERENCES

Anderson, J. T., Van Holliday, D., Kloser, R., Reid, D. G., and Simard, Y. (**2008**). "Acoustic seabed classification: current practice and future directions," ICES J. Mar. Sci. **65**(6), 1004–1011. doi:10.1093/icesjms/fsn061

Armitage, D. W., and Ober, H. K. (**2010**). "A comparison of supervised learning techniques in the classification of bat echolocation calls," Ecol. Inform. **5**, 465–473. doi:10.1016/j.ecoinf.2010.08.001

Beck, M. W. (**2018**). "NeuralNetTools: Visualization and analysis tools for neural networks," J. Stat. Softw. **85**(11), 1–20. doi:10.18637/jss.v085.i11

Bee, M. A., Reichert, M. S., and Tumulty, J. (**2016**). "Chapter Four - Assessment and Recognition of Rivals in Anuran Contests," In M. Naguib, J. C. Mitani, L. W. Simmons, L. Barrett, S. Healy, and M. Zuk (Eds.), Adv. Stud. Behav. Academic Press, Vol. 48, pp. 161–249. doi:10.1016/bs.asb.2016.01.001

Betz, N. E. (**1987**). "Use of discriminant analysis in counseling psychology research," J. Couns. Psychol., **34**(4), 393–403. doi:10.1037/0022-0167.34.4.393

Boersma, P., and van Heuven, V. (**2001**). “Speak and unSpeak with PRAAT,” **5**, 7.

Bradbury, J. W., and Vehrencamp, S. L. (**2011**). *Principles of animal communication*, Sinauer Associates, Sunderland, Mass, 2nd ed., 697 pages.

Branch, C. L., and Pravosudov, V. V. (**2015**). “Mountain chickadees from different elevations sing different songs: acoustic adaptation, temporal drift or signal of local adaptation?” Roy. Soc. Open Sci. **2**(4), 150019. doi:10.1098/rsos.150019

Branch, C. L., and Pravosudov, V. V. (**2019**). “Variation in song structure along an elevation gradient in a resident songbird,” Behav. Ecol. Sociobiol. **74**(1), 9. doi:10.1007/s00265-019-2786-5

Bravo Sanchez, F. J., Hossain, M. R., English, N. B., and Moore, S. T. (**2021**). “Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture,” Sci. Rep. **11**(1), 15733. doi:10.1038/s41598-021-95076-6

Brooker, S. A., Stephens, P. A., Whittingham, M. J., and Willis, S. G. (**2020**). “Automated detection and classification of birdsong: An ensemble approach,” Ecol. Indic. **117**, 106609. doi:10.1016/j.ecolind.2020.106609

Campbell, K. A., Hahn, A. H., Congdon, J. V., and Sturdy, C. B. (**2016**). “An investigation of sex differences in acoustic features in black-capped chickadee (*Poecile atricapillus*) *chick-a-dee* calls,” J. Acoust. Soc. Am. **140**(3), 1598–1608. doi:10.1121/1.4962281

Caycedo-Rosales, P. C., Ruiz-Muñoz, J. F., and Orozco-Alzate, M. (**2013**). “Automated recognition of bioacoustic signals: A review of methods and applications,” Ingeniería y Ciencia, **9**(18), 171–195.

Chen, R., and Goldberg, J. H. (**2020**). “Actor-critic reinforcement learning in the songbird,” Curr. Opin. Neurobio. **65**, 1–9. doi:10.1016/j.conb.2020.08.005

Chicco, D., and Jurman, G. (**2020**). “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” BMC Genomics, **21**, 6. doi:10.1186/s12864-019-6413-7

Chou, C.-H., and Liu, P.-H. (**2009**). “Bird species recognition by wavelet transformation of a section of birdsong,” 2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, 189–193. Presented at the 2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing. doi:10.1109/UIC-ATC.2009.85

Collias, N. E. (**1963**). “A spectrographic analysis of the vocal repertoire of the African village Weaverbird,” The Condor. **65**(6), 517–527. doi:10.2307/1365510

Collias, N., and Joos, M. (**1953**). “The spectrographic analysis of sound signals of the domestic fowl,” Behaviour, **5**, 175–188. doi:10.1163/156853953X00104

Collobert, R., and Bengio, S. (**2004**). “Links between perceptrons, MLPs and SVMs,” Twenty-first international conference on Machine learning - ICML ’04, ACM Press, Banff, Alberta, Canada, 23. Presented at the Twenty-first international conference. doi:10.1145/1015330.1015415

Congdon, J. V., Hahn, A. H., Campbell, K. A., Scully, E. N., Yip, D. A., Bayne, E. M., and Sturdy, C. B. (**2020**). “Acoustic discrimination of predators by black-capped chickadees (*Poecile atricapillus*),” Anim. Cogn. **23**(3), 595–611. doi:10.1007/s10071-020-01364-5

Cortes, C., and Vapnik, V. (**1995**). “Support-vector networks,” Mach. Learn. **20**(3), 273–297. doi:10.1007/BF00994018

D’Amelio, P. B., Klumb, M., Adreani, M. N., Gahr, M. L., and ter Maat, A. (**2017**). “Individual recognition of opposite sex vocalizations in the zebra finch,” Sci. Rep. **7**(1), 5579. doi:10.1038/s41598-017-05982-x

Dawson, M. R. W., Charrier, I., and Sturdy, C. B. (**2006**). “Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types,” J. Acoust. Soc. Am., **119**(5), 3161–3172. doi:10.1121/1.2189028

Delgado, R. A. (**2006**). “Sexual selection in the loud calls of male primates: Signal content and function,” Int. J. Primatol. **27**(1), 5–25. doi:10.1007/s10764-005-9001-4

Doupe, A. J., and Kuhl, P. K. (**1999**). “Birdsong and human speech: Common themes and mechanisms,” Annu. Rev. Neurosci. **22**(1), 567–631. doi:10.1146/annurev.neuro.22.1.567

Dufour, O., Artieres, T., Glotin, H., and Giraudet, P. (**2014**). “Clusterized Mel Filter Cepstral Coefficients and support vector machines for bird song identification,” Soundscape Semiotics - Localization and Categorization. Retrieved from https://www.intechopen.com/chapters/45589

Elie, J. E., and Theunissen, F. E. (**2016**). “The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals,” Anim. Cogn. **19**(2), 285–315. doi:10.1007/s10071-015-0933-6

Elie, J. E., and Theunissen, F. E. (**2018**). “Zebra finches identify individuals using vocal signatures unique to each call type,” Nat. Commun. **9**(1), 4026. doi:10.1038/s41467-018-06394-9

Engler, J. O., Rödder, D., Stiels, D., and Förschler, M. I. (**2014**). “Suitable, reachable but not colonised: seasonal niche duality in an endemic mountainous songbird,” J. Ornithol. **155**(3), 657–669. doi:10.1007/s10336-014-1049-5

Erbe, C. (**2016**). “What is animal bioacoustics?” J. Acoust. Soc. Am. **139**(4), 2004–2004. doi:10.1121/1.4949882

Fedurek, P., and Slocombe, K. E. (**2011**). “Primate vocal communication: A useful tool for understanding human speech and language evolution?” Hum. Biol. **83**, 153–173. doi:10.3378/027.083.0202

Ficken, M. S., Ficken, R. W., and Witkin, S. R. (**1978**). “Vocal repertoire of the back-capped chickadee,” The Auk, **95**(1), 34–48. doi:10.2307/4085493

Friendly, M., and Fox, J. (**2021**). *candisc: Visualizing generalized canonical discriminant and canonical correlation analysis* (manual). Retrieved from https://CRAN.R-project.org/package=candisc

Garcia, M., and Favaro, L. (**2017**). “Animal vocal communication: function, structures, and production mechanisms,” Curr. Zool. **63**(4), 417–419. doi:10.1093/cz/zox040

Gentry, K. E., Lewis, R. N., Glanz, H., Simões, P. I., Nyári, Á. S., and Reichert, M. S. (**2020**). “Bioacoustics in cognitive research: Applications, considerations, and recommendations,” WIREs Cogn. Sci. doi: 10.1002/wcs.1538. doi:10.1002/wcs.1538

Gobes, S. M. H., ter Haar, S. M., Vignal, C., Vergne, A. L., Mathevon, N., and Bolhuis, J. J. (**2009**). “Differential responsiveness in brain and behavior to sexually dimorphic long calls in male and female zebra finches,” J. Comp. Neurol. **516**(4), 312–320. doi:10.1002/cne.22113

Goldstein, L. (**2021**). *Measuring F0 Contours using Praat*, Available: https://sail.usc.edu/~lgoldste/General_Phonetics/Tone/Measurement/measurement.html, (date last viewed: 06-Apr-21). Retrieved April 6, 2021, from https://sail.usc.edu/~lgoldste/General_Phonetics/Tone/Measurement/measurement.html

Gómez, W. E., Isaza, C. V., and Daza, J. M. (**2018**). "Identifying disturbed habitats: A new method from acoustic indices," Ecol. Inform. **45**, 16–25. doi:10.1016/j.ecoinf.2018.03.001

Gorman, B. (**2018**). *mltools: Machine learning tools* (manual). Retrieved from https://CRAN.R-project.org/package=mltools

Günther, F., and Fritsch, S. (**2010**). "neuralnet: Training of Neural Networks," The R Journal, **2**, 30. doi:10.32614/RJ-2010-006

Hahn, A. H., Guillette, L. M., Hoeschele, M., Mennill, D. J., Otter, K. A., Grava, T., Ratcliffe, L. M., et al. (**2013a**). "Dominance and geographic information contained within black-capped chickadee (*Poecile atricapillus*) song," Behav. **150**(13), 1601–1622. doi:10.1163/1568539X-00003111

Hahn, A. H., Krysler, A., and Sturdy, C. B. (**2013b**). "Female song in black-capped chickadees (*Poecile atricapillus*): Acoustic song features that contain individual identity information and sex differences," Behav. Process. **98**, 98–105. doi:10.1016/j.beproc.2013.05.006

Henderson, E. E., Hildebrand, J. A., and Smith, M. H. (**2011**). "Classification of behavior using vocalizations of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*)," J. Acoust. Soc. Am. **130**, 557–567. doi:10.1121/1.3592213

Herbst, C. T., Herzel, H., Švec, J. G., Wyman, M. T., and Fitch, W. T. (**2013**). "Visualization of system dynamics using phasegrams," J. Roy. Soc. Interface. **10**(85), 20130288. doi:10.1098/rsif.2013.0288

Hernandez, A. M., Perez, E. C., Mulard, H., Mathevon, N., and Vignal, C. (**2016**). "Mate call as reward: Acoustic communication signals can acquire positive reinforcing values during adulthood in female zebra finches (*Taeniopygia guttata*).," J. Comp. Psychol. **130**(1), 36–43. doi:10.1037/a0040027

Hopp, S. L., Owren, M. J., and Evans, C. S. (Eds.) (**1998**). *Animal Acoustic Communication: Sound Analysis and Research Methods*, Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-76220-8

Ibrahim, O. M. (**2013**). "A comparison of methods for assessing the relative importance of input variables in artificial neural networks.," J. Appl. Sci. Res. **9**(11), 5692-5700

Ivanenko, A., Watkins, P., Gerven, M. A. J. van, Hammerschmidt, K., and Englitz, B. (**2020**). "Classifying sex and strain from mouse ultrasonic vocalizations using deep learning," PLOS Comput. Biol. **16**(6), e1007918. doi:10.1371/journal.pcbi.1007918

Izenman, A. J. (**2008**). *Modern Multivariate Statistical Techniques*, Springer Texts in Statistics, Springer New York, New York, NY. doi:10.1007/978-0-387-78189-1

Jaiswara, R., Nandi, D., and Balakrishnan, R. (**2013**). "Examining the effectiveness of discriminant function analysis and cluster analysis in species identification of male field crickets based on their calling songs," PLOS ONE, **8**(9), e75930. doi:10.1371/journal.pone.0075930

Jakkula, V. (**2011**). "Tutorial on support vector machine (SVM).," School of EECS, Washington State University, **37**.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (**2013**). *An Introduction to Statistical Learning*, Springer Texts in Statistics, Springer New York, New York, NY, Vol. 103. doi:10.1007/978-1-4614-7138-7

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., et al. (**2016**). "Acoustic sequences in non-human animals: a tutorial review and prospectus: Acoustic sequences in animals," Biol. Rev. **91**(1), 13–52. doi:10.1111/brv.12160

Khan, A. A., and Qureshi, I. Z. (**2017**). "Vocalizations of adult male Asian koels (*Eudynamys scolopacea*) in the breeding season," PLOS ONE, **12**(10), e0186604. doi:10.1371/journal.pone.0186604

Knight, E., Hannah, K., Foley, G., Scott, C., Brigham, R., and Bayne, E. (**2017**). "Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs," Avian Conserv. Ecol. doi: 10.5751/ACE-01114-120214.

Koumura, T., and Okanoya, K. (**2016**). "Automatic recognition of element classes and boundaries in the birdsong with variable sequences," PLOS ONE, **11**(7), e0159188. doi:10.1371/journal.pone.0159188

Laiolo, P., Claudia, P., and Antonio, R. (**2000**). "A study of Choughs' vocal repertoire: variability related to individuals, sexes and ages," J. Ornithol. **141**(2), 168. doi:10.1046/j.1439-0361.2000.00074.x

Laiolo, P., Vögeli, M., Serrano, D., and Tella, J. L. (**2008**). "Song diversity predicts the viability of fragmented bird populations," PLOS ONE, **3**(3), e1822. doi:10.1371/journal.pone.0001822

Ligout, S., Dentressangle, F., Mathevon, N., and Vignal, C. (**2016**). "Not for parents only: Begging calls allow nest-mate discrimination in juvenile zebra finches," Ethology, **122**(3), 193–206. doi:https://doi.org/10.1111/eth.12450

Lohr, B., Dooling, R. J., and Bartone, S. (**2006**). “The discrimination of temporal fine structure in call-like harmonic sounds by birds.,” J. Comp. Psychol. **120**, 239–251. doi:10.1037/0735-7036.120.3.239

Lohr, B., Wright, T. F., and Dooling, R. J. (**2003**). “Detection and discrimination of natural calls in masking noise by birds: estimating the active space of a signal,” Anim. Behav. **65**, 763–777. doi:10.1006/anbe.2003.2093

Manteuffel, G., Puppe, B., and Schön, P. C. (**2004**). “Vocalization of farm animals as a measure of welfare,” Appl. Anim. Behav. Sci. **88**(1), 163–182. doi:10.1016/j.applanim.2004.02.012

Matthews, B. W. (**1975**). “Comparison of the predicted and observed secondary structure of T4 phage lysozyme,” Biochim. et Biophys. Acta (BBA) - Protein Structure, **405**(2), 442–451. doi:10.1016/0005-2795(75)90109-9

Mcloughlin, M. P., Stewart, R., and McElligott, A. G. (**2019**). “Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring,” J. R. Soc. Interface. **16**(155), 20190225. doi:10.1098/rsif.2019.0225

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (**2019**). *e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), TU wien* (manual), Retrieved from https://CRAN.R-project.org/package=e1071

Montenegro, C., Sahu, P. K., and Sturdy, C. B. (**2021**). “Individual acoustic differences in female black-capped chickadee (*Poecile atricapillus*) fee-bee songs,” J. Acoust. Soc. Am. **150**, 3038–3046. doi:10.1121/10.0006532

Mouterde, S. C., Theunissen, F. E., Elie, J. E., Vignal, C., and Mathevon, N. (**2014**). “Acoustic communication and sound degradation: How do the individual signatures of male and female zebra finch calls transmit over distance?" (M. J. Coleman, Ed.) PLoS ONE, **9**(7), e102842. doi:10.1371/journal.pone.0102842

Mundry, R., and Sommer, C. (**2007**). "Discriminant function analysis with nonindependent data: Consequences and an alternative," Anim. Behav. **74**(4), 965–976. doi:10.1016/j.anbehav.2006.12.028

Nowicki, S., and Nelson, D. A. (**1990**). "Defining natural categories in acoustic signals: Comparison of three methods applied to '*Chick-a-dee'* call notes," Ethology, **86**(2), 89–101. https://doi.org/10.1111/j.1439-0310.1990.tb00421.x

Piczak, K. J. (**2016**). "Recognizing bird species in audio recordings using deep convolutional neural networks," BirdCLEF, CLEF 2016 Conference, Évora, Portugal, 534-543

Prior, N. H., Smith, E., Lawson, S., Ball, G. F., and Dooling, R. J. (**2018**). "Acoustic fine structure may encode biologically relevant information for zebra finches," Sci. Rep. **8**, 6212. doi:10.1038/s41598-018-24307-0

Priyadarshani, N., Marsland, S., and Castro, I. (**2018**). "Automated birdsong recognition in complex acoustic environments: a review," J. Avian Biol. **49**(5), jav-01447. doi:10.1111/jav.01447

Průchová, A., Jaška, P., and Linhart, P. (**2017**). "Cues to individual identity in songs of songbirds: testing general song characteristics in Chiffchaffs *Phylloscopus collybita*," J. Ornithol. **158**(4), 911–924. doi:10.1007/s10336-017-1455-6

R Core Team (**2019**). *R: A language and environment for statistical computing* (manual), Vienna, Austria. Retrieved from https://www.R-project.org/

Robakis, E., Watsa, M., and Erkenswick, G. (**2018**). “Classification of producer characteristics in primate long calls using neural networks,” J. Acoust. Soc. Am. **144**, 344–353. doi:10.1121/1.5046526

Röttgen, V., Schön, P. C., Becker, F., Tuchscherer, A., Wrenzycki, C., Düpjan, S., and Puppe, B. (**2020**). “Automatic recording of individual oestrus vocalisation in group-housed dairy cattle: Development of a cattle call monitor,” Animal, **14**(1), 198–205. doi:10.1017/S1751731119001733

Salamon, J., Bello, J. P., Farnsworth, A., Robbins, M., Keen, S., Klinck, H., and Kelling, S. (**2016**). “Towards the automatic classification of avian flight calls for bioacoustic monitoring,” PLOS ONE, **11**(11), e0166866. doi:10.1371/journal.pone.0166866

Salmi, R., Hammerschmidt, K., and Doran-Sheehy, D. M. (**2013**). “Western gorilla vocal repertoire and contextual use of vocalizations,” Ethology. **119**(10), 831–847. doi:10.1111/eth.12122

Schön, P., Puppe, B., and Manteuffel, G. (**2004**). “Automated recording of stress vocalisations as a tool to document impaired welfare in pigs,” Anim. Welfare. **13**(2), 105-110

Shettleworth, S. J. (**2009**). *Cognition, Evolution, and Behavior*, Oxford University Press, Oxford, New York, Second Edition., 720 pages.

Slocombe, K. E., and Zuberbühler, K. (**2005**). “Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) Vary as a Function of Social Role.,” J. Comp. Psychol. **119**(1), 67–77. doi:10.1037/0735-7036.119.1.67

Tachibana, R. O., Oosugi, N., and Okanoya, K. (**2014**). “Semi-automatic classification of birdsong elements using a linear support vector machine,” (J. J. Bolhuis, Ed.) PLoS ONE, **9**(3), e92584. doi:10.1371/journal.pone.0092584

Tallet, C., Linhart, P., Policht, R., Hammerschmidt, K., Šimeček, P., Kratinova, P., and Špinka, M. (**2013**). "Encoding of situations in the vocal repertoire of piglets (Sus scrofa): A comparison of discrete and graded classifications," PLOS ONE, **8**(8), e71841. doi:10.1371/journal.pone.0071841

Teixeira, D., Maron, M., and van Rensburg, B. J. (**2019**). "Bioacoustic monitoring of animal vocal behavior for conservation," Conserv. Sci. Prac. **1**(8), e72. doi:10.1111/csp2.72

Tooze, Z. J., Harrington, F. H., and Fentress, J. C. (**1990**). "Individually distinct vocalizations in timber wolves, *Canis lupus*," Anim. Behav., **40**, 723–730. doi:10.1016/S0003-3472(05)80701-8

Tuncer, T. (**2013**). *Geographic variation of song within and between subspecies of the Mangrove Warbler (Setophaga petechia castaneiceps and S. p. rhizophorae)* (MS), West Virginia University Libraries. Retrieved from https://researchrepository.wvu.edu/etd/574

Tung, H.-Y., Huang, D.-A., Xie, X.-F., Zhou, Y., and Russino, J. (**2003**). "Identification of Songbird Species in Field Recordings.," Carnegie Mellon University, ML-10701 Project.

Valletta, J. J., Torney, C., Kings, M., Thornton, A., and Madden, J. (**2017**). "Applications of machine learning in animal behaviour studies," Anim. Behav., **124**, 203–220. doi:10.1016/j.anbehav.2016.12.005

Venables, W. N., and Ripley, B. D. (**2002**). *Modern applied statistics with s*, Springer, New York, 4th ed. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vicario, D. S., Naqvi, N. H., and Raksin, J. N. (**2001**). "Sex differences in discrimination of vocal communication signals in a songbird," Anim. Behav. **61**(4), 805–817. doi:10.1006/anbe.2000.1651

Vignal, C., and Mathevon, N. (**2011**). “Effect of acoustic cue modifications on evoked vocal response to calls in zebra finches (*Taeniopygia guttata*).,” J. Comp. Psychol. **125**(2), 150–161. doi:10.1037/a0020865

Vignal, C., Mathevon, N., and Mottin, S. (**2004**). “Audience drives male songbird response to partner’s voice,” Nature, **430**(6998), 448–451. doi:10.1038/nature02645

Vignal, C., Mathevon, N., and Mottin, S. (**2008**). “Mate recognition by female Zebra finch: Analysis of individuality in male call and first investigations on female decoding process,” Behav. Process. **77**(2), 191–198. doi:10.1016/j.beproc.2007.09.003

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., and Fischer, J. (**2015**). “Characterizing vocal repertoires--Hard vs. Soft classification approaches,” PLoS One, **10**(4), e0125785. doi:10.1371/journal.pone.0125785

Walcott, C., Mager, J. N., and Piper, W. (**2006**). “Changing territories, changing tunes: male loons, Gavia immer, change their vocalizations when they change territories,” Anim. Behav. **71**(3), 673–683. doi:10.1016/j.anbehav.2005.07.011

Warwick, A. R., Travis, J., and Lemmon, E. M. (**2015**). “Geographic variation in the Pine Barrens Treefrog (*Hyla andersonii*): Concordance of genetic, morphometric and acoustic signal data,” Mol. Ecol. **24**(13), 3281–3298. doi:10.1111/mec.13242

Weihs, C., Ligges, U., Luebke, K., and Raabe, N. (**2005**). “klaR analyzing german business cycles,” In D. Baier, R. Decker, and L. Schmidt-Thieme (Eds.), Data analysis and decision support, Springer-Verlag, Berlin, 335–343.

Wellock, C. D., and Reeke, G. N. (**2012**). “Quantitative tools for examining the vocalizations of juvenile songbirds,” Comput. Intel. Neurosci. **2012**, e261010. doi:10.1155/2012/261010

Xu, M., and Shaw, K. L. (**2019**). "The genetics of mating song evolution underlying rapid speciation: Linking quantitative variation to candidate genes for behavioral isolation," Genetics, **211**(3), 1089–1104. doi:10.1534/genetics.118.301706

Zann, R. A. (**1996**). *The zebra finch: a synthesis of field and laboratory studies*, Oxford ornithology series, Oxford University Press, Oxford; New York, 335 pages.