Methodological Comparisons of Classification Techniques and Refinement of an Operant go/no-go Procedure

by

Prateek Kumar Sahu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Psychology
University of Alberta

Abstract

Bioacoustic analyses 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. Here, I 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 our models. My results presented here agree with previous results and identified highly convergent acoustic features involved in this classification. My comparison of classification methods gives researcher much needed information to select an optimal classification method.

Refining and modifying experimental procedures play a vital role in improving methodology while also reducing animal distress. Standardized experimental procedures are often modified for obtaining optimal responses. Many automated operant conditioning procedures have been developed and modified over time. Here, I asked if an increase in feed time duration affects discrimination in an operant go/no-go task. Specifically, we used zebra finches' sexually dimorphic distance calls as acoustic stimuli to test whether there were any significant differences in performance on an operant discrimination task requiring zebra finches to classify calls according to the sex of the producer when a key experimental parameter, feed time duration, was increased from 1 second to 2 seconds. We found no significant differences in learning speed (trials to criterion) between birds that were given 1 sec or 2 sec of food access following a correct go response. Our results indicate doubling food access duration did not impact the speed of acquisition of distance call discrimination in zebra finches. These findings suggest that we can provide twice as much time for zebra finches to access food, potentially improving animal welfare, whilst still being able to compare new results with historical results.

Preface

All procedures followed the Animal Care (CCAC) Guidelines and Policies and were approved by the Animal Care and Use Committee for Biosciences at the University of Alberta (AUP 1937 and AUP 2923).

Chapter 2 of this thesis has been accepted for publication as P.K. Sahu, K.A. Campbell, A. Oprea, L.S. Phillmore, & C.B. Sturdy, “Comparing methodologies for classification of zebra finch distance calls”, Journal of the Acoustical Society of America, *accepted*. I was responsible for the concept formation, experimental design, data analysis, and manuscript composition. K.A. Campbell helped in parts of data analysis and manuscript edits and revisions. A. Oprea and L.S. Phillmore contributed parts of data collection and manuscript edits and revisions. C.B. Sturdy was supervisory author and was involved in concept formation and manuscript edits and revision.

Chapter 3 of this thesis has been submitted as a publication as P.K. Sahu, C. Montenegro, C. T. Lambert, A. Oprea, M. Deimeke, V. Rennie, S. Smeltz, T.J. Benowicz, D. Patel, L.S. Phillmore, L.M. Guillette, & C.B. Sturdy, “Effect of feed-time duration on discrimination of vocalizations in a go/no-go operant paradigm”, Behavioural Processes, *submitted.* I was responsible for the concept formation, experimental design, data analysis, and manuscript composition. C. Montenegro, M. Deimeke, V. Rennie, S. Smeltz, T.J. Benowicz, D. Patel, A. Oprea, L.S. Phillmore helped with data collection and manuscript edits. C. T. Lambert & L.M. Guillette helped in experimental design, and manuscript edits. C.B. Sturdy was supervisory author and was involved in concept formation and manuscript edits and revision. The data obtained from the experiments of this chapter was a part of another study and has been published as C.T Lambert, P.K. Sahu, C.B. Sturdy & L.M. Guillette, “Among-individual differences in auditory and physical cognitive abilities in zebra finches” Learning & Behavior.

Dedication

My thesis is dedicated to all my cousins, Sapna, Trupti, Shruti, Ishi, Aryan, and Prabhat who always get excited about my work.

Acknowledgments

I would like to thank my supervisor Dr. Christopher B. Sturdy for his constant guidance and support throughout. I would also like to thank my supervisory committee, Dr. Lauren Guillette and Dr. Pete Hurd for their time, mentorship, and advice throughout the past three years. I thank Dr. Dana Hayward, Dr. Kimberley Mathot, and Dr. Kyle Nash for serving on my committee.

I thank all my former and current lab members, Dr. Carolina Montenegro, Dr. Kimberley Campbell, Dr. Jenna Congdon, Dr. Erin Scully, William Service, Moriah Deimeke, Katharine Stenstrom, Victoria Rennie, Sarah Smeltz, all other undergraduate students, WISEST students, and all volunteers. Thanks for answering all my questions, helping me transitioning to this new place, always there to talk, and guiding me throughout.

I would like to thank all my collaborators, Dr. Leslie Phillmore, and Dr. Lauren Guillette whose help allowed me to develop the projects and whose comments helped grow me as a researcher.

Thank you to my family who always supported me, for what I do. I thank all my friends who are always there when needed to talk. Thank you for all the encouragement and support.

I thank our shop personnel, Isaac Lank, Philip May and late Al Dennington for their amazing help in running and fixing our instruments. Thanks to Sarah Collard, Brianna Blake, and entire Science Animal Support Services team for taking care of our birds and helping me whole time.

Table of Contents

List of Tables

List of Figures

Chapter 1. General Introduction

Zebra finches (*Taeniopygia guttata*) are well established animal models for understanding mechanisms of vocal learning using behavioural, neurobiological, hormonal, and mathematical paradigms (Brainard & Doupe, 2002; Zann, 1996). The zebra finch vocal repertoire contains many vocalizations including songs and calls (Elie & Theunissen, 2016). For this thesis, I will focus on distance calls, a type of sexually dimorphic contact call emitted by zebra finches on take-off, in flight, and prominently when a bird is isolated from partner or flock (Zann, 1996). In this chapter of Introduction, I begin by reviewing various methods of classification used in bioacoustic analysis, keeping focus on binary classification a method which entails separation of the two categories based on acoustic features e.g., differences in sex. Later in this chapter, I outline the importance of refinement and standardization of experimental procedures with focus on operant conditioning in birds. In Chapter 2, I compare the performance of the classification methodologies using distance call datasets. In Chapter 3, I test the effect of increasing in feed time duration in an operant paradigm on discrimination of distance calls. I conclude Chapter 1 by detailing the aims and questions that I will focus on this thesis.

1.1 Bioacoustic analysis

Animal communication can be defined as sending and receiving specific signals from one individual to other individual(s) of same or different species, usually for a specific function (Bradbury & Vehrencamp, 2011). The modality of the signal depends on animal species and its social organization, primary sensory system, environment, behavioural context, and function of the signal (Ryan, 1990). The signal can be visual, acoustic, chemical and/or tactical in nature depending on type of signal, its biological function and presence of environmental constraints (Bradbury & Vehrencamp, 2011; Hopp et al., 1998). Acoustic communication is widespread throughout the animal kingdom from insects to birds and mammals (Hopp et al., 1998). The field of *Bioacoustic* includes the study of sound production, transmission, and reception in animals (Hopp et al., 1998). A major branch of bioacoustic research focuses on classification of animal vocalizations into vocal repertoires and trying to describe and define the classes that characterize animal vocalizations. The classification can be interspecies, identifying species based on vocalization or intraspecies (Chou et al., 2007; Chou & Liu, 2009; Fagerlund, 2007; Piczak, 2016), classifying vocalizations into vocal repertoires based on type of vocalizations (Elie & Theunissen, 2016, 2018), or based on sex (Campbell et al., 2016), geography (Hahn et al., 2013; Tuncer, 2013; Warwick et al., 2015), and elevation (Branch & Pravosudov, 2015, 2019). The classification of vocalizations can also aid in species monitoring biodiversity research and understanding the functional role of vocalizations (Mcloughlin et al., 2019; Penar et al., 2020; Teixeira et al., 2019).

1.2 Methods of classification

In the field of classification, observations with similar characteristics are classified into classes (Izenman, 2008). The classes may include presence or absence of a particular disease, spam or non-spam email, blood groups, sets of neurons, species of birds, or types of vocalizations. In classification, these classes can be separated by an algorithm, known as *classifier*. The information from the labeled observations is used for the construction of the classifier which separates the classes. The classifier can predict the class of new unlabeled observations, and this is known as supervised learning. The variables which are included for classification are called *feature variables*. Each variable may contain several observations. The feature variables with the observations are known as *feature vectors* (Izenman, 2008; James et al., 2013). The problem of classification can be divided into two broad categories, binary classification, and multiclass classification. As the names suggest, binary classification involves two classes and multiclass classification involves classifying observations into one of several classes. In this thesis, we are primarily interested in binary classification.

We see binary classification problems in our everyday life, whether it's presence or absence of a disease through medical testing, credit risk calculation, quality control decision or email spam detection (Duda et al., 2001). We select the particular method of binary classification depending on properties of the feature vector. Some of the popular methods of classification are Logistic regression, Probit model, Discriminant function analysis, Decision trees, Random Forest, Bayes's rule classifier, Support vector machines, and Artificial neural networks (Duda et al., 2001; Izenman, 2008). Each method has its advantages and disadvantages. For our interest and limited scope of the thesis, we will focus on three of these methods: Discriminant function analysis (DFAs), Support vector machines (SVMs) and Artificial neural network (ANN).

1.2.1 Discriminant function analysis (DFA)

In 1936 R. A. Fisher developed discriminant functions for a taxonomic problem where Fisher classified plants to its species based on multiple measurements of the flowers (Fisher, 1936a). Fisher's discriminant function uses dimensionality reduction technique where the ratio of the between-class variance to the within-class variance is maximized. Here, the linear combination of feature vectors separates the two classes as much as possible. When the assumption of normality for the probability density function of feature variables for each class is met, the discriminant function analysis is known as *Gaussian* linear discriminant analysis or linear discriminant analysis (LDA) (Izenman, 2008).

Bayes's rule classifier is used for classifying observations between two classes. Each observation belongs to a class with *prior probability*. From Bayes's theorem, one can obtain the *posterior probability* for an observation belonging to a class. The posterior probability for an observation is calculated for both the classes. A ratio of posterior probability for both the classes is used for classifying an observation to a class. Let's say, the ratio of posterior probability of class 1 to class 2 is higher than 1, then the observation will belong to class 1 else to class 2. In linear discriminant function analysis (LDA), the logarithm of the ratio of the two posterior probabilities is used, where a ratio greater than zero belongs to one class; else belongs to the other class.

The linear discriminant function (LDF) can be calculated using training datasets. Then, LDF can be cross validated using various methods. In leave-one-out cross-validation procedure, one observation is dropped from the dataset and LDF is calculated from remaining (n-1) observations. Then, the omitted observation can be classified using the LDF. This procedure is repeated until all the observations are classified. In the holdout method of cross-validation, the whole dataset is divided into two sets, training set, and testing set. The LDFs are calculated on the training set and performance of this model is evaluated with the testing set. We can then construct a *confusion matrix* showing true classifications and misclassifications out of total classifications for the classes. Another way to calculate LDF is using multiple regression. The target variable in multiple regression can be the classes and the feature vector as predictor variables. We can regress the classes on the variables to obtain regression coefficients. This can be used in variable selection especially for high-dimension data based on significant coefficients obtained from multiple regression.

The requirements for a dataset for it to be used in DFA are rarely satisfied. This can potentially lead to over or under representation of the datasets. Randomization testing is one of the ways to tackle this problem (Manly & Navarro Alberto, 2020). Randomization testing helps in determining whether the null hypothesis, in which the pattern observed in the data is purely a chance event, is a reasonable assumption. In this randomization testing, a test statistic measure is chosen to obtain the extent of pattern observed in the data. The same test statistic can be obtained for the null distribution of the data which is constructed from randomization. Now, we compare both test statistics for observed data and random null distribution data. The point being, if the null hypothesis is true, the observed distribution is one of the possible outcomes. Thus, we can observe this while comparing test statistics for both null and observed distribution of data.

Mundry and Sommer, 2007 developed permuted DFA (pDFA) procedure to solve the dataset misrepresentation issue for bioacoustics analysis. In the real world the data are messy, and they rarely fit for various discriminant function analyses for strict requirements as mentioned above. In bioacoustics analysis, one of the major problems is having multiple vocalizations from the same individual. Permuted DFA (pDFA) solves this by constructing a test statistic which takes care of this issue while comparing to the null distribution (Mundry & Sommer, 2007).

1.2.2 Support vector machines (SVM)

Support vector machines (SVMs) are a relatively modern methodology for binary classification which can perform better than discriminant functions in terms of classification accuracy depending on data structure. SVMs were first introduced by Boser, Guyon, and Vapnik in 1992 (Boser et al., 1992; Cortes & Vapnik, 1995). SVMs use supervised learning methods for the binary classification, where training datasets are used for class prediction of testing datasets. SVMs work well for both linear and non-linear data, giving SVM clear advantages to other methods like neural networks and decision trees (Cortes & Vapnik, 1995; Jakkula, 2011). SVMs use a new mathematical tool known as *kernel methods* which can be very successful with non-linear data.

SVM uses *maximal margin classifier* as a decision rule for classification as compared to DFA which uses Bayes's rule classifier as described above in section 1.2.1 (Izenman, 2008). The primary problem of binary classification is to construct a function using observed data, which can classify new data to either of the classes. Suppose f (x) is a separating function, then with a new datum x, the sign of f (x) will determine which class the datum x belongs to. The function f (x) will assign all points with positive values to one class and negative values to the other class. We can construct a hyperplane with a linear combination of observed data that can separate positive and negative functions. There can be an infinite number of such separating hyperplanes. We are interested in a hyperplane which has *maximum margin*. The *margin* is the sum of the shortest distance from the separating hyperplane to the nearest negative data point and to the nearest positive data point. Thus, when we maximize this margin, we can say this as a *maximal* *margin classifier* (Figure 1.1).

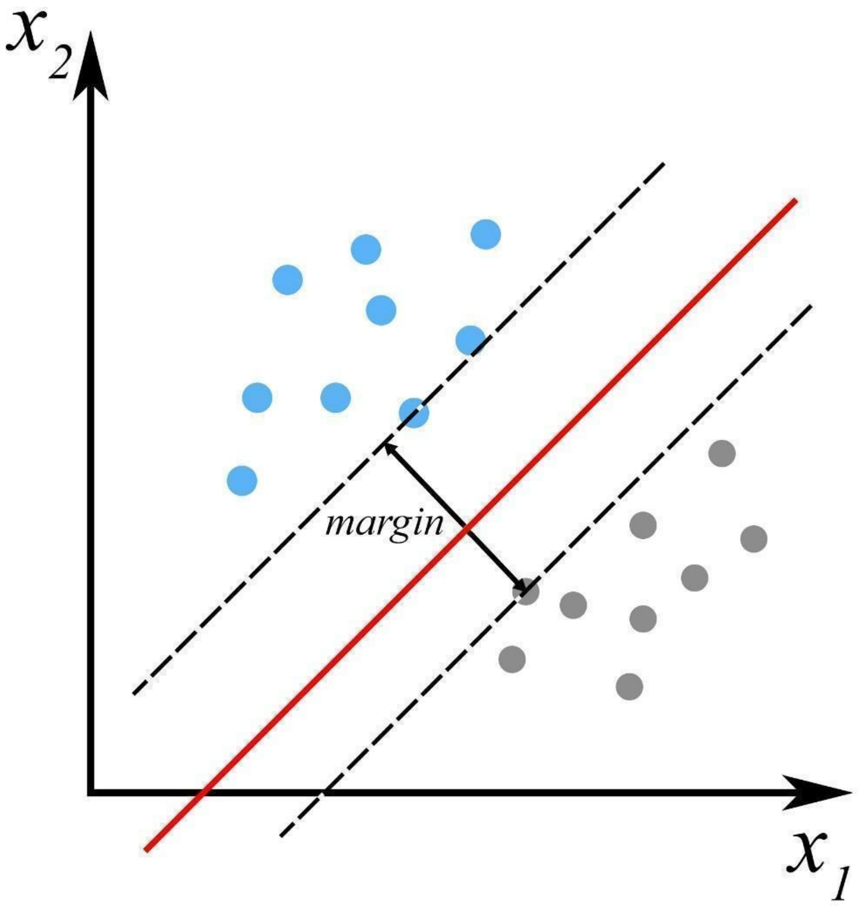

Figure 1.1 Showing a maximum margin classifier. Data points obtained from a separating function $f(x)$ are plotted in x_1 and x_2 axes. The blue data points can correspond to all the data points having negative sign $f(x)$ and gray data points to positive sign $f(x)$. The separating hyperplane is shown in a red line. The dotted lines are two other hyperplanes, and points on those hyperplanes are support vectors. The black line with arrowheads is the margin of the separating hyperplane.

Similar to DFA, we can use the holdout method of cross-validation dividing the whole dataset into training and testing sets and use testing sets for cross-validation. Then a *confusion matrix* can be constructed to represent true classifications and misclassifications.

1.2.3 Artificial neural network (ANN)

Initially, ANNs were developed to model human intelligence and mimic neural activities in the brain. Now, ANNs are used to solve problems of pattern recognition, classification, and prediction such as speech recognition, image recognition, and robotics (Haykin, 2009). In 1943, McCulloch and Pitts constructed a simplified model of neural activities in the human brain, *artificial neural networks* (McCulloch & Pitts, 1943). In this simplified model, the artificial neurons consisted of multiple inputs (equivalent to dendrites of a biological neuron) and a single output (axon). The input units could take values of 1(on) or 0 (off). Multiple input units with 1 and 0 are summed through logical functions (“AND” and “OR”) to produce a target output. The value of output is compared with a predefined *threshold value*. If the value of target output is larger than threshold value, the neuron will fire. The McCulloch–Pitts neurons are not a good approximation of biological neurons and not flexible with variation in input parameters.

How does ANN learning take place? When there is repeated excitation of synapses, the synapses get strengthened. The strength of synaptic connection depends on the firing history of those neurons, the strength gets weaker with less firing. This is known as the “Hebb learning rule”. In 1949, Donald O. Hebb described this theory in detail, which became the foundation on how the brain functions (Hebb, 1949). Later, Frank Rosenblatt improved Hebb’s theory and built a *perceptron* system known as single layer perceptron (Rosenblatt, 1958). The perceptron system is similar to McCulloch–Pitts neuron except now the input units have connection weights. The connection weights represent strength of the neural connections. The network of neural connections can be represented in graphs as nodes and edges. The *activation value* of the output nodes can be calculated with a linear combination of input units. These are known as linear activation functions. We can filter the activation values for each node through a nonlinear *threshold activation function*. Meaning firing of the neuron will depend on these activation functions. Some common nonlinear activation functions include linear, ReLU, step function, logistic, and various sigmoid functions (Izenman, 2008; James et al., 2013).

In binary classification, the neural network consists of layers of input units and a single output unit. The output variable from the output until takes on value 0 or 1, depending on whether the neuron fires (1) or does not fire (0). Suppose there are two groups of observations which are linearly separable. This means there exist a set of weight connection vectors, which can separate the observations into two groups through linear combination with observations. We can use a learning algorithm to iteratively update learning, ultimately resulting in a separation of the observations. The *gradient-descent algorithm* is one such algorithm where the input observations are analyzed one at a time and classified to one of the two groups. Then the true group of the observation is revealed, and the classification procedure is updated according to this classification error (Duda et al., 2001; Izenman, 2008).

Recent research into ANNs from the 1980s to present time focused on multilayer perceptrons with the newer backpropagation algorithms (Haykin, 2009), mainly driven by increased computational power of more modern computers. A multilayer perceptron network consists of an input layer, hidden layer, and output layer. Multivariate statistics is used on the input vector (nonlinearly) to predict output. The hidden layer acts as a node in between input and output layer. A schematic is shown in Figure 2. For binary classification the output layer has one variable with value 0 (not fire) or 1 (fire).

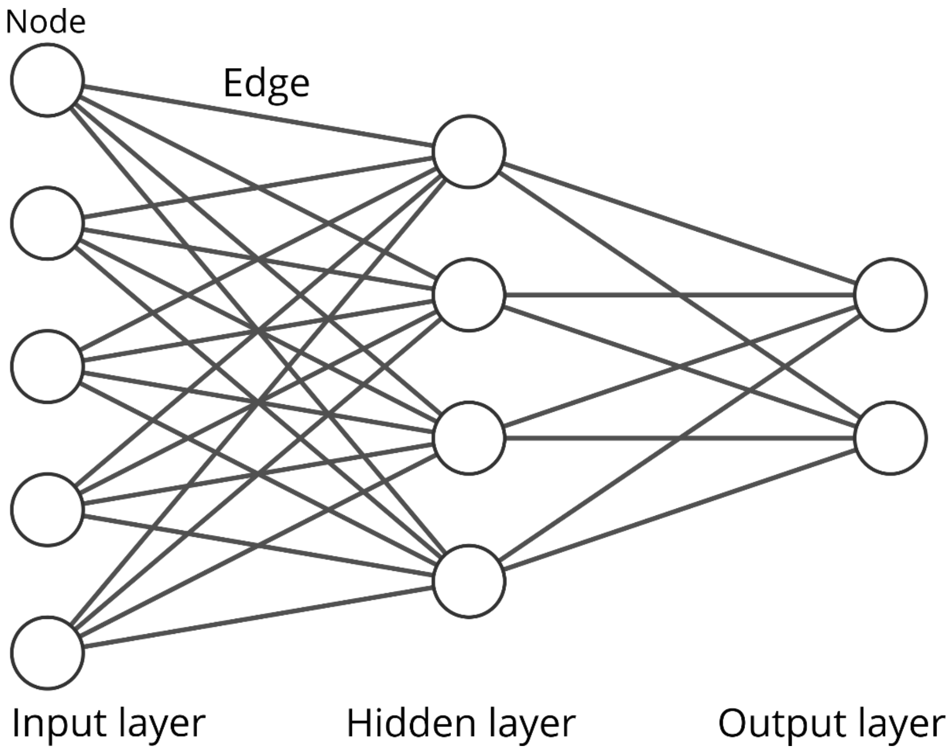

Figure 1.2 Showing schematic of an artificial neural network with input layer, one hidden layer and binary output layer. Individual neurons are represented as nodes (circles), and connections as edges (lines).

1.3 Refinement of experimental designs

In 1935, R. A. Fisher published “The Design of Experiments” on the principle of experimentation with a focus on statistical aspects of the design (Fisher, 1936b). We should design experiments with respect to scientific methods, while following standard laboratory procedures to produce valid reproducible data (Diamond, 2001; Johnson & Besselsen, 2002; Kirk, 2012; Larsson, 2001; Myers, 1979; Ryan & Morgan, 2007; Sproull, 2002; Weber & Skillings, 2017). Design of an animal experiment involves literature search, research questions and objectives, selection of animal model, standard research methods, and statistical analyses etc. with consideration to 3Rs (Replacement, Reduction, and Refinement) (*CCAC Guidelines:*

Animal Welfare Assessment, 2021; Johnson & Besselsen, 2002). Standardized research methods are essential for reproducibility (Richter et al., 2009; Würbel, 2000). Modifying factors in a standardized research procedure might impact validity, and reproducibility of the experiments (Bailoo et al., 2014; Richter et al., 2009; Würbel, 2000). Commonly used standard research procedures in behavioural testing are often reviewed and refined by manipulating various factors of the procedures. Studies have looked at biological factors (age, gender, body weight, strain etc.), environmental factors, food manipulations, schedule of treatments, dosage and type of drugs etc. affecting behavioural and physiological responses in the forced swim test (FST), a widely used behavioural test in rodents for basic research and pharmaceutical research (Bogdanova et al., 2013). The olfactory conditioning of proboscis extension response (PER), a paradigm to study learning and memory in honeybees, has been standardized (Bitterman et al., 1983) and reviewed (Matsumoto et al., 2012) as variation in conditioning procedure (schedule, intertrial interval, interstimulus interval, number of trials, intensities of conditioned & unconditioned stimulus etc.) may result in significant differences in acquisition and retention performances. Change in a consistently used experimental procedure requires investigating the impact on experiment outcomes.

1.3.1 Operant conditioning in small birds

In the 1970s, operant conditioning procedures began to be used successfully to study sensory processes in laboratory animals like rats, pigeons, and monkeys (Stebbins, 1970). In 1980s, Hienz et al. developed operant conditioning techniques to study auditory cognition and processing in birds such as redwing blackbirds (*Agelaius phoeniceus*) and brown-headed cowbird (*Molothrus ater*) (Buchholz & Hothersall, 1976; Hienz et al., 1977). Later, Park, Okanoya and Dooling successfully applied the standard operant conditioning techniques for other bird species including budgerigars (*Melopsittacus undulatus*), canaries (*Serinus canarius*), zebra finches (*Taeniopygia castanotis*), swamp sparrows (*Zonotrichia georgiana*), and song sparrows (*Zonotrichia melodia*) (Klump et al., 1995; Park et al., 1985). Auditory perception research with operant conditioning involved two types of response methodologies one in which birds make a response by hopping to the perch (Park & Dooling, 1991; Shy et al., 1986) and one in which they make their response by key pecking response buttons (Hulse & Cynx, 1986; Klump et al., 1995; Okanoya & Dooling, 1987; Park et al., 1985). Since then, automated systems for operant conditioning procedures (hopping to the perch and food delivery) have been developed and refined over the time (Berkhoudt et al., 1987; Brown & Riede, 2017; Gess et al., 2011; Nagel et al., 2010; Njegovan et al., 1994; Sturdy & Weisman, 2006; Varnon et al., 2018). An automated motor driven feeder (Njegovan et al., 1994) has been used extensively for operant paradigm in zebra finches (*Taeniopygia guttata*), wild-caught black-capped chickadees (*Parus atricapillus*), and wild-caught white-throated sparrows (*Zonotrichia albicollis*) (Congdon et al., 2019; Hoeschele et al., 2013; Klump et al., 1995; Montenegro, Service, et al., 2021; Njegovan & Weisman, 1997; Sturdy et al., 1999, 2001; Weisman et al., 2004). As discussed earlier, various operant conditioning factors, such as schedules of reinforcement, intertrial interval, interstimulus interval, number of trials, and intensity of reinforcement impact the response rate, acquisition, and retention performances (Gess et al., 2011; Klump et al., 1995). Duration of food access during operant trials is one of the important factors determining the performance in an operant task with respect to motivation. Here, I was interested in doubling food access time in an operant task, for a consistently used operant system (Njegovan et al., 1994; Sturdy & Weisman, 2006), and determining the impact on discrimination performance for zebra finches in a task discriminating female and male distance calls (Gess et al., 2011; Zann, 1996). Chapter 3 goes in details about the effect of increasing feed time duration on discrimination.

1.4 Aims of the thesis

In this thesis, I aim to answer the following questions:

1. Presenting a methodological comparison of various classification methods using zebra finch distance calls, and their bioacoustic measurements, as a test set.
2. Are there any impacts of changing feed time duration in an operant paradigm on discrimination of vocalizations?

First, I will describe DFA, SVM, and ANN classification methods for vocalization classification, then I will compare classification performance and consistency among these methods using sex-based classification of zebra finch distance calls from measured acoustic features from spectrograms. I will also analyze variables of relative importance for acoustic features used in the classification. The comparative bioacoustic analysis will provide a baseline and framework for future vocalization classification problems. The ranked acoustic features, ranked by relative variable of importance, can be used in future perceptual, neurobiological or playback studies to better understand biological basis of distance calls. Chapter 2 goes into details about the comparative bioacoustic analysis.

Second, I will increase the feed time duration in a go/no-go operant task and look at its effect on discrimination performance. I will use zebra finches as model system and zebra finch distance calls for a discrimination task. Zebra finches will discriminate between female distance calls from male distance calls in a go/no-go operant paradigm. I will compare the discrimination performance between birds with 1 sec food access to that of 2 sec food access during operant trials. This is an attempt to refine the consistently used operant paradigm (Njegovan et al., 1994; Sturdy & Weisman, 2006). If I find no significant differences in the discrimination performance, we can double the feed time duration during operant trials with no negative impact on discrimination performance. In doubling the feed time duration, birds will have more time to eat during trials, which can potentially improve their wellbeing. The manipulation of feed time duration and examining its effect on discrimination performance will allow for comparisons of past auditory discrimination experiments with future experiments (with modified operant procedures). In Chapter 3, I will discuss in detail about the effect of change in feed time duration on discrimination of female and male zebra finch distance calls.

Finally, I will conclude the thesis with Chapter 4, the discussions, and conclusions from above two studies. Overall, I compare bioacoustic classification techniques with advantages and disadvantages and I examine the impact of increase in feed time duration in an operant task on discrimination performance.

Chapter 2: Comparing methodologies for classification of zebra finch distance calls

2.1. 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 & 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 & 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 & 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., 2013; Tuncer, 2013) based on ecological habitats (Anderson et al., 2008; Gómez et al., 2018) or based on the individuals (Elie & Theunissen, 2016; Hahn et al., 2013; 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 & 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 & Zuberbühler, 2005), and individual recognition (D'Amelio et al., 2017; Elie & 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 & 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 & 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 & Theunissen, 2016; Wadewitz et al., 2015), and classifying based on sex (Campbell et al., 2016), geography (Hahn et al., 2013; Tuncer, 2013), and individuals (Elie & 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 sexually dimorphic in a number of important aspects. Only male zebra finches produce songs, though both sexes produce a variety of calls (Elie & 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 2.1). 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 & 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 & 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 & Theunissen, 2016). This analysis showed that females produced longer, and lower pitched distance calls compared to male distance calls (Elie & 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 measured and 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, I 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 when analyzing the entire repertoire of the Zebra finch. I used 10 acoustic measurements in our analyses including both temporal and spectral measures. I hypothesize similar classification performance in all three classification methods given the past successes using these methods for similar tasks (Elie & Theunissen, 2016; Mouterde et al., 2014). Furthermore, I 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 & Theunissen, 2016; Vignal & Mathevon, 2011) suggest that these are the key features facilitating sex-based call discrimination.

2.2. Methods and Results

2.2.1 Recordings

In total, N=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 depending on availability. 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.

2.2.2 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 & Nelson, 1990). The fundamental frequency was measured in Praat 6.1.38 (Boersma & van Heuven, 2001; Goldstein, 2021). Figure 2.1 shows the measured acoustic features from the spectrograms of male and female zebra finch distance calls.

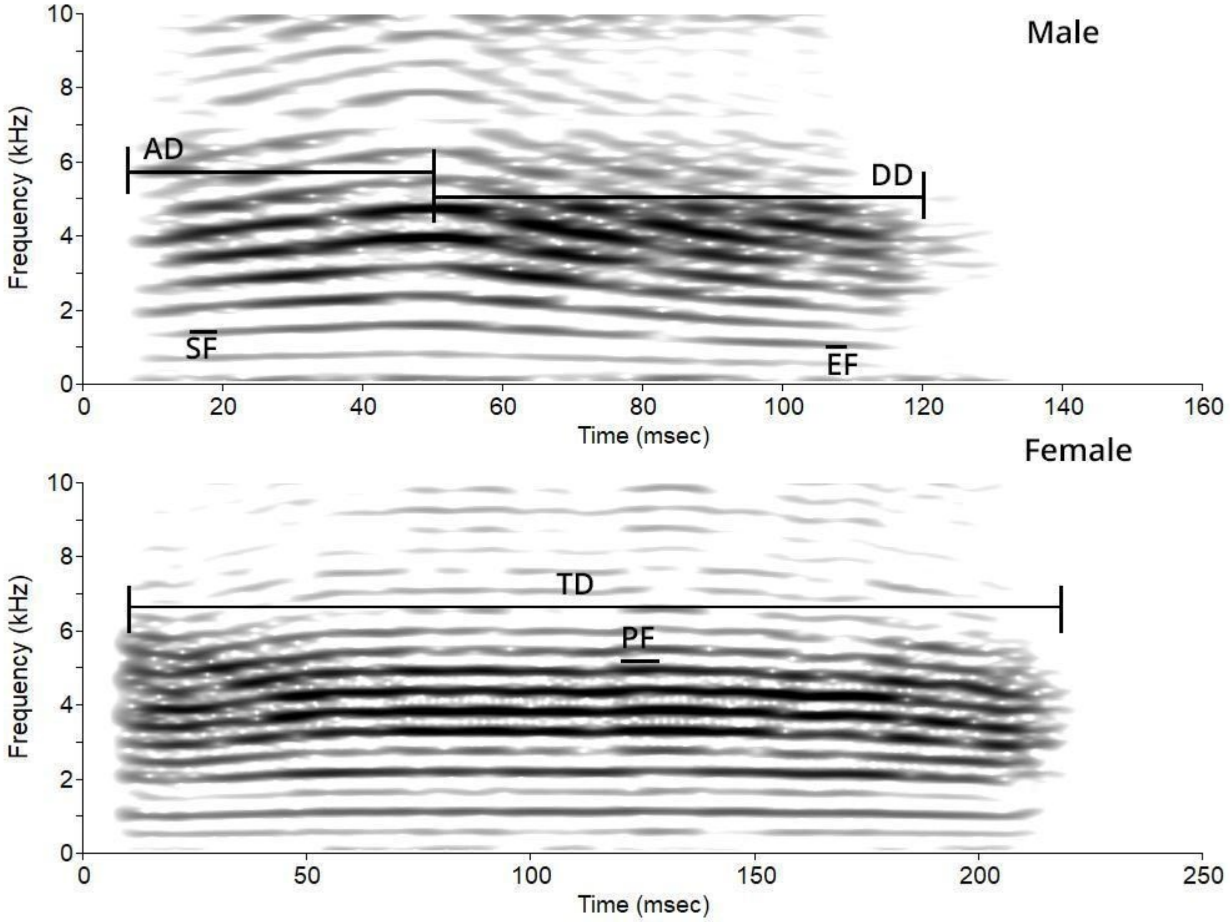

Figure 2.1 Measured acoustic features from the spectrogram showing Total duration (TD), Start frequency (SF), End frequency (EF), Peak frequency (PF), Ascending duration (AD), Descending duration (DD) with male distance call at top and female distance call at bottom. F0 was measured in Praat (not pictured here).

2.2.3 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 & 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 & 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 & 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* which evaluates variable importance through input-hidden and hidden-output connection weights (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 ($r \geq 0.75$ and significant). Table 2.1 shows correlation across the measured acoustic features.

Table 2.1 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	

2.2.3.1. 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 & Goldberg, 2020; Mundry & Sommer, 2007). For example, DFA has been used to classify vocalizations of mountain chickadees (*Poecile gambeli*) based on elevation gradient (Branch & 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).

I 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 (i.e. permutation test), thus giving us overall percent correct classification (Betz, 1987; Mundry & 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 & Sommer, 2007). A permuted DFA (pDFA) can be used to account for potential pseudoreplication if any is present (Mundry & Sommer, 2007). With pDFA, I 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 & 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) in order, 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. I 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.

I 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 on classification accuracy. Low MCC can potentially pose a problem resulting in a larger dataset overestimating the classifier. Figure 2.2 shows the distribution of the individual distance calls according to the first discriminant function, LD1, male and female calls are well separated.

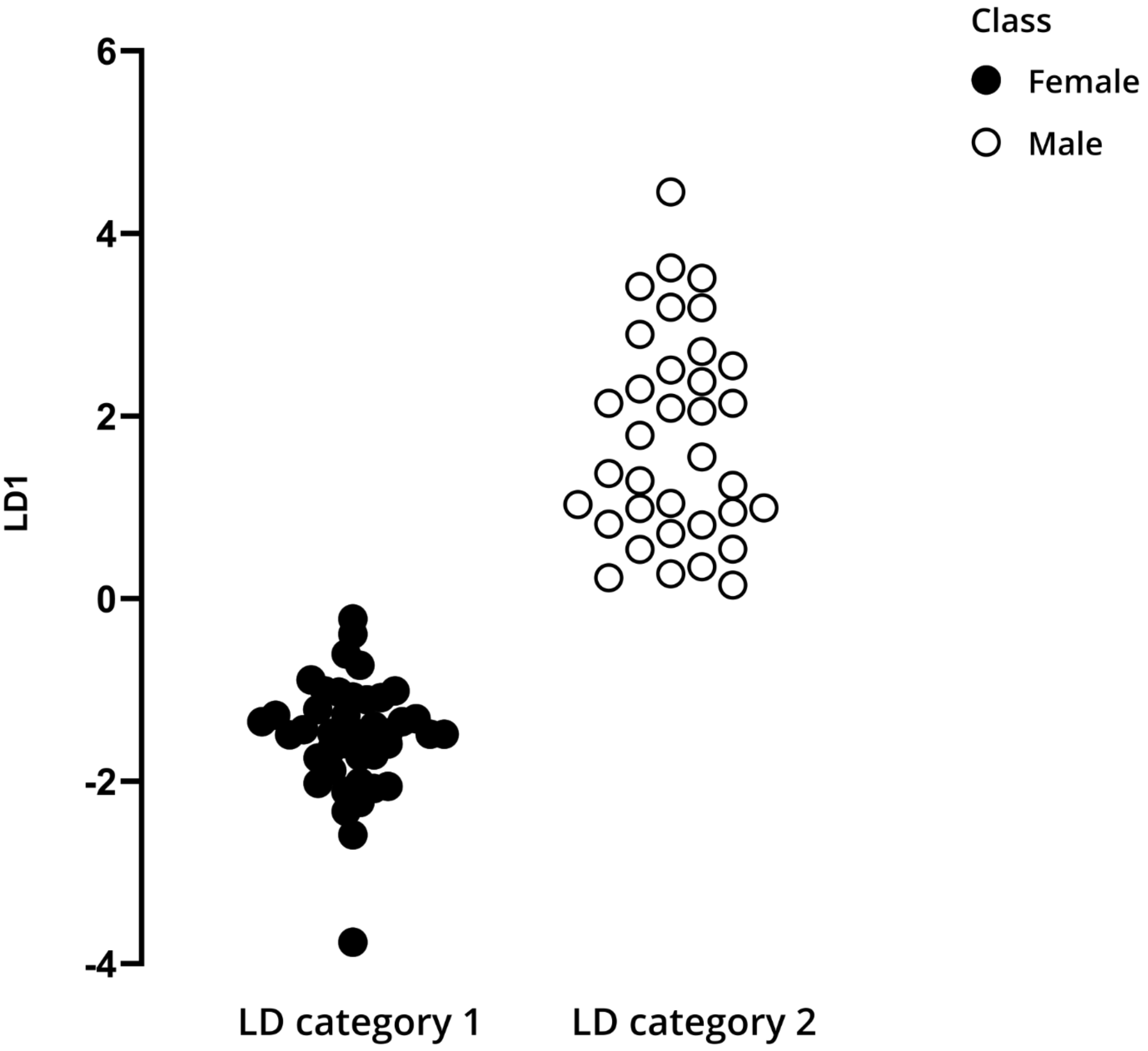

Figure 2.2 Distribution of the first discriminant function (LD1) for all male and female distance calls.

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. I 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.

2.2.3.2 SVM

Support vector machines (SVMs) are supervised learning algorithms used for mainly two-group classification problems (Cortes & 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, I 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

& Okanoya, 2016; Tachibana et al., 2014). In this study, I 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, I 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.

I 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).

2.2.3.3 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 & Liu, 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 & Bengio, 2004; Jakkula, 2011).

I used a supervised ANN, which used the eight measured acoustic features from distance calls as input, to classify the calls. I built an artificial neural network using the *neuralnet* package in R with the default logistic activation function (Günther & 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). I 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 & Fritsch, 2010). A schematic of the neural network is shown in Figure 2.3.

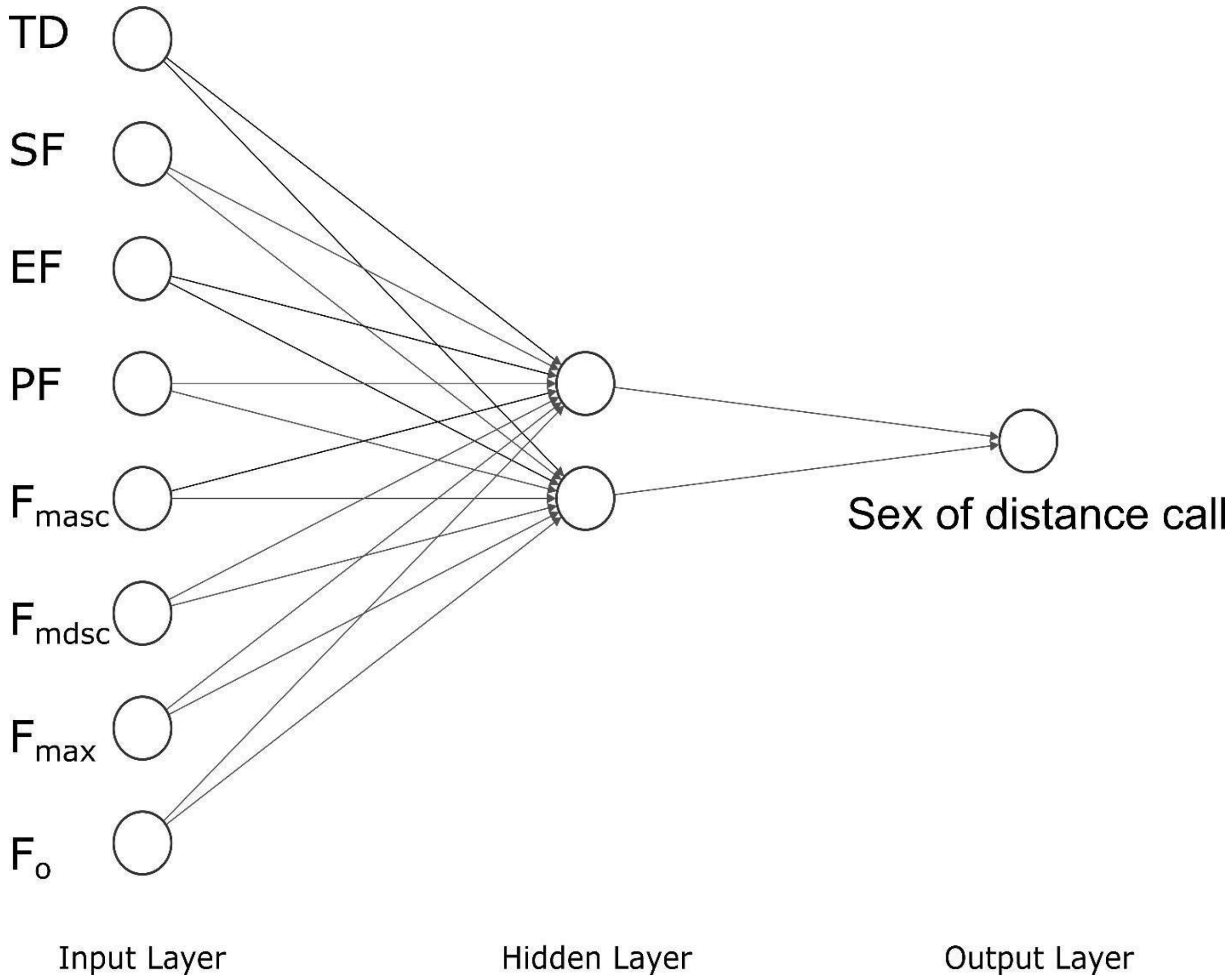

Figure 2.3 Schematics of the neural network showing acoustic features as input, two neurons in hidden layer and output layer where TD: Total duration, SF: Start frequency, EF: End frequency, PF: Peak frequency, Fmasc: Frequency modulation of ascending frequency, Fmdsc: Frequency modulation of descending 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. I 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 & 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.

2.2.3.4. 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), I 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 & Jurman, 2020; Matthews, 1975; Wellock & Reeke, 2012). MCC generates high score only if the classification predictor can correctly predict most of both classification categories. 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, I 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.

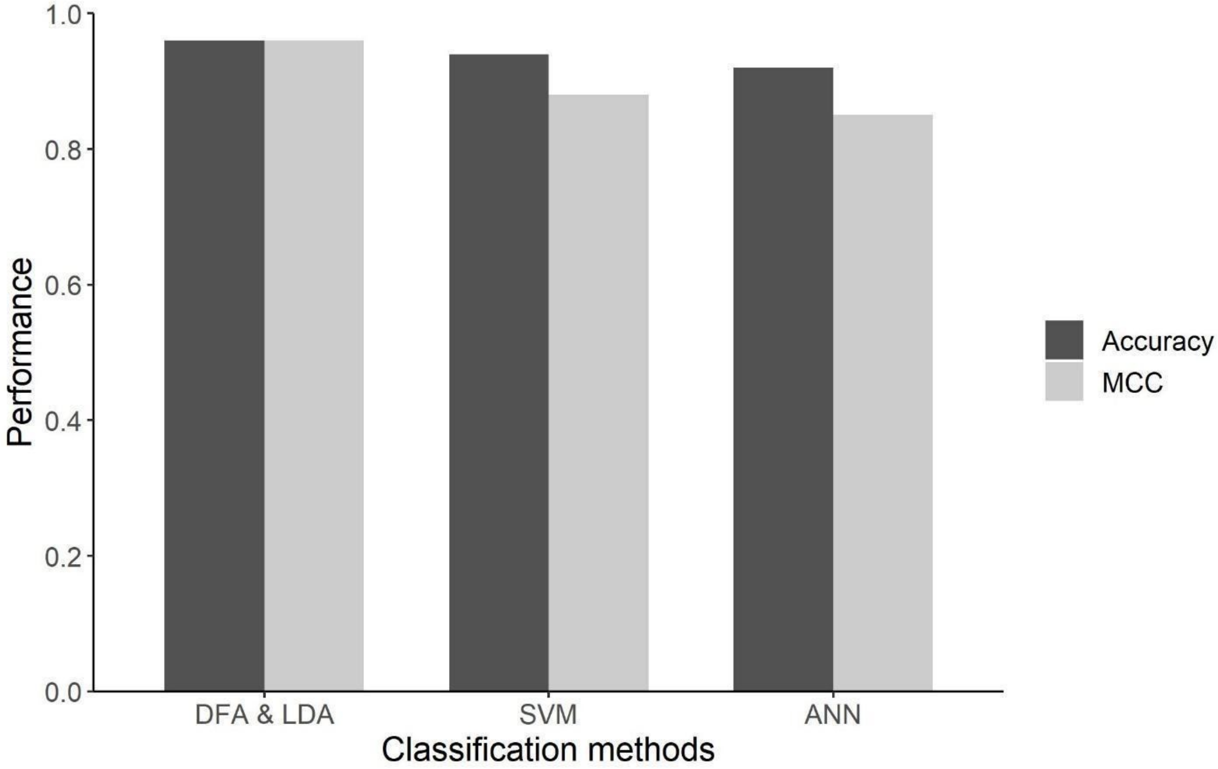

Figure 2.4 A bar graph showing classification accuracy (scaled to 1) and MCC (Matthews correlation coefficient) values for all classification methods; pDFA & LDA, SVM, and ANN. Higher value is better.

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 2.2 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 2.2. 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.

2.3 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, I recommend the use of a combination of these methods to account for stochasticity of real-world data. Pragmatically, if I 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 & 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 I 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). I 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., 2003, 2006; 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. Birds might use both spectral and temporal acoustic features for sex identity of the caller. 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 & Qureshi, 2017; Tooze et al., 1990), Gini index, or mean decrease accuracy for Random Forest algorithm (Armitage & Ober, 2010; Elie & 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. I used similar measures for variable importance and expanded with the connection weight algorithm (Olden & 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, I 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. I 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).

Chapter 3. Effect of feed-time duration on discrimination of vocalizations in a go/no-go operant paradigm

3.1. Introduction

Experimental methodologies often focus on optimizing behavioural responses and minimizing distress to animals, whether for a newly developed paradigm, or to refine existing paradigms (Klump et al., 1995). Operant conditioning procedures are often optimized for high rates of engagement by balancing high levels of motivation and optimal food access (Gess et al., 2011; Goltstein et al., 2018; Kim et al., 2017; Phillips et al., 2017). Many laboratories have developed operant paradigms to test auditory perception in songbirds (Gentner et al., 2000; Gess et al., 2011; Houx & Ten cate C, 1999; Nagel et al., 2010; Njegovan et al., 1994; Park et al., 1985; Scharff et al., 1998; Sturdy & Weisman, 2006). Refinement and improvement of an extensively used operant paradigm (Njegovan et al., 1994; Sturdy & Weisman, 2006) can involve manipulation of reinforcement factors such as type, strength, and schedule, coupled with investigating the effect of these variations on performance in the operant task (Scheiner et al., 2004; Stebbins et al., 1959; Trosclair-Lasserre et al., 2008). Duration of food access during operant conditioning trials may impact motivation, task performance, and even the wellbeing of an animal.

In the current study, I investigated whether changing the duration of food access after a correct response in an operant go/no-go discrimination task impacted the discrimination performance of zebra finches (*Taeniopygia guttata*), a widely used model species for neurobiological and behavioural studies of song development and auditory perception (e.g.

Adret, 1993; Brainard & Doupe, 2002; Geberzahn & Derégnaucourt, 2020). Previously, each bird received 1 sec of food after rewarded responses during discrimination trials (Campbell et al., 2020; Congdon et al., 2021; Guillette et al., 2013; Hahn et al., 2015; Hoeschele et al., 2013; Montenegro et al., 2021; Scully et al., 2020). In an attempt to refine our procedures, and potentially increase welfare of our experimental animals, I increased the feed duration from 1 sec to 2 sec which allow us to directly compare discrimination performance between conditions. If birds perform similarly in both 1 sec and 2 sec conditions, we could refine our experimental procedures by increasing the feed duration without compromising the comparability of future studies with past results. Two groups of birds were trained to discriminate zebra finch distance calls (Elie & Theunissen, 2016; Gess et al., 2011; Zann, 1996) with either 1 sec and 2 sec of reward; both groups of birds discriminated female distance calls (go or S+) from male distance calls (no-go or S-).

3.2. Methods

3.2.1. Subjects

In total, 25 adult zebra finches were tested (14 male, 11 female) between February and August 2021. One female bird died during the training stage due to natural causes. Six birds (two males and four females) failed to learn to use the perch and feeder to obtain food during the training stage so were removed from the experiment and any further analyses. Birds were bred and raised at the University of Alberta, Canada. The housing rooms were maintained on a 14:10 light:dark cycle (lights on 0700, full spectrum lights - Standard, 32W, T8 Daylight) at ~ 21°C temperature and humidity ~ 40%. Birds were provided with spray millet once per week and spinach and Prime Vitamin Supplement (Hagen) three times a week. The birds (25 in total) were naïve to the experimental procedures, including the acoustic operant conditioning task, and call stimuli, but had experience with other cognitive tasks (Camacho-Alpízar et al., 2021, Lambert et al., 2022). For detailed housing information refer to Lambert et al. 2022.

3.2.2. Apparatus

During the experiment, each bird was housed in a modified cage (30 × 40 × 40 cm) placed inside a ventilated, sound-attenuating operant chamber, and maintained with a 14:10 hour light:dark cycle. For all other details about apparatus refer to Lambert et al. 2022. For a diagram and detailed description of the apparatus, see Sturdy and Weisman, 2006.

3.2.3. Acoustic stimuli

A total of 60 zebra finch distance calls were used in the experiment: 30 male and 30 female calls produced by 18 males and 18 females with 1-2 calls used per individual (Lambert et al. 2022).

3.2.4. Procedure

3.2.4.1. Initial shaping and training

Once each bird learned to use the request perch and feeder to obtain food, training to discriminate a tone (1,000 Hz; to receive food access) from no tone (0 Hz) began. At this point in tone/no tone training, each bird was randomly assigned to one of two treatment groups (1s group: 1 sec food access and 2s group: 2 sec food access). After successfully learning to discriminate (DR ≥ 0.80 for three 500 trial bins; see definition of DR in section 2.5) tone from no tone, each bird moved to Non-differential training where they were exposed to and reinforced for responding to each of the 60 stimuli. The goal of this phase was to ensure each bird responded to all stimuli equivalently prior to discrimination training. For a detailed training criteria for the initial shaping and training, refer to Lambert et al. 2022.

3.2.4.2. Discrimination training

In Discrimination training, 40 of the 60 calls were randomly selected and presented as training stimuli. Responses to half of the stimuli (S+ = 20 female distance calls) were reinforced with 1 sec or 2 sec access to food, according to the treatment group, and responses to the other half (S- = 20 male distance calls) were punished with a 30 sec intertrial interval with lights off. Discrimination training continued until each bird completed six 320-trial blocks with a discrimination ratio (DR, see definition in section 2.5)) ≥ 0.80 with the last two blocks out of six blocks being consecutive with DR ≥ 0.80.

3.2.5. Response measures

For tone/no tone training (DR = (response to tone trials)/sum(response to tone and no tone trials) in a 500 trial block) and discrimination training, a discrimination ratio (DR) was calculated. The DR is a measure of how accurately a bird discriminates rewarded calls (S+) from unrewarded calls (S-). A DR was calculated using the formula: (R+S+)/sum(R+S+ and R+S-), where R+S+ is the mean proportion of responses for block of 320 trials when rewarded calls (S+) were played and R+S- is the mean proportion response when unrewarded calls (S-) were played. A DR of 0.50 indicates equal response to rewarded (S+) and unrewarded (S-) stimuli, and a DR of 1.00 indicates a bird only responded to S+, thus a perfect discrimination. Average number of trials per day was measured as a proxy to compare motivation and total trial blocks required to reach the criterion in the Discrimination training stage was measured to compare the speed of acquisition between birds in the treatment groups.

3.2.6. Statistical analysis

All statistical analyses were conducted in R v3.6.2 (R Core Team, 2019). Below data are represented as mean ± SD. I conducted a mixed-model Analysis of Variance (ANOVA) on average number of trials per day as a proxy for motivation to compare across feed time groups and sex in different experimental stages. If there are significant differences in average trial numbers for each stage, it would mean a change in feed time duration affecting completion time of an experiment. The Greenhouse-Geisser sphericity correction was applied due to a violation of sphericity assumptions. The assumption of normality was violated (from the Shapiro–Wilks test) for the distribution of a few groups below. Therefore, I used a robust two-way ANOVA with the WRS2 package (Mair & Wilcox, 2020) using trimmed means (20%) to examine the main effect of feed time groups (1s and 2s) and sex (female and male), and the interaction between feed time group and sex based on the total number of trial blocks required to reach criterion. An alpha of p =.05 was used as the cutoff for significance.

3.3. Results

3.3.1. Trials per day

I used the average number of trials completed per day for different stages (Non-differential and Discrimination training) as a proxy to compare motivation for speed of task completion for feed time groups and for females and males. I conducted a Group × Experiment Stage × Sex, mixed model ANOVA on average number of trials per day with experiment stage as the within-subject factor and feed time groups and subject sex were between-subject factors. There was a significant main effect of the experimental stage ($F = 10.6$, $p = 0.002$). There were no other significant main effects or interactions (all *p*s > 0.07; Figure 3.1). Birds performed significantly more trials per day in Discrimination (1391.7 ± 374.5) than Non-differential training (883.65 ± 167.1).

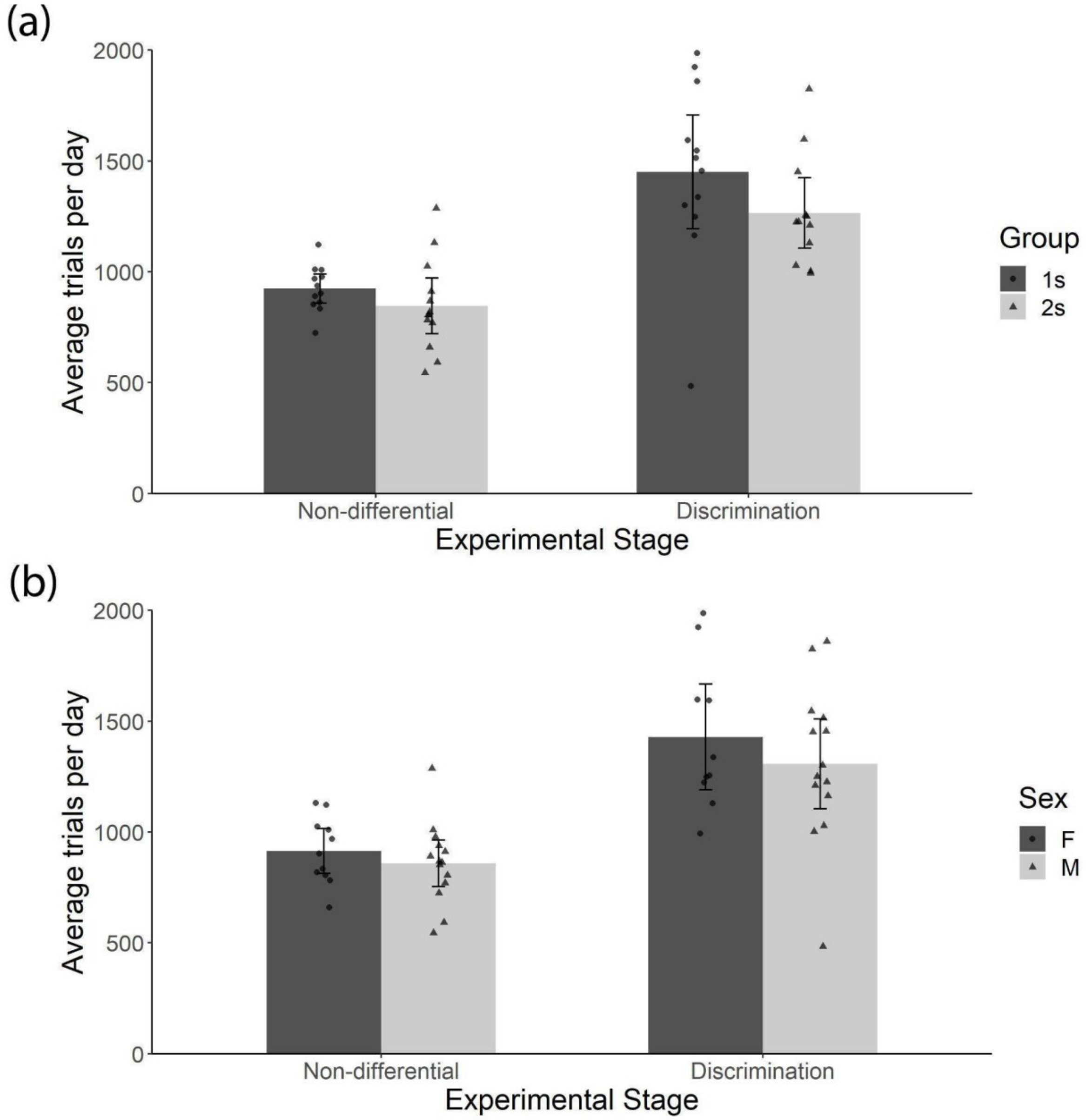

Figure 3.1 Bar and scatter (jittered) plot showing average trials per day for feed time groups (a), and for females and males (b). The bars represent average number of trials completed per day for experimental stages (Non-differential and Discrimination) for feed time groups and sex (dark

gray: 1s and female and light gray: 2s and males). No significant difference across feed time groups and sex was found. Each filled circle (1s group and females) and filled triangle (2s group and males) represents individual birds. Error bars represent 95% confidence intervals. Note: same data is shown in both (a) and (b).

I conducted pairwise comparisons using *t*-tests with Bonferroni corrections to determine if 1s and 2s groups and females and males differed in the average trials per day in each experiment stage. There were no significant differences in the average number of trials completed per day for 1s vs 2s group and female vs male, in each individual experimental stage (all *p*s > 0.42).

3.3.2. Trials to criterion in Discrimination phase

I used the total number of 320-trial blocks required to reach the criterion (DR $\geq$ 0.80 for six blocks, last two blocks consecutive) to compare the speed of acquisition for feed time groups and for females and males. I conducted a two-way ANOVA with trial blocks required to reach the criterion as the dependent variable with feed time group and sex as factors. There were no significant main effects of the feed time group (adj. critical value = 0.49, $p = 0.5$) or sex (adj. critical value = 3.3, $p = 0.1$), nor was there a significant interaction between feed time group and sex (adj. critical value = 0.15, $p = 0.71$), see Figure 3.2.

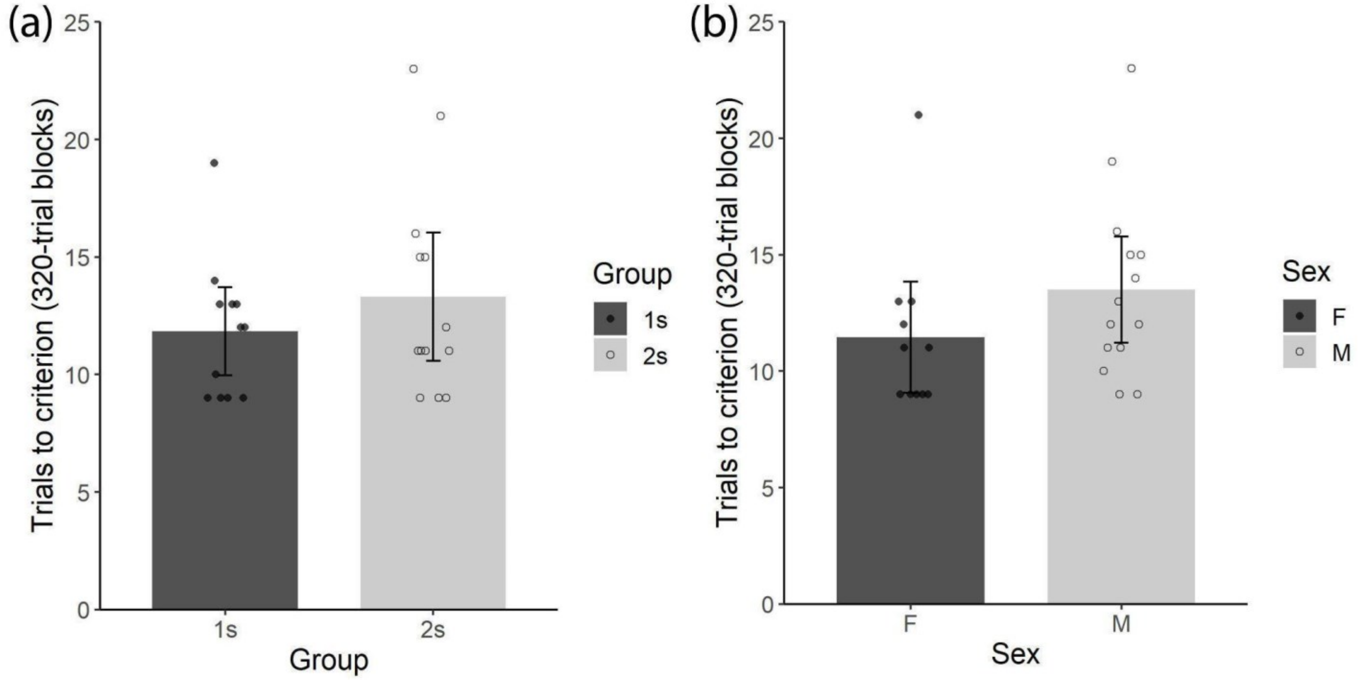

Figure 3.2 Bar and scatter (jittered) plot showing trial blocks required to reach criterion for Discrimination training for feed time groups (a), and for females and males (b). The bars represent the average number of trial blocks required for each group (dark gray: 1s and female and light gray: 2s and males). Each filled circle (1s group and female) and open circle (2s group and male) represents individual birds. No significant difference across feed time groups and sex was found. Error bars represent 95% confidence intervals. Note: same data is shown in both (a) and (b).

3.4. Discussion

In the current study, I asked whether modifying feed time duration from 1 sec to 2 sec influences the performance of zebra finches learning to discriminate between female and male distance calls. Birds given either 1 or 2 seconds to obtain food after responding to a rewarded stimulus did not differ in number of trials completed per day from, nor did the groups differ in the speed of acquisition of the discrimination task.

I used the number of trials per day as a proxy measure for motivation which was important to measure since the duration of food access might be expected to impact motivation. However, the increase in feed duration did not significantly impact the number of trials per day performed in either Non-differential or Discrimination stages. However, birds differed in the number of trials per day between experimental stages (i.e., Non-differential and Discrimination). This is likely due to differing degrees of difficulty of the experimental stages, as there is punishment for responding to S- stimuli in the Discrimination stage, compared to responding in the Non-differential stage where responses to all the stimuli are rewarded. Nevertheless, this means I can double the feed duration without significantly affecting the number of operant trials performed per day, which can impact the completion of an experimental stage. Birds can have more access to food reward during an experiment, potentially reducing the stress to the demands of having to eat quickly during a trial. The operant experiments with the current new feed durations (i.e., 2 seconds) can take a similar number of trials for completion and previous work with 1 second feed access, aiding in comparisons across studies. Our results show that at a minimum, I can successfully double the access time to food for birds without a significant impact on desired outcomes for sex-based discrimination of distance calls, at least in terms of the speed of learning. Doubling the feed access duration ensures birds get more time to eat food, which can potentially impact the well-being of the animals.

In an animal welfare study, Starlings when deprived of enrichments they show negative affective state through more pessimistic responses (Bateson, 2016; Bateson and Matheson, 2007; Matheson et al., 2008). Food and water restriction in a visual discrimination task in mice significantly influenced the task engagement (trials per session) and task acquisition (number of sessions for learning criteria) (Goltstein et al., 2018). Operant conditioning in mice has been standardized for optimal reproducibility and engagement through standardization of reinforcer's caloric content (Kim et al., 2017, Phillips et al., 2017). Our study adds to these studies, looking at optimizing engagement and food access in consideration with animal wellbeing. Future studies should consider examining the effect of an increase in feed duration on the well-being of the animals with comparisons of welfare measurements like weight, fat content, etc. before and after an experiment. These studies can help guide decisions about food access for long-term experiments where animals have limited food access for a longer period. Further studies with different species and more challenging discriminations for longer periods are required to get a complete picture of the effect of feed duration on operant-based discrimination tasks (Sturdy & Weisman, 2006). Nevertheless, our study takes an important step towards improving and refining experimental operant procedures without sacrificing discrimination performance.

Our study used calls as discriminative stimuli to look at the effect of feed time duration. Overall, it illustrates the effect of changing feed duration in an operant discrimination task. In the future, similar studies on the manipulation of other experimental factors can be conducted with other passerine bird species, or with more complex song discrimination in order to further improve the operant paradigm. Here, we have taken a step forward in improving experimental methodology and possibly animal-wellbeing while providing information about sex-based discrimination.

Ethical note

All the procedures were conducted in accordance with the Canadian Council on Animal Care (CCAC) Guidelines and Policies with approval from the Animal Care and Use Committee for Biosciences for the University of Alberta (AUP 1937 and AUP 2923), which is consistent with the Animal Care Committee Guidelines for the Use of Animals in Research.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank Isaac Lank and Philip May for their technical assistance. This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (NSERC RGPIN-2016-04591) and Discovery Accelerator Supplement (NSERC RGPAS-2016-412311), an Alberta Ingenuity Fund (AIF) New Faculty Grant, and the Canada Foundation for Innovation (CFI) New Opportunities Fund (NOF). CBS is also a member of the Neuroscience and Mental Health Institute, University of Alberta, Edmonton, T6G 2R3, Canada.

Chapter 4. Discussion

In this thesis, I addressed two issues which are crucial for researchers studying auditory cognition, communication: 1. vocalization classification and 2. refinement of experimental procedures. Classification of vocalizations helps in our understanding of and sheds light into the functional roles of the vocalizations through finding commonalities and patterns. Researchers have used many classification techniques to group and describe animal vocalizations. In this thesis, I compared binary classification methodologies, using sexually dimorphic zebra finch distance calls to understand advantages and disadvantages of each classification method. Secondly, refinement of experimental procedures is often required for fine-tuning and improvement of an experimental procedure. Here, I refined the feed time duration, an important factor in go/no-go operant conditioning procedure and looked at its effect on discrimination of vocalizations (female vs male zebra finch distance calls) which helps us by improving of an operant procedure by allowing increased food access to the birds performing an operant discrimination task.

4.1 What does comparing bioacoustic-classification methodologies tell us?

Classification of vocalizations is one of the major aspects of bioacoustic research (Garcia & Favaro, 2017). Many studies have been conducted in a variety of species and contexts in order to answer relevant questions of classification. These studies involve one or many statistical classification methodologies. The application and comparison of multiple classification methods does two things: 1. It provides a measure of classification accuracy via multiple methodologies and 2. It provides options to choose optimal classification methods for the current as well as future research problems.

In this thesis, I focused on binary classification, which is relatively simple but useful in contexts ranging from medical science like detection of diseases to computer science like spam detection (Abi-Haidar & Rocha, 2008; Izenman, 2008; Zhang et al., 2014). I compared Discriminant function analyses (DFAs), Support vector machines (SVMs), and Artificial neural networks (ANNs), which are widely used across disciplines for vocalization classification. Chapter 2 described how each classification method works, what are the differences in their classification algorithms, and what are the advantages and disadvantages of these methods across various contexts of vocalization classification. A detailed description and comparison of these methods helps in understanding the classification process of the different groups of stimuli. Biological datasets are usually complex and often context dependent, thus it becomes difficult to choose an optimal classification method. Studies applying and comparing various classification methodologies help in answering a particular research question and provides a base for solving future research problem involving similar datasets (Armitage & Ober, 2010; Caycedo-Rosales et al., 2013; Colonna et al., 2016; Ivanenko et al., 2020; Keen et al., 2014). I used zebra finch distance call (a sexual dimorphic call) dataset to evaluate the classification performance of DFAs, LDAs, SVMs, and ANNs, where the goal was to classify the calls based on sex of the caller. I found similar classification performance across all the methods for the zebra finch distance call classification based on sex. Previous bioacoustic studies have shown that distance calls can be classified based on sex in different contexts (Elie & Theunissen, 2016, 2018). The current study validates these findings by classifying distance calls with high accuracy with a dataset constructed of calls from different sources (i.e., from the USA, Germany, and Canada). I also provide a detailed description and foundation for future classification problems especially those involving binary classification, songbird vocalizations, and multiple vocalizations from the same individual (i.e., pseudoreplication).

The current study adds to the literature of comparative bioacoustic classification. Previously, various classification techniques have been compared for bat echolocation classification based on species and genera (Armitage & Ober, 2010), mouse ultrasonic vocalizations based on sex and strain (Ivanenko et al., 2020), and wood-warblers flight call based on species (Keen et al., 2014) to name a few. These studies used DFAs, LDAs, SVMs, ANNs, and Random Forest (RF) algorithms to compare classification performance of the tasks. In zebra finches, PCA (principal component analysis), RFs, and LDAs were used to classify vocalization types (Elie & Theunissen, 2016) and for classification based on individual identity (Elie & Theunissen, 2018). In black-capped chickadees, DFAs, and ANNs were used for sex-based classification of *chick-a-dee* calls (Campbell et al., 2016) and individual identity-based classification of *fee-bee* songs (Montenegro et al., 2021). Here, I compared classification performance of DFAs, LDAs, SVMs, and ANNs using sex-based classification of zebra finch distance calls. This study validates the usage of these classification methodologies for various classification problems and with a variety of datasets.

Later in Chapter 2 goes into the relative importance of variables (in the present case, acoustic features) for the classification of calls based on sex of the caller. I compared previously used methods and extended these with more methods, not commonly used in songbird bioacoustics research, to obtain ranked acoustic features in the order of relative importance for the sex-based classification. Previous studies primarily obtained the order of relative importance through random forests' Gini index (Armitage & Ober, 2010; Elie & Theunissen, 2016). I used standardized coefficients for LDAs, weight vectors for SVMs, and Olden function for ANNs to compare the relative importance of acoustic features. This gives us the option to choose an optimal method depending on the classification technique used.

One concern in using various classification techniques has been the number of calls in the dataset. Specifically, many numbers of calls per classification category is required to perform optimally. Though the current study had a good number of calls (range: 38-45) per classification categories (female and male calls) based on previous studies (Armitage & Ober, 2010: range 20-62 calls); and classification problems based on individuals would require more calls from each individual. Here, I measured the acoustic features manually for all the calls and with large datasets this becomes a challenging task. An easier alternative would be to use automated methods for measurements with human supervision. It would be interesting to compare the classification methods used here with more challenging classification tasks as the current one involved relatively easy task of classifying sexually dimorphic female vs male distance calls. Nevertheless, the current study provides a base for future classification comparisons. Overall, I believe this study will benefit researchers, students, and everyone else who are new to this area of research by providing an introduction and comparison of techniques used for the classification of vocalizations.

4.2 Modifying a standardized go/no-go operant conditioning procedure

Refinements are often required for fine-tuning and improvement of experimental procedures. Modifying any standardized research procedure requires studies looking at effects of manipulation of experimental factors on outcomes. Operant conditioning procedures have been developed to study auditory cognition in small birds by many labs. A go/no-go operant conditioning procedure has been consistently used in our and others’ labs to study auditory cognition in chickadees and zebra finches (Congdon et al., 2020; Gentner et al., 2000; Gess et al., 2011; Hahn et al., 2015; Houx & Ten cate C, 1999; Nagel et al., 2010; Njegovan et al., 1994; Park et al., 1985; Scharff et al., 1998; Sturdy & Weisman, 2006).

In Chapter 3, I doubled the feed time duration, a crucial factor in a go/no-go operant conditioning procedure (Njegovan et al., 1994; Sturdy & Weisman, 2006) and looked at its effect on discrimination of vocalizations. Zebra finches discriminated female from male zebra finch distance calls. Birds with 2 seconds feed time duration performed similarly in terms of trials per day and trials required for discrimination as birds with 1 second feed time duration. For a particular experimental stage (Non-differential or Discrimination stage), both feed time group birds (1s and 2s access to food) performed a similar number of trials per day. It could be that birds had more time, thus reducing the stress to eat quickly, or eating twice as much as compared to 1s group birds. Direct measurements of birds' health parameters (such as mass, fat deposits, etc.) are required to determine definitively whether this manipulation had any of these effects. Learning speed, in terms of trials required to complete the discrimination, was similar for the feed time groups, which could imply doubling the feed time duration had a negligible effect on this parameter, which in this case is desired since it would allow more direct comparison of 2 second feed access with studies conducted with 1 second feed access.

It would be interesting to see how other species such as black-capped chickadees perform with new 2 second time window and how birds' performance compares with more challenging discrimination tasks. Future studies could focus more on the animals' wellbeing aspect of the experiment. Animal well-being measures before and after the experiment are required to comment on the improvement of operant procedure in accordance with improving animal welfare. Nevertheless, comparatively we could reduce potential stress to eat food quickly by increasing feed time duration.

4.3 Conclusion

In this thesis, I compared methodologies for binary classification of vocalizations to provide a baseline and framework for future bioacoustic analyses. I performed comparative analyses of classification methods, which included DFAs, SVMs, and ANNs, for a binary classification problem. I used zebra finch distance calls to conduct sex-based classification using these methods. All classification methods successfully classified distance calls according to sex with high accuracy and consistency. I also calculated and compared the variables of relative importance for measured acoustic features of distance calls which separates female and male distance calls from all methods. Many automated and semi-automated methods have been used for vocalization classification (Brooker et al., 2020; Priyadarshani et al., 2018). Comparisons of classification techniques inform us about the applicability of individual techniques, advantages & disadvantages of techniques according to data sets, significance of classifying features, and provide a framework for future researchers. In this thesis, I extend the methodological comparisons for vocalization classification with zebra finch distance calls which provides base, and framework for future bird vocalizations classification problems with giving insights about ranked features involved in distance call classification.

Modification in standardized research methods requires manipulation of experimental factors and investigating its effect on research outcome (Bitterman et al., 1983; Bogdanova et al., 2013; Klump et al., 1995). In the second part of this thesis, I investigated the effect of increase in the feed time duration during trials on discrimination performance in a go/no-go operant conditioning paradigm. Zebra finches discriminated female distance calls from male distance calls equivalently when the feed time duration was doubled, from 1 sec to 2 sec. Thus, here I modified a go/no-go operant paradigm which is widely used to test auditory perception, where now birds can receive more time to eat, potentially increasing the well-being of birds.

In sum, I provide a comparisons of classification methodologies, and show increase in feed time duration in a go/no-go operant procedure has no significant effect on discrimination performance.

Bibliography

Abi-Haidar, A., & Rocha, L. M. (2008). *Adaptive Spam Detection Inspired by a Cross-Regulation Model of Immune Dynamics: A Study of Concept Drift* (Vol. 5132, pp. 36–47). https://doi.org/10.1007/978-3-540-85072-4_4

Adret, P. (1993). Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Animal Behaviour*, *46*(1), 149–159. https://doi.org/10.1006/anbe.1993.1170

Anderson, J. T., Van Holliday, D., Kloser, R., Reid, D. G., & Simard, Y. (2008). Acoustic seabed classification: Current practice and future directions. *ICES Journal of Marine Science*, *65*(6), 1004–1011. https://doi.org/10.1093/icesjms/fsn061

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, *5*(6), 465–473. https://doi.org/10.1016/j.ecoinf.2010.08.001

Bailoo, J. D., Reichlin, T. S., & Würbel, H. (2014). Refinement of experimental design and conduct in laboratory animal research. *ILAR Journal*, *55*(3), 383–391. https://doi.org/10.1093/ilar/ilu037

Bateson, M., 2016. Optimistic and pessimistic biases: a primer for behavioural ecologists. *Current Opinion in Behavioral Sciences*, *12*, 115–121. https://doi.org/10.1016/j.cobeha.2016.09.013

Bateson, M., Matheson, S., 2007. Performance on a categorisation task suggests that removal of environmental enrichment induces “pessimism” in captive European starlings (*Sturnus vulgaris*). *Animal Welfare 16*.

Beck, M. W. (2018). NeuralNetTools: Visualization and analysis tools for neural networks. *Journal of Statistical Software*, *85*(11), 1–20. https://doi.org/10.18637/jss.v085.i11

Bee, M. A., Reichert, M. S., & 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, & M. Zuk (Eds.), *Advances in the Study of Behavior* (Vol. 48, pp. 161–249). Academic Press. https://doi.org/10.1016/bs.asb.2016.01.001

Beeman, K. (2017). *SIGNAL reference guide. – Engineering design, Berkley, CA.* [Manual].

Berkhoudt, H., van der Reijden, D., & Heijmans, M. (1987). An Automatic Pellet dispenser for precise control of feeding topography in granivorous birds. *Journal of the Experimental Analysis of Behavior*, *48*(3), 435–438. https://doi.org/10.1901/jeab.1987.48-435

Betz, N. E. (1987). Use of discriminant analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 393–403. https://doi.org/10.1037/0022-0167.34.4.393

Bitterman, M. E., Menzel, R., Fietz, A., & Schäfer, S. (1983). Classical conditioning of proboscis extension in honeybees (*Apis mellifera*). *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *97*(2), 107–119.

Boersma, P., & van Heuven, V. (2001). *Speak and unSpeak with PRAAT*. *5*(9), 7.

Bogdanova, O. V., Kanekar, S., D'Anci, K. E., & Renshaw, P. F. (2013). Factors influencing behavior in the forced swim test. *Physiology & Behavior*, *118*, 227–239. https://doi.org/10.1016/j.physbeh.2013.05.012

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152. https://doi.org/10.1145/130385.130401

Bradbury, J. W., & Vehrencamp, S. L. (2011). *Principles of animal communication* (2nd ed). Sinauer Associates.

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351–358. https://doi.org/10.1038/417351a

Branch, C. L., & Pravosudov, V. V. (2015). Mountain chickadees from different elevations sing different songs: Acoustic adaptation, temporal drift or signal of local adaptation? *Royal Society Open Science*, *2*(4), 150019. https://doi.org/10.1098/rsos.150019

Branch, C. L., & Pravosudov, V. V. (2019). Variation in song structure along an elevation gradient in a resident songbird. *Behavioral Ecology and Sociobiology*, *74*(1), 9. https://doi.org/10.1007/s00265-019-2786-5

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

Brooker, S. A., Stephens, P. A., Whittingham, M. J., & Willis, S. G. (2020). Automated detection and classification of birdsong: An ensemble approach. *Ecological Indicators*, *117*, 106609. https://doi.org/10.1016/j.ecolind.2020.106609

Brown, C., & Riede, T. (2017). *Comparative Bioacoustics: An Overview*.

Buchholz, D., & Hothersall, D. (1976). Acquisition and retention of two operant responses in redwinged blackbirds. *Bulletin of the Psychonomic Society*, *8*(3), 159–162. https://doi.org/10.3758/BF03335112

Camacho-Alpízar, A., Eckersley, T., Lambert, C. T., Balasubramanian, G., & Guillette, L. M. (2021). Learning about construction behaviour from observing an artefact: Can experience with conspecifics aid in artefact recognition? *Animal Cognition*, *24*(6), 1267–1277. https://doi.org/10.1007/s10071-021-01519-y

Campbell, K. A., Hahn, A. H., Congdon, J. V., Scully, E. N., & Sturdy, C. B. (2020). Perception of sex differences in black-capped chickadee (*Poecile atricapillus*) *chick-a-dee* calls. 17. *7*(4), 514-530. https://doi.org/10.26451/abc.07.04.04.2020

Campbell, K. A., Hahn, A. H., Congdon, J. V., & Sturdy, C. B. (2016). An investigation of sex differences in acoustic features in black-capped chickadee (*Poecile atricapillus*) *chick-a-dee* calls. *The Journal of the Acoustical Society of America*, *140*(3), 1598–1608. https://doi.org/10.1121/1.4962281

Caycedo-Rosales, P. C., Ruiz-Muñoz, J. F., & Orozco-Alzate, M. (2013). Reconocimiento automatizado de señales bioacústicas: Una revisión de métodos y aplicaciones. *Ingeniería y Ciencia*, *9*(18), 171–195.

CCAC guidelines: Animal welfare assessment. (2021). 31.

Chen, R., & Goldberg, J. H. (2020). Actor-critic reinforcement learning in the songbird. *Current Opinion in Neurobiology*, *65*, 1–9. https://doi.org/10.1016/j.conb.2020.08.005

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Chou, C.-H., Lee, C.-H., & Ni, H.-W. (2007). Bird Species Recognition by Comparing the HMMs of the Syllables. *Second International Conference on Innovative Computing, Information and Control (ICICIC 2007)*, 143–143. https://doi.org/10.1109/ICICIC.2007.199

Chou, C.-H., & 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. https://doi.org/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. https://doi.org/10.2307/1365510

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

Collobert, R., & Bengio, S. (2004). Links between perceptrons, MLPs and SVMs. *Twenty-First International Conference on Machine Learning - ICML '04*, 23. https://doi.org/10.1145/1015330.1015415

Colonna, J. G., Gama, J., & Nakamura, E. F. (2016). How to Correctly Evaluate an Automatic Bioacoustics Classification Method. In O. Luaces, J. A. Gámez, E. Barrenechea, A. Troncoso, M. Galar, H. Quintián, & E. Corchado (Eds.), *Advances in Artificial Intelligence* (Vol. 9868, pp. 37–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-44636-3_4

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

Congdon, J. V., Hahn, A. H., Campbell, K. A., Scully, E. N., Yip, D. A., Bayne, E. M., & Sturdy, C. B. (2021). Can you hear me now? The effect of signal degradation on perceived predator threat in black-capped chickadees (*Poecile atricapillus*). *Animal Cognition*, *24*(1), 193–204. https://doi.org/10.1007/s10071-020-01433-9

Congdon, J. V., Hahn, A. H., Filippi, P., Campbell, K. A., Hoang, J., Scully, E. N., Bowling, D. L., Reber, S. A., & Sturdy, C. B. (2019). Hear them roar: A comparison of black-capped chickadee (*Poecile atricapillus*) and human (*Homo sapiens*) perception of arousal in vocalizations across all classes of terrestrial vertebrates. *Journal of Comparative Psychology*, *133*(4), 520–541. https://doi.org/10.1037/com0000187

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

D'Amelio, P. B., Klumb, M., Adreani, M. N., Gahr, M. L., & ter Maat, A. (2017). Individual recognition of opposite sex vocalizations in the zebra finch. *Scientific Reports*, *7*(1), 5579. https://doi.org/10.1038/s41598-017-05982-x

Dawson, M. R. W., Charrier, I., & Sturdy, C. B. (2006). Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types. *The Journal of the Acoustical Society of America*, *119*(5), 3161–3172. https://doi.org/10.1121/1.2189028

Delgado, R. A. (2006). Sexual selection in the loud calls of male primates: Signal content and function. *International Journal of Primatology*, *27*(1), 5–25. https://doi.org/10.1007/s10764-005-9001-4

Diamond, W. J. (2001). *Practical Experiment Designs: For Engineers and Scientists*. John Wiley & Sons.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed). Wiley.

Dufour, O., Artieres, T., Glotin, H., & Giraudet, P. (2014). Clusterized Mel Filter Cepstral Coefficients and Support Vector Machines for Bird Song Identification. In *Soundscape Semiotics—Localization and Categorization*. https://www.intechopen.com/books/soundscape-semiotics-localisation-and-categorisation/clusterized-mel-filter-cepstral-coefficients-and-support-vector-machines-for-bird-song-identificatio

Elie, J. E., & 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. *Animal Cognition*, *19*(2), 285–315. https://doi.org/10.1007/s10071-015-0933-6

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

Engler, J. O., Rödder, D., Stiels, D., & Förschler, M. I. (2014). Suitable, reachable but not colonised: Seasonal niche duality in an endemic mountainous songbird. *Journal of Ornithology*, *155*(3), 657–669. https://doi.org/10.1007/s10336-014-1049-5

Erbe, C. (2016). What is animal bioacoustics? *The Journal of the Acoustical Society of America*, *139*(4), 2004–2004. https://doi.org/10.1121/1.4949882

Fagerlund, S. (2007). Bird species recognition using Support Vector Machines. *EURASIP Journal on Advances in Signal Processing*, *2007*(1), 038637. https://doi.org/10.1155/2007/38637

Ficken, M. S., Ficken, R. W., & Witkin, S. R. (1978). Vocal repertoire of the black-capped chickadee. *The Auk*, *95*(1), 34–48. https://doi.org/10.2307/4085493

Fisher, R. A. (1936a). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Fisher, R. A. (1936b). *Design of Experiments* (Vol. 1). http://tankona.free.fr/fisher1935.pdf

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

Garcia, M., & Favaro, L. (2017). Animal vocal communication: Function, structures, and production mechanisms. *Current Zoology*, *63*(4), 417–419. https://doi.org/10.1093/cz/zox040

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

Gentner, T. Q., Hulse, S. H., Bentley, G. E., & Ball, G. F. (2000). Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *Journal of Neurobiology*, *42*(1), 117–133. https://doi.org/10.1002/(sici)1097-4695(200001)42:1<117::aid-neu11>3.0.co;2-m

Gentry, K. E., Lewis, R. N., Glanz, H., Simões, P. I., Nyári, Á. S., & Reichert, M. S. (2020). Bioacoustics in cognitive research: Applications, considerations, and recommendations. *WIREs Cognitive Science*, *11*(5). https://doi.org/10.1002/wcs.1538

Gess, A., Schneider, D. M., Vyas, A., & Woolley, S. M. N. (2011). Automated auditory recognition training and testing. *Animal Behaviour*, *82*(2), 285–293. https://doi.org/10.1016/j.anbehav.2011.05.003

Goldstein, L. (n.d.). *Measuring F0 Contours using Praat*. Retrieved April 6, 2021, from https://sail.usc.edu/~lgoldste/General_Phonetics/Tone/Measurement/measurement.html

Goltstein, P. M., Reinert, S., Glas, A., Bonhoeffer, T., & Hübener, M. (2018). Food and water restriction lead to differential learning behaviors in a head-fixed two-choice visual discrimination task for mice. *PLoS ONE*, *13*(9), e0204066. https://doi.org/10.1371/journal.pone.0204066

Gómez, W. E., Isaza, C. V., & Daza, J. M. (2018). Identifying disturbed habitats: A new method from acoustic indices. *Ecological Informatics*, *45*, 16–25. https://doi.org/10.1016/j.ecoinf.2018.03.001

Gorman, B. (2018). *mltools: Machine learning tools* [Manual]. https://CRAN.R-project.org/package=mltools

Guillette, L. M., Hoeschele, M., Hahn, A. H., & Sturdy, C. B. (2013). Heterospecific discrimination of Poecile vocalizations by zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, *127*(3), 227–236. https://doi.org/10.1037/a0029992

Günther, F., & Fritsch, S. (2010). neuralnet: Training of Neural Networks. *The R Journal*, *2*(1), 30. https://doi.org/10.32614/RJ-2010-006

Hahn, A. H., Guillette, L. M., Hoeschele, M., Mennill, D. J., Otter, K. A., Grava, T., Ratcliffe, L. M., & Sturdy, C. B. (2013). Dominance and geographic information contained within black-capped chickadee (Poecile atricapillus) song. *Behaviour*, *150*(13), 1601–1622. https://doi.org/10.1163/1568539X-00003111

Hahn, A. H., Hoang, J., McMillan, N., Campbell, K., Congdon, J., & Sturdy, C. B. (2015). Biological salience influences performance and acoustic mechanisms for the discrimination of male and female songs. *Animal Behaviour*, *104*, 213–228. https://doi.org/10.1016/j.anbehav.2015.03.023

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

Haykin, S. S. (2009). *Neural networks and learning machines* (3rd ed). Prentice Hall.

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory* (pp. xix, 335). Wiley.

Henderson, E. E., Hildebrand, J. A., & Smith, M. H. (2011). Classification of behavior using vocalizations of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*). *The Journal of the Acoustical Society of America*, *130*(1), 557–567. https://doi.org/10.1121/1.3592213

Herbst, C. T., Herzel, H., Švec, J. G., Wyman, M. T., & Fitch, W. T. (2013). Visualization of system dynamics using phasegrams. *Journal of The Royal Society Interface*, *10*(85), 20130288. https://doi.org/10.1098/rsif.2013.0288

Hernandez, A. M., Perez, E. C., Mulard, H., Mathevon, N., & Vignal, C. (2016). Mate call as reward: Acoustic communication signals can acquire positive reinforcing values during adulthood in female zebra finches (Taeniopygia guttata). *Journal of Comparative Psychology*, *130*(1), 36–43. https://doi.org/10.1037/a0040027

Hienz, R. D., Sinnott, J. M., & Sachs, M. B. (1977). Auditory sensitivity of the redwing Blackbird (*Agelaius phoeniceus*) and brown-headed cowbird (*Molothrus ater*). *Journal of Comparative and Physiological Psychology*, *91*(6), 1365–1376. https://doi.org/10.1037/h0077403

Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H., & Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Animal Cognition*, *16*(4), 599–609. https://doi.org/10.1007/s10071-013-0597-z

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

Houx, B.B, & Ten cate C. (1999). Song learning from playback in zebra finches: Is there an effect of operant contingency? *Animal Behaviour*, *57*(4), 837–845. https://doi.org/10.1006/anbe.1998.1046

Hulse, S. H., & Cynx, J. (1986). Interval and contour in serial pitch perception by a passerine bird, the European starling (*Sturnus vulgaris*). *Journal of Comparative Psychology*, *100*(3), 215–228. https://doi.org/10.1037/0735-7036.100.3.215

Ibrahim, O. M. (2013). A comparison of methods for assessing the relative importance of input variables in artificial neural networks. *Journal of Applied Sciences Research*, *9*(11), 5692–5700.

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

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

Jaiswara, R., Nandi, D., & 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. https://doi.org/10.1371/journal.pone.0075930

Jakkula, V. (2011). *Tutorial on Support Vector Machine (SVM)*. *13*. 1-13. https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf

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

Johnson, P. D., & Besselsen, D. G. (2002). Practical aspects of experimental design in animal research. *ILAR Journal*, *43*(4), 202–206. https://doi.org/10.1093/ilar.43.4.202

Keen, S., Ross, J. C., Griffiths, E. T., Lanzone, M., & Farnsworth, A. (2014). A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). *Ecological Informatics*, *21*, 25–33. https://doi.org/10.1016/j.ecoinf.2014.01.001

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., … Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus: Acoustic sequences in animals. *Biological Reviews*, *91*(1), 13–52. https://doi.org/10.1111/brv.12160

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

Kim, E. W., Phillips, B. U., Heath, C. J., Cho, S. Y., Kim, H., Sreedharan, J., Song, H.-T., Lee, J. E., Bussey, T. J., Kim, C. H., Kim, E., & Saksida, L. M. (2017). Optimizing reproducibility of operant testing through reinforcer standardization: Identification of key nutritional constituents determining reward strength in touchscreens. *Molecular Brain*, *10*(1), 31. https://doi.org/10.1186/s13041-017-0312-0

Kirk, R. E. (2012). *Experimental Design: Procedures for the Behavioral Sciences*. SAGE Publications.

Klump, G. M., Dooling, R. J., Fay, R. R., & Stebbins, W. C. (Eds.). (1995). *Methods in Comparative Psychoacoustics*. Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-7463-2

Knight, E., Hannah, K., Foley, G., Scott, C., Brigham, R., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology*, *12*(2). https://doi.org/10.5751/ACE-01114-120214

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

Laiolo, P., Claudia, P., & Antonio, R. (2000). A study of Choughs' vocal repertoire: Variability related to individuals, sexes and ages. *Journal Für Ornithologie*, *141*(2), 168. https://doi.org/10.1046/j.1439-0361.2000.00074.x

Laiolo, P., Vögeli, M., Serrano, D., & Tella, J. L. (2008). Song Diversity Predicts the Viability of Fragmented Bird Populations. *PLOS ONE*, *3*(3), e1822. https://doi.org/10.1371/journal.pone.0001822

Lambert, C., Sahu, P.K., Sturdy, C.B., Guillette, L.M., 2022. Among-individual differences in auditory and physical cognitive abilities in zebra finches. *Learning & Behavior*. https://doi.org/10.3758/s13420-022-00520-w

Larsson, N. O. (2001). A Design View on Research in Social Sciences. *Systemic Practice and Action Research*, *14*(4), 383–405. https://doi.org/10.1023/A:1011399918214

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

Lohr, B., Dooling, R. J., & Bartone, S. (2006). The discrimination of temporal fine structure in call-like harmonic sounds by birds. *Journal of Comparative Psychology*, *120*(3), 239–251. https://doi.org/10.1037/0735-7036.120.3.239

Lohr, B., Wright, T. F., & Dooling, R. J. (2003). Detection and discrimination of natural calls in masking noise by birds: Estimating the active space of a signal. *Animal Behaviour*, *65*(4), 763–777. https://doi.org/10.1006/anbe.2003.2093

Mair, P., & Wilcox, R. (2020). *Robust Statistical Methods Using WRS2*. 41.

Manly, B. F. J., & Navarro Alberto, J. A. (2020). *Randomization, bootstrap and monte carlo methods in biology* (Fourth edition). CRC Press / Taylor & Francis Group.

Manteuffel, G., Puppe, B., & Schön, P. C. (2004). Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science*, *88*(1), 163–182. https://doi.org/10.1016/j.applanim.2004.02.012

Matsumoto, Y., Menzel, R., Sandoz, J.-C., & Giurfa, M. (2012). Revisiting olfactory classical conditioning of the proboscis extension response in honey bees: A step toward standardized procedures. *Journal of Neuroscience Methods*, *211*(1), 159–167. https://doi.org/10.1016/j.jneumeth.2012.08.018

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

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Mcloughlin, M. P., Stewart, R., & McElligott, A. G. (2019). Automated bioacoustics: Methods in ecology and conservation and their potential for animal welfare monitoring. *Journal of The Royal Society Interface*, *16*(155), 20190225. https://doi.org/10.1098/rsif.2019.0225

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

Montenegro, C., Sahu, P. K., & Sturdy, C. B. (2021). Individual acoustic differences in female black-capped chickadee (Poecile atricapillus) fee-bee songs. *The Journal of the Acoustical Society of America*, *150*(4), 3038–3046. https://doi.org/10.1121/10.0006532

Montenegro, C., Service, W. D., Scully, E. N., Mischler, S. K., Sahu, P. K., Benowicz, T. J., Fox, K. V. R., & Sturdy, C. B. (2021). The impact of anthropogenic noise on individual identification via female song in Black-capped chickadees (Poecile atricapillus). *Scientific Reports*, *11*(1), 17530. https://doi.org/10.1038/s41598-021-96504-3

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

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

Myers, J. L. (1979). *Fundamentals of Experimental Design*. Allyn and Bacon.

Nagel, K. I., McLendon, H. M., & Doupe, A. J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of Neurophysiology*, *104*(3), 1426–1437. https://doi.org/10.1152/jn.00028.2010

Njegovan, M., Hilhorst, B., Ferguson, S., & Weisman, R. (1994). A motor-driven feeder for operant training in song birds. *Behavior Research Methods, Instruments & Computers*, *26*(1), 26–27. https://doi.org/10.3758/BF03204558

Njegovan, M., & Weisman, R. (1997). Pitch discrimination in field- and isolation-reared black-capped chickadees (Parus atricapillus). *Journal of Comparative Psychology*, *111*(3), 294–301. https://doi.org/10.1037/0735-7036.111.3.294

Nowicki, S., & 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

Okanoya, K., & Dooling, R. J. (1987). Hearing in passerine and psittacine birds: A comparative study of absolute and masked auditory thresholds. *Journal of Comparative Psychology*, *101*(1), 7–15. https://doi.org/10.1037/0735-7036.101.1.7

Olden, J. D., & Jackson, D. A. (2002). Illuminating the ‘“black box”’: A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 16.

Park, T. J., & Dooling, R. J. (1991). Sound localization in small birds: Absolute localization in azimuth. *Journal of Comparative Psychology*, *105*(2), 125–133. https://doi.org/10.1037/0735-7036.105.2.125

Park, T., Okanoya, K., & Dooling, R. (1985). Operant conditioning of small birds for acoustic discrimination. *Journal of Ethology*, *3*(1), 5–9. https://doi.org/10.1007/BF02348160

Penar, W., Magiera, A., & Klocek, C. (2020). Applications of bioacoustics in animal ecology. *Ecological Complexity*, *43*, 100847. https://doi.org/10.1016/j.ecocom.2020.100847

Phillips, B. U., Heath, C. J., Ossowska, Z., Bussey, T. J., & Saksida, L. M. (2017). Optimisation of cognitive performance in rodent operant (touchscreen) testing: Evaluation and effects of reinforcer strength. *Learning & Behavior*, *45*(3), 252–262. https://doi.org/10.3758/s13420-017-0260-7

Piczak, K. J. (2016). Recognizing bird species in audio recordings using deep convolutional neural networks. *BirdCLEF*, 10.

Prior, N. H., Smith, E., Lawson, S., Ball, G. F., & Dooling, R. J. (2018). Acoustic fine structure may encode biologically relevant information for zebra finches. *Scientific Reports*, *8*(1), 6212. https://doi.org/10.1038/s41598-018-24307-0

Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: A review. *Journal of Avian Biology*, *49*(5), jav-01447. https://doi.org/10.1111/jav.01447

Průchová, A., Jaška, P., & Linhart, P. (2017). Cues to individual identity in songs of songbirds: Testing general song characteristics in Chiffchaffs Phylloscopus collybita. *Journal of Ornithology*, *158*(4), 911–924. https://doi.org/10.1007/s10336-017-1455-6

R Core Team. (2019). *R: A language and environment for statistical computing* [Manual]. https://www.R-project.org/

Richter, S. H., Garner, J. P., & Würbel, H. (2009). Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nature Methods*, *6*(4), 257–261. https://doi.org/10.1038/nmeth.1312

Robakis, E., Watsa, M., & Erkenswick, G. (2018). Classification of producer characteristics in primate long calls using neural networks. *The Journal of the Acoustical Society of America*, *144*(1), 344–353. https://doi.org/10.1121/1.5046526

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Röttgen, V., Schön, P. C., Becker, F., Tuchscherer, A., Wrenzycki, C., Düpjan, S., & 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. Scopus. https://doi.org/10.1017/S1751731119001733

Ryan, M. J. (1990). Signals, Species, and Sexual Selection. *American Scientist*, *78*(1), 46–52.

Ryan, T. P., & Morgan, J. P. (2007). Modern Experimental Design. *Journal of Statistical Theory and Practice*, *1*(3–4), 501–506. https://doi.org/10.1080/15598608.2007.10411855

Salamon, J., Bello, J. P., Farnsworth, A., Robbins, M., Keen, S., Klinck, H., & Kelling, S. (2016). Towards the Automatic Classification of Avian Flight Calls for Bioacoustic Monitoring. *PLOS ONE*, *11*(11), e0166866. https://doi.org/10.1371/journal.pone.0166866

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

Scharff, C., Nottebohm, F., & Cynx, J. (1998). Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *Journal of Neurobiology*, *36*(1), 81–90.

Scheiner, R., Page, R., & Erber, J. (2004). Sucrose responsiveness and behavioral plasticity in honey bees (Apis mellifera). *Http://Dx.Doi.Org/10.1051/Apido:2004001*, *35*. https://doi.org/10.1051/apido:2004001

Schön, P., Puppe, B., & Manteuffel, G. (2004). Automated recording of stress vocalisations as a tool to document impaired welfare in pigs. *Animal Welfare*, 6.

Scully, E. N., Campbell, K. A., Congdon, J. V., & Sturdy, C. B. (2020). Discrimination of black-capped chickadee (*Poecile atricapillus*) chick-a-dee calls produced across seasons. *Animal Behavior and Cognition*, *7*(2), 247–256. https://doi.org/10.26451/abc.07.02.14.2020

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

Shy, E., McGregor, P. K., & Krebs, J. R. (1986). Discrimination of song types by male great tits. *Behavioural Processes*, *13*(1–2), 1–12. https://doi.org/10.1016/0376-6357(86)90012-4

Slocombe, K. E., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social Role. *Journal of Comparative Psychology*, *119*(1), 67–77. https://doi.org/10.1037/0735-7036.119.1.67

Sproull, N. L. (2002). *Handbook of Research Methods: A Guide for Practitioners and Students in the Social Sciences*. Scarecrow Press.

Stebbins, W. C. (Ed.). (1970). *Animal Psychophysics: The design and conduct of sensory experiments*. Springer US. https://doi.org/10.1007/978-1-4757-4514-6

Stebbins, W. C., Mead, P. B., & Martin, J. M. (1959). The relation of amount of reinforcement to performance under a fixed-interval schedule. *Journal of the Experimental Analysis of Behavior*, *2*(4), 351–355. https://doi.org/10.1901/jeab.1959.2-351

Sturdy, C. B., Phillmore, L. S., Price, J. L., & Weisman, R. G. (1999). Song-note discriminations in zebra finches (*Taeniopygia guttata*): Categories and pseudocategories. *Journal of Comparative Psychology*, *113*(2), 204–212. https://doi.org/10.1037/0735-7036.113.2.204

Sturdy, C. B., Phillmore, L. S., Sartor, J. J., & Weisman, R. G. (2001). Reduced social contact causes auditory perceptual deficits in zebra finches, *Taeniopygia guttata*. *Animal Behaviour*, *62*(6), 1207–1218. https://doi.org/10.1006/anbe.2001.1864

Sturdy, C. B., & Weisman, R. G. (2006). Rationale and methodology for testing auditory cognition in songbirds. *Behavioural Processes*, *72*(3), 265–272. https://doi.org/10.1016/j.beproc.2006.03.007

Tachibana, R. O., Oosugi, N., & Okanoya, K. (2014). Semi-automatic classification of birdsong elements using a linear Support Vector Machine. *PLoS ONE*, *9*(3), e92584. https://doi.org/10.1371/journal.pone.0092584

Tallet, C., Linhart, P., Policht, R., Hammerschmidt, K., Šimeček, P., Kratinova, P., & 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. https://doi.org/10.1371/journal.pone.0071841

Teixeira, D., Maron, M., & van Rensburg, B. J. (2019). Bioacoustic monitoring of animal vocal behavior for conservation. *Conservation Science and Practice*, *1*(8), e72. https://doi.org/10.1111/csp2.72

Tooze, Z. J., Harrington, F. H., & Fentress, J. C. (1990). Individually distinct vocalizations in timber wolves, Canis lupus. *Animal Behaviour*, *40*(4), 723–730. https://doi.org/10.1016/S0003-3472(05)80701-8

Trosclair-Lasserre, N. M., Lerman, D. C., Call, N. A., Addison, L. R., & Kodak, T. (2008). Reinforcement magnitude: An evaluation of preference and reinforcer efficacy. *Journal of Applied Behavior Analysis*, *41*(2), 203–220. https://doi.org/10.1901/jaba.2008.41-203

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]. https://doi.org/10.33915/etd.574

Tung, H.-Y., Huang, D.-A., Xie, X.-F., Zhou, Y., & Russino, J. (2003). *Identification of songbird species in field recordings*. Carnegie Mellon University. ML-10701 Project. http://www.wiomax.com/team/xie/paper/CMU10701.pdf

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005

Varnon, C. A., Lang, H., & Abramson, C. I. (2018). Automated research in comparative psychology: Limitations and new directions. *International Journal of Comparative Psychology*, *31*(0). https://escholarship.org/uc/item/8gf4d5tx

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

Vicario, D. S., Naqvi, N. H., & Raksin, J. N. (2001). Sex differences in discrimination of vocal communication signals in a songbird. *Animal Behaviour*, *61*(4), 805–817. https://doi.org/10.1006/anbe.2000.1651

Vignal, C., & Mathevon, N. (2011). Effect of acoustic cue modifications on evoked vocal response to calls in zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, *125*(2), 150–161. https://doi.org/10.1037/a0020865

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

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

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires—Hard vs. Soft classification approaches. *PloS One*, *10*(4), e0125785. https://doi.org/10.1371/journal.pone.0125785

Walcott, C., Mager, J. N., & Piper, W. (2006). Changing territories, changing tunes: Male loons, Gavia immer, change their vocalizations when they change territories. *Animal Behaviour*, *71*(3), 673–683. https://doi.org/10.1016/j.anbehav.2005.07.011

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

Weber, D. C., & Skillings, J. H. (2017). *A First Course in the Design of Experiments: A Linear Models Approach*. Routledge. https://doi.org/10.1201/9781315136318

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

Weisman, R. G., Njegovan, M. G., Williams, M. T., Cohen, J. S., & Sturdy, C. B. (2004). A behavior analysis of absolute pitch: Sex, experience, and species. *Behavioural Processes*, *66*(3), 289–307. https://doi.org/10.1016/j.beproc.2004.03.010

Wellock, C. D., & Reeke, G. N. (2012). Quantitative tools for examining the vocalizations of juvenile songbirds. *Computational Intelligence and Neuroscience*, *2012*, e261010. https://doi.org/10.1155/2012/261010

Würbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics*, *26*(3), 263–263. https://doi.org/10.1038/81541

Xu, M., & 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. https://doi.org/10.1534/genetics.118.301706

Zann, R. A. (1996). *The zebra finch: A synthesis of field and laboratory studies*. Oxford University Press.

Zhang, Y., Wang, S., Phillips, P., & Ji, G. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowledge-Based Systems*, *64*, 22–31. https://doi.org/10.1016/j.knosys.2014.03.015