

RESEARCH ARTICLE

Comparing context-dependent call sequences employing machine learning methods: an indication of syntactic structure of greater horseshoe bats

Kangkang Zhang¹, Tong Liu¹, Muxun Liu¹, Aoqiang Li¹, Yanhong Xiao¹, Walter Metzner² and Ying Liu^{1,*}

ABSTRACT

For analysis of vocal syntax, accurate classification of call sequence structures in different behavioural contexts is essential. However, an effective, intelligent program for classifying call sequences from numerous recorded sound files is still lacking. Here, we employed three machine learning algorithms (logistic regression, support vector machine and decision trees) to classify call sequences of social vocalizations of greater horseshoe bats (*Rhinolophus ferrumequinum*) in aggressive and distress contexts. The three machine learning algorithms obtained highly accurate classification rates (logistic regression 98%, support vector machine 97% and decision trees 96%). The algorithms also extracted three of the most important features for the classification: the transition between two adjacent syllables, the probability of occurrences of syllables in each position of a sequence, and the characteristics of a sequence. The results of statistical analysis also supported the classification of the algorithms. The study provides the first efficient method for data mining of call sequences and the possibility of linguistic parameters in animal communication. It suggests the presence of song-like syntax in the social vocalizations emitted within a non-breeding context in a bat species.

KEY WORDS: Aggressive call, Animal communication, Distress call, *Rhinolophus ferrumequinum*, Syntax

INTRODUCTION

Understanding how human language evolved from earlier forms of animal communication is the key to fully appreciating its unique capacity in human communication (Scarantino and Clay, 2015). One of the most powerful features of human language is grammatical rules such as word ordering under various backgrounds (Fitch, 2010). Multisyllabic vocalizations with syntax (the set of rules for combining words into phrases as in human sentences) have been found in several species of birds and mammals (Collier et al., 2014; Schlenker et al., 2016). Furthermore, social context has been shown to be essential for song structure and note use in songbirds (Catchpole and Slater, 2003; Byers and Kroodsma, 2009). In mammals, although the ability to construct calls/songs from vocal units has been reported in various species (Clarke et al., 2006; Bohn et al., 2009; Fedurek and

Slocombe, 2011; Green et al., 2011; Filatova et al., 2012; Kershenbaum et al., 2012), evidence of altering vocal composition and structure owing to social context has been reported in only a few species (Bohn et al., 2008; Candiotti et al., 2012; Cäsar et al., 2013; Chabout et al., 2015). Exploring the ways in which vocal elements are ordered and combined under various social cues or behavioural contexts in mammalian species is critical to advancing our understanding of the function and evolutionary origin of syntax.


Previous research described and compared animal syntax based on various sound features, including the spectral and temporal parameters of syllables, numbers and types of syllables, probability of different syllable types occurring, temporal emission patterns and probability of syllable transitions (Briefer et al., 2013; Deslandes et al., 2014; Chabout et al., 2015; Lin et al., 2016a). These features are based on different subunits such as syllables, phrases and bouts of animal vocalizations, and they can reveal the call structure and the rules of syllable ordering in animal vocalizations. However, most studies have used components of these features in one research paradigm (Gadziola et al., 2012; Bohn et al., 2013; Suzuki et al., 2018). For example, Chabout et al. (2015) mainly used the probability of occurrence of a transition type between syllables to compare male mice song syntax depending on social contexts, while Cäsar et al. (2013) compared the types of call sequences between study groups to state that Titi monkey calls varied with predator location and type. Moreover, analyses that are conducted at the level of sequences, such as Markov chains and Zipf's law, have usually only adopted call sequences of sufficient length (Berwick et al., 2011; Deslandes et al., 2014), which might cause information to be neglected owing to the different lengths of sequences. In addition, limited work has been conducted for exploring the contribution of different features for syntactic classification. Therefore, before further analysis of the context-dependent syntax in animal vocalization can proceed, it would be efficient if a method could be constructed to classify the call sequences in different contexts by integrating the features of different subunit levels and evaluating the contribution of each feature to the classification from a large quantity of sound recording data.

Machine learning is a subset of artificial intelligence in the field of computer science, often using statistical techniques to give computers the ability to 'learn' from data (Michalski et al., 2013). In recent years, machine learning has been shown to be a powerful tool and has been widely applied to make predictions and discover hidden structure within large datasets, as well as to deal with classification problems in many circumstances (Skowronski and Harris, 2006; Acevedo et al., 2009; Stathopoulos et al., 2018). In bioacoustics, numerous studies have applied various machine learning tools for species recognition in environmental monitoring (Huang et al., 2009; Walters et al., 2012; Shamir et al., 2014; Aodha et al., 2018) and automatic classification of

¹Jilin Provincial Key Laboratory of Animal Resource Conservation and Utilization, No. 2555, Street Jingyue, Northeast Normal University, Changchun 130117, China.

²Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA 90095, USA.

*Author for correspondence (liuy252@nenu.edu.cn)

 Y.L., 0000-0003-3359-0716

animal vocalizations (Ranjard and Ross, 2008; Armitage and Ober, 2010; Pozzi et al., 2010; Turesson et al., 2016). The successful application of machine learning tools indicates their importance for constructing a transparent, fast and accurate algorithm for animal acoustic signals analysis.

Bats have a suite of features that indicate a neural substrate supporting vocal plasticity and complexity, such as neural adaptations to support laryngeal echolocation (Siemers et al., 2011; Fenton et al., 2012), vocal learning (Knörnschild, 2014) and geographical divergence of vocalizations (Lin et al., 2015; Prat et al., 2017). Furthermore, some species have been documented to possess the capacity to vary social calls in response to social cues and behavioural context. For example, Brazilian free-tailed bats, *Tadarida brasiliensis*, quickly varied song composition to meet the specific demands of different social functions (Bohn et al., 2013). Big brown bats, *Eptesicus fuscus*, emitted bouts of vocalizations that could be assigned to specific aggressive behaviours (Gadziola et al., 2012). Mexican free-tailed bats (also known as Brazilian free-tailed bats) produced fixed vocal compositions, including irritation calls, protest calls and warning calls during agonistic interactions (Bohn et al., 2008). The complexity and plasticity of acoustic communication systems observed in bats make them an important template for studies of acoustic communication. However, to date, syntactic structures of only songs and aggressive calls have been reported in bats. It remains unclear whether syntax exists in other social contexts of other bat species and whether the syntax is the same or differs between different contexts.

Aggressive encounters and distress calls occur in a wide range of animal groups. Aggressive encounters occur when the individuals compete for limited resources such as mates, food, shelter or territories (Bradbury and Vehrencamp, 2011). Distress calls as a categorical alarm signal are usually produced by vertebrates when cornered, attacked or captured by a predator (Magrath et al., 2015). In echolocating bats that have weak vision and that normally live in dark environments, acoustic signals play a primary role in information exchange in the two behavioural contexts (Gillam and Fenton, 2016). As mentioned above, syntax of aggressive calls in Mexican free-tailed bats and big brown bats has been reported (Bohn et al., 2008; Gadziola et al., 2012), but it is still unclear in other species. For distress calls, previous studies have revealed both interspecific and conspecific acoustic similarity of the call structures, and the bats could recognize and respond to the acoustic similarity (Russ et al., 2004; Eckenweber and Knörnschild, 2016; Huang et al., 2018). These findings suggested the potential of existing rules in which the vocal units were ordered and combined, that is, existing syntax in a distress context.

Greater horseshoe bats (*Rhinolophus ferrumequinum*) have a large vocal repertoire (Ma et al., 2006; Jiang et al., 2017). The bats roost in the tens to hundreds and have all-female or mixed-sex colonies that fluctuate in size across seasons. Previous studies revealed a broad diversity of the vocalizations in this species that may reflect the existing rules for syllable ordering and combining responses to different social and behavioural contexts (Jones and Siemers, 2011; Luo et al., 2013; Lin et al., 2016b). Therefore, we recorded the social calls of greater horseshoe bats in aggressive and distress contexts as study templates. Our aim was to employ machine learning methods to classify the call sequences in different behavioural contexts by integrating the sound features of all subunit levels (syllables, transitions and sequences) and extracting the features that play large roles in the classification. The analysis may provide a fast and efficient path to mining useful information from

large amounts of vocalizational data, so that the experimenter can select the important features for further analysis.

MATERIALS AND METHODS

Animals

In May 2016, we captured eight adult *Rhinolophus ferrumequinum* (Schreber 1774) (4 males, 4 females) with mist nets from Dalazi Cave in Zhi'an Village, Jilin, Peoples Republic of China. Bats were housed in a laboratory with regulated temperature (20–25°C), humidity (50%–70%) and light:dark cycles (natural photoperiod in Changchun). Experimental bats had free access to sufficient mealworms and fresh water in dishes every day. All experimental procedures complied with the ABS/ASAB guidelines for the Use of Animals in Research and were approved by the Committee on the Use and Care of Animals at the Northeast Normal University (approval number: NENU-W-2010–101). All bats were released into their roosts after the experiment.

Acoustic and behavioural recording

Distress calls

Bats often emit echolocation calls and distress calls when they are captured by predators or experimenters. Distress calls are usually recorded from handheld bats, as predation events of bats are very rare (Russ et al., 2004; Luo et al., 2013; Lin et al., 2015; Huang et al., 2018). Therefore, we recorded distress calls from the handheld bats. Distress calls were recorded using an UltrasoundGate 116 (Avisoft Bioacoustics, Berlin, Germany) connected to a laptop computer (at a sampling rate of 250 kHz at 16 bits per sample). The condenser microphone, with a flat frequency response between 10 Hz and 200 kHz (± 3 dB), was set on a small tripod 1 m from the hand-held bat. During recording, each bat was held gently and its lower back was gently massaged by the researcher. In this case, a 4 min sound file of distress calls was produced for every individual.

Aggressive calls

After distress call recording, each bat was marked with 4.2-mm numbered aluminium alloy band (Porzana Ltd, East Sussex, UK). Recent studies in our laboratory (Jiang et al., 2017; Sun et al., 2019) have confirmed that the bands do not change the normal behaviour of the bats. For each recording trial, four individuals (one male and three females) were randomly selected and housed in a small cage (33×33×25 cm). To obtain natural calls, we adopted the natural paradigm (Gadziola et al., 2012) in which bats were undisturbed and recorded for several hours during their active period. Each recording trial lasted from 17:00 to 06:00 h on the next day, and data were recorded via the UltrasoundGate system at a sampling rate of 250 kHz at 16 bits per sample. Synchronized videos were filmed via an infrared digital video camera (Sony HDR PJ760E). Bats typically produce aggressive calls when one individual (intruder) disturbs another (resident) as it might jostle for a roost position within the group (Zhao et al., 2018, 2019; Sun et al., 2019), but the aggressive call structure varies with the degree of agonistic encounter (Gadziola et al., 2012). To obtain the aggressive calls under relatively the same aggression degree, we only used the call sequences produced during physical conflict for analysis. During the physical conflict, only the resident that was disturbed vocalized, while the intruder or other undisturbed residents did not produce any sounds. Thus, vocalizations recorded during aggressive contexts rarely contained overlapping signals from multiple animals. In addition, the caller could be visually identified in the video, because it would open its mouth when producing aggressive calls. The

recording trials were repeated until no new syllable types were found in the call sequences.

Terminology of call sequences

We analysed the recorded call sequences using Avisoft SASLab Pro (Avisoft Bioacoustics, Berlin, Germany). The social vocalizations were described and classified following the nomenclature given by Kanwal et al. (1994) and followed by others, e.g. Ma et al. (2006) and Gadziola et al. (2012). Simple syllable types were named according to call structure, with a prefix denoting secondary spectral features [e.g. SFM, sinusoidal FM; BNB, broadband noise burst (NB); DFM, downward FM; UFM, upward FM] and a suffix denoting secondary temporal features (e.g. BNBI, long BNB; BNBs, short BNB). The composite syllable types were named according to the empirically established combination of simple syllables and abbreviated accordingly (e.g. NB-DFM, noise burst-DFM). The composited syllables with more than two components were named according to the combination of the first letter of names of each syllable (e.g. NSND for NB-SFM-NB-DFM).

Because the purpose of the present study was to compare the sequence structures, we used each call sequence as an analysis unit. During distress contexts, the held bats emitted distress call sequences with multiple syllables separated by inter-syllable intervals. A distress call sequence was determined if it was separated from other calls by intervals exceeding four times the average inter-syllable intervals within this sequence (Jiang et al., 2017).

During the physical contact in aggressive contexts, the disturbed residents usually fought back with wing flapping or boxing moves. We determined a separate agonistic behaviour when it started from the first wing flap or boxing move given by any disturbed individual and ended when every individual had calmed down. A sequence of multiple syllables emitted by the disturbed resident during each separate agonistic behaviour was defined as an aggressive call sequence for the following analyses.

Feature selection

All sequences recorded in aggressive and distress contexts were used for machine learning classification models. We extracted 12 features used frequently in bat acoustic analysis from each sequence using a self-written Python program. These features were as follows: (a) total number of syllable types occurring in a sequence; (b) total number of syllables of all types in a sequence; (c) total number of transitions (transitions between two adjacent syllables) types in a sequence; (d) a/c , addressing the linearity of the way syllables are ordered in a sequence (Scharff and Nottebohm, 1991); (e) $c/\text{total number of transition types under the behaviour context}$, expressing the consistency of the occurrence frequency of one transition type (Scharff and Nottebohm, 1991); (f) entropy, calculated with

$$E = -\sum_{i=1}^n P_i (-\log_2 P_i), \text{ where } P_i \text{ is the probability of occurrence of}$$

the i th syllable type, and n is the number of syllable types; (g) product of the probabilities of each syllable occurring in a certain position of the sequence; (h) the product of probabilities of each transition occurring in the current context; (i) a/b , representing the versatility of a sequence; (j) uncertainty of transitions, calculated as

$$H(x) = -\sum_{i=1}^n P_i \log_2(P_i), \text{ where } H(x) \text{ measures a given syllable } x \text{ to}$$

the rest of the n syllables that follow, and P_i represents the probability of the transition from x to i (Hailman et al., 1985); (k) gender of bats that emitted the sequence; and (l) marker label of the bats, representing the bat individual. Features a and b were related

to syllables; features c, e, h and j describe transitions in one sequence; features d, f and i describe characteristics of sequence structure; feature g describes positions of syllables occurring in a sequence; and features k and l concern individuals' information. The class names representing aggressive and distress contexts were labelled with integers (0, 1) as training targets.

Nominal features such as gender and marker label were converted to new dummy variables via the one-hot encoding technique to avoid technical glitches. Because many linear models, such as the logistic regression and SVM, initialize the weights to 0 or small random values close to 0, standardization was used to center the feature columns at a mean of 0 with a standard deviation of 1 so that the feature columns take the form of a normal distribution, which makes it easier to learn the weights.

Classification algorithms

All the classification algorithms were implemented with Scikit-learn (version 0.19.1), a Python module integrating classical machine learning algorithms (Pedregosa et al., 2011; Raschka and Mirjalili, 2017).

Logistic regression

The logistic regression classifier is a powerful and widely used algorithm for linear and binary classification problems. As the name suggests, in the logistic regression model, the weighted input features are fed to the logistic function. Then, the probability of each sequence belonging to a given context was calculated with the odds ratio. Finally, the predicted probability can then be converted into a binary outcome. The logistic regression model has been used for research on habitat selection of animal populations (Prugh et al., 2008; Duchesne et al., 2010) and other scenarios of data mining. Here, we trained the classifier with the regularization parameter $C=1000.0$ and regularized using the L2 norm of the classifier weights. We used the implementation in `sklearn.linear_model.LogisticRegression`.

Support vector machine (SVM)

The SVM is a sophisticated kernel-based machine learning classifier and has attracted much attention as a new classification technique with good generalization ability (Cristianini and Shawe-Taylor, 2000). SVMs have been widely applied to species recognition and acoustical classification in animals (Fagerlund, 2007; Chen et al., 2012). In SVMs, the optimization objective is to maximize the margin, the distance between the separating hyperplane (each hyperplane representing the call sequences of each context) and the training samples that are closest to this hyperplane, the so-called support vectors. The SVM induction process aims to establish an optimal discriminative function between two classes of call sequences in two contexts while accomplishing the trade-off between generalization and overfitting. Here, we used the `sklearn.svm.SVC` method with `kernel='linear'` and hyperparameter $C=1.0$.

Decision trees

Decision trees are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. Using the decision algorithm, we started at the tree root and split the data on the feature that results in the largest information gain. Feature scaling (standardization) is not a requirement for decision tree algorithms. Decision trees are readily interpretable, directly used to generate rules, and computationally inexpensive to train, evaluate and store

Table 1. Overview of sequences under aggressive and distress contexts

Context	Number of bats	Number of syllables	Number of syllable types	Number of transition types	Number of sequences	Average length of sequences
Aggressive	8	4692	30	156	2014	3 (2.3)
Threatening	8	2552	19	122	665	4 (3.8)

The average length of sequences in the aggressive context was 2.3 and in the distress context was 3.8. The value was rounded up to an integer.

(Valletta et al., 2017). Here, we used DecisionTreeClassifier in sklearn.tree to train a decision tree with criterion='entropy' and max_depth=5.

After the model construction and feature selection, all 12 features of each call sequence in two contexts were standardized by the StandardScaler function in the scikit-learn package and collected as a dataset. Then the dataset was randomly partitioned into a separate test dataset (804 samples) and a training dataset (1875 samples) with the ratio of 3:7. The three models for sequence classification were trained with the training set and tested with the test set. Hyperparameters specific to each machine learning algorithm were

tuned by cross-validation to strike a balance between underfitting and overfitting.

The performance of the machine learning models was characterized with area under the curve in receiver operator characteristic graphs, a useful tool for estimating model performance with respect to the false positive and true positive rates (Raschka and Mirjalili, 2017).

Feature importance estimation

The contribution of each feature to the classification of call sequences between contexts was measured as the average impurity

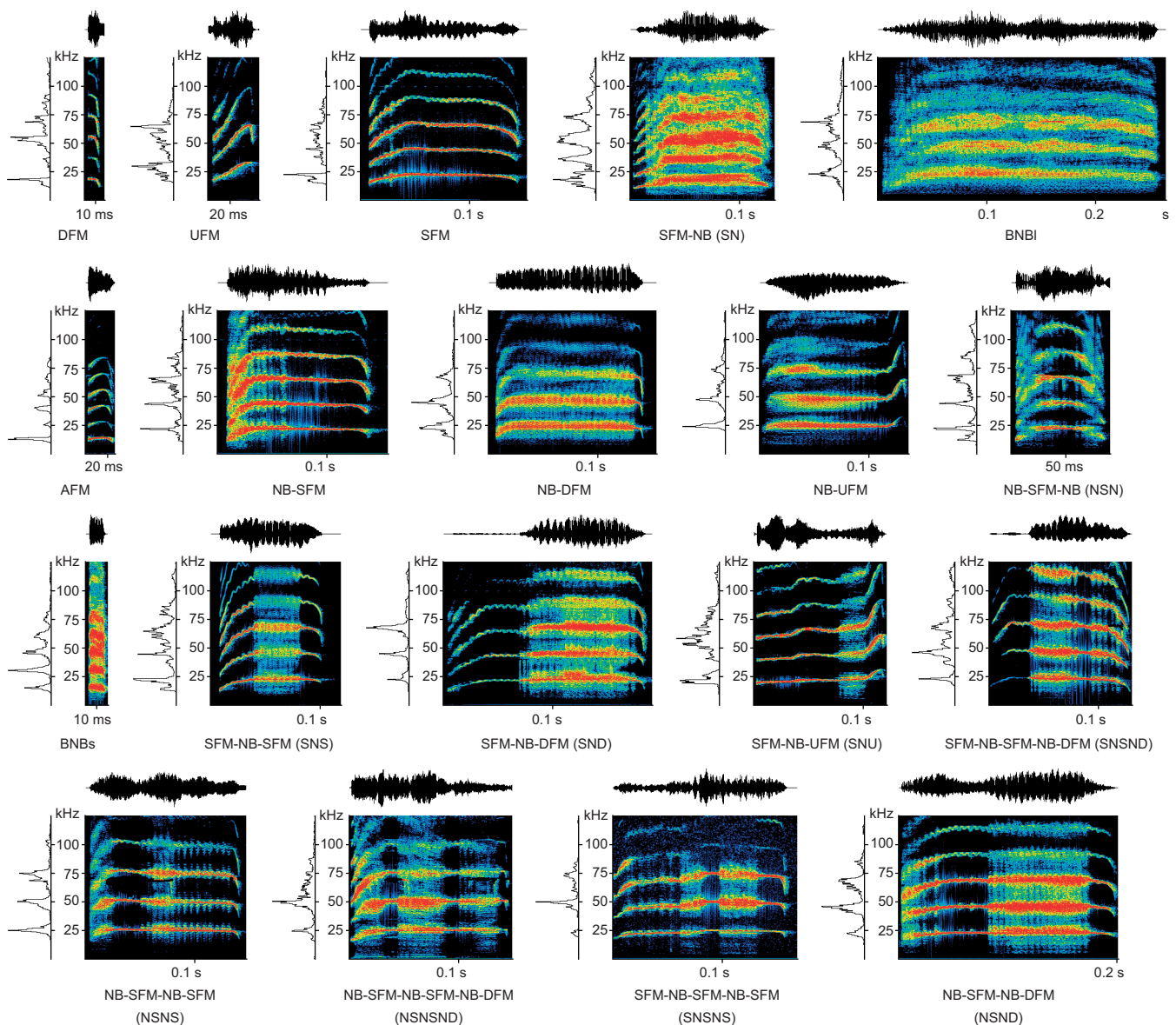


Fig. 1. Spectrograms of syllable types investigated. Only syllables that constitute more than 1% of all vocalizations were plotted.

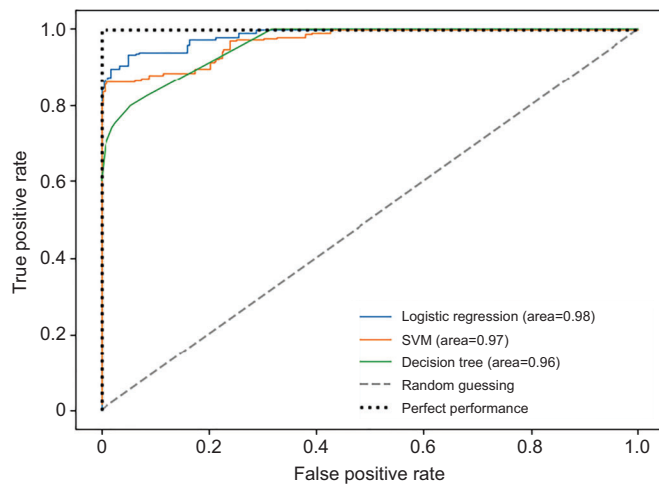


Fig. 2. Receiver operator characteristic (ROC) graphs of three models. The resulting ROC curve indicates that there was a low degree of variance between the different models, and the ROC area under the curve (average 0.97) fell closer to a perfect score (1.0) than random guessing (0.5).

decrease in the random forest model. It was computed from all decision trees in the forest without making any assumptions, e.g. whether our data were linearly separable. Random forests can handle thousands of mixed categorical and continuous predictors and are robust to outliers, and therefore they are often used to compute an estimate of the importance of every predictor (Valletta et al., 2017). We assessed the importance of the features via the feature_importance attribute after fitting a random forest classifier with 10,000 trees.

Statistical analyses

Statistical analyses were conducted to test the difference of sequence structures between aggressive and distress calls. Pearson's chi-square statistic (chisquare and chi2_consistency) were used to test the occurrence frequencies of different syllable types, the probabilities of occurrences of transition types and the probabilities of syllable types occurring in each position of a sequence. Features related to sequence characteristics were tested with Wilcoxon rank-sum statistics (ranksums in scipy). The statistical analyses were performed using the scipy.stats package

in IPython for Windows version 5.1.0. The significance level of all tests was set at 0.05.

RESULTS

A total of 2014 aggressive call sequences were collected from 12,337-min recordings, consisting of 4692 syllables belonging to 30 different types. In the distress context, 665 call sequences were collected containing 2552 syllables belonging to 19 types (Table 1, Fig. 1). First, we employed machine learning models to classify the syntactic structures in different behavioural contexts and extracted the most important features for the classification. Then, statistical analysis was employed to test the results of the machine learning models.

Call sequence classification of machine learning algorithms

The 12 sound features extracted from various levels of call subunits including syllables, phrases and sequences were used to train the logistic regression, SVM and decision trees algorithms and perform the classification. The classification performance of the three algorithms showed impressively high accuracies on categorizing all sequences to two contexts: logistic regression 98%, SVM 97% and decision trees 96% (Fig. 2).

The contribution of each feature to the discrepancy was measured by the random forest model (Fig. 3). The most important features were related to syllable transitions (features e, h, j and c, 35.7%). The second most important features concerned the characteristics of call sequences (features d, f and i, 21.4%), and the third most important were about the positions where each syllable occurred in a sequence (feature g, 16.1%). The features concerning the sequences structure were more important than features about number or types of syllables (features a and b, 4.7%), bat individuals (feature l, 14.3%) and gender (feature k, 7.7%).

Statistical comparison of important features

Because syllables were fundamental units of call sequences, numbers and types of syllables in two contexts were compared first, although features related to syllables contributed the least percentages. The percentages of shared syllable types in the two contexts were both greater than 90% (aggressive: 95.0%, distress: 90.5%; each syllable type with a frequency of occurrence greater than two was counted), which showed high similarity. However, in both aggressive and distress contexts, greater horseshoe bats had

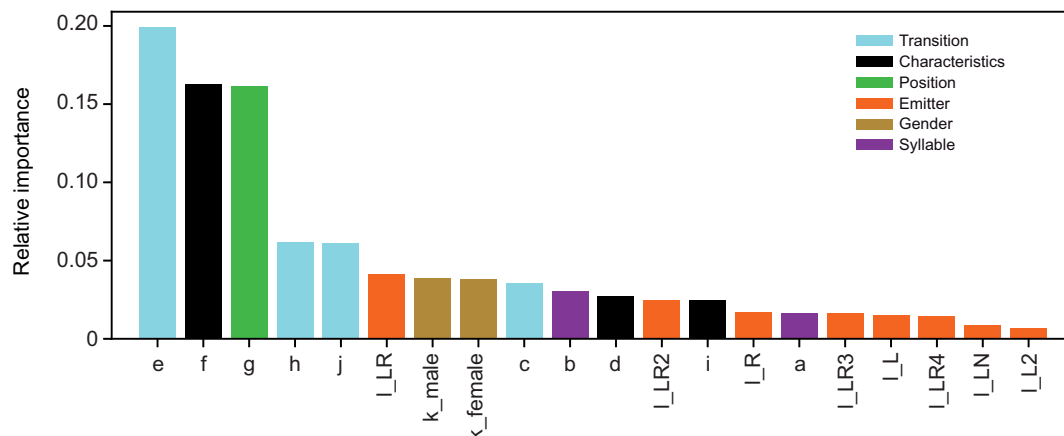


Fig. 3. Feature importance analysis and related acoustic parameters. The ranking of different features in the dataset by their relative importance. The features' importance was normalized, so that they sum to 1.0. a to l are the parameters clarified in feature selection. L, R, LR, LN and the numbers after them are the markers of eight bat individuals.

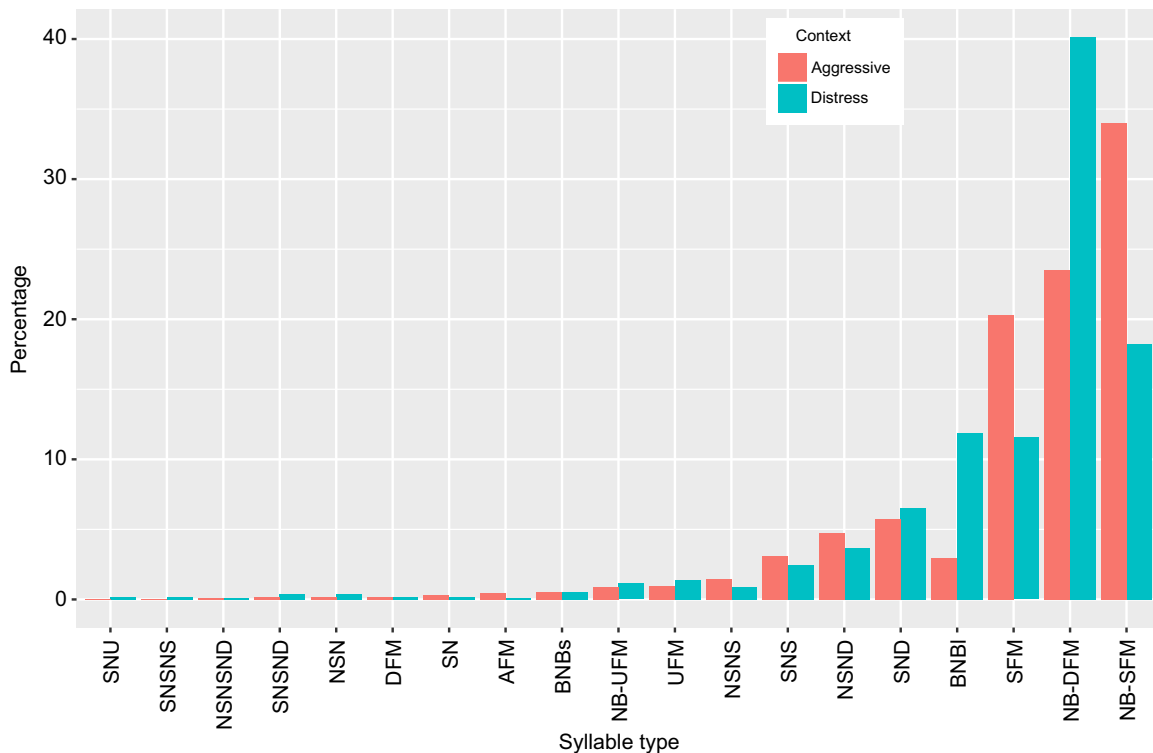


Fig. 4. Syllable types of communication vocalizations of greater horseshoe bats in aggressive and distress contexts. Syllables that constitute more than 1% of all vocalizations are plotted.

intra-context selective preferences in using different syllable types (chi-square test: aggressive: $\chi^2=26,275$, d.f.=29, $P<0.001$, distress: $\chi^2=8524.1$, d.f.=18, $P<0.001$). NB-SFM, NB-DFM and SFM were used the most by bats (aggressive: 77.8%, distress: 69.9%). Some syllable types such as SNU and SNSNS were only found several times. The syllable types occurring most frequently between two contexts were different. Bats tended to use NB-SFM in aggressive contexts but tended to use NB-DFM in distress contexts (chi-square test: $\chi^2=632.8$, d.f.=29, $P<0.001$; Fig. 4).

Transition types of two adjacent syllables in each context were also selectively used by greater horseshoe bats (Fig. 5). In aggressive calls, the transitions used frequently were self-transitions of NB-DFM ($n=402$), NB-SFM ($n=368$) and SFM ($n=220$) (chi-square test: $\chi^2=24,858.4$, d.f.=155, $P<0.001$; Fig. 5A). In distress calls, the transitions used most frequently were self-transitions of NB-DFM ($n=485$), NB-SFM ($n=126$) and BNBI ($n=138$) (chi-square test: $\chi^2=19,485$, d.f.=121, $P<0.001$; Fig. 5B). Although self-transition of NB-DFM was the most frequently used transition type in both contexts, it occupied a higher percentage in distress calls (25.7%) than in aggressive calls (15%) (chi-square test: $\chi^2=856$, d.f.=184, $P<0.001$).

Features related to characteristics of call sequences were sequence entropy (d), linearity (f) and versatility (i) (Fig. 6). Difference in entropy indicated more variable syllables occurring in distress contexts than in aggressive contexts ($F=467,414$, $P<0.001$). Sequences in aggressive context had low linearity ($F=506,663$, $P<0.001$), meaning that the ways in which syllables were ordered were more variable (or had more transition types), but high versatility ($F=876,051$, $P<0.001$), meaning there were more syllable types or shorter sequence lengths in call sequences of the aggressive context than of the distress context. These results were consistent with the overview of sequences under the two contexts presented in Table 1.

To obtain the occurrence frequency of syllable types in different positions of a sequence, we counted the numbers of the first three positions of call sequences in the aggressive context and the first four positions of call sequences in the distress context based on the average length of all sequences (Fig. 7). In aggressive calls, NB-SFM had the highest occurrence frequency in the first three positions (Fig. 7). In distress calls, NB-DFM had the highest occurrence frequency in the first four positions (Fig. 7). The probability of occurrence of syllable types in each position within a context were significantly different (aggressive context: chi-square test: position 1: $\chi^2=7695.4$, d.f.=21, $P<0.001$, position 2: $\chi^2=3735.2$, d.f.=18, $P<0.001$, position 3: $\chi^2=1745.8$, d.f.=16, $P<0.001$; distress context: position 1: $\chi^2=2496.5$, d.f.=18, $P<0.001$, position 2: $\chi^2=1358.2$, d.f.=14, $P<0.001$, position 3: $\chi^2=674.75$, d.f.=13, $P<0.001$, position 4: $\chi^2=603.4$, d.f.=13, $P<0.001$). Significant differences in the occurrence probability of syllable types in the first three positions also existed between the two contexts (chi-square test: position 1: $\chi^2=228.7$, d.f.=22, $P<0.001$, position 2: $\chi^2=144.8$, d.f.=19, $P<0.001$, position 3: $\chi^2=93.8$, d.f.=18, $P<0.001$).

DISCUSSION

This study employed machine learning methods to classify the call sequences between aggressive and distress contexts in greater horseshoe bats and to extract the features that play important roles in the classification. Logistic regression, SVM and decision trees were trained using 12 features, and each method obtained accurate classification rates greater than 95% (logistic regression 98%, SVM 97% and decision trees 96%; Fig. 2). The top three most important features for classification were all related to the structures of call sequences, including syllable transitions, positions and characteristics of call sequences (versatility, consistency and entropy). Moreover, the statistical comparison of selective

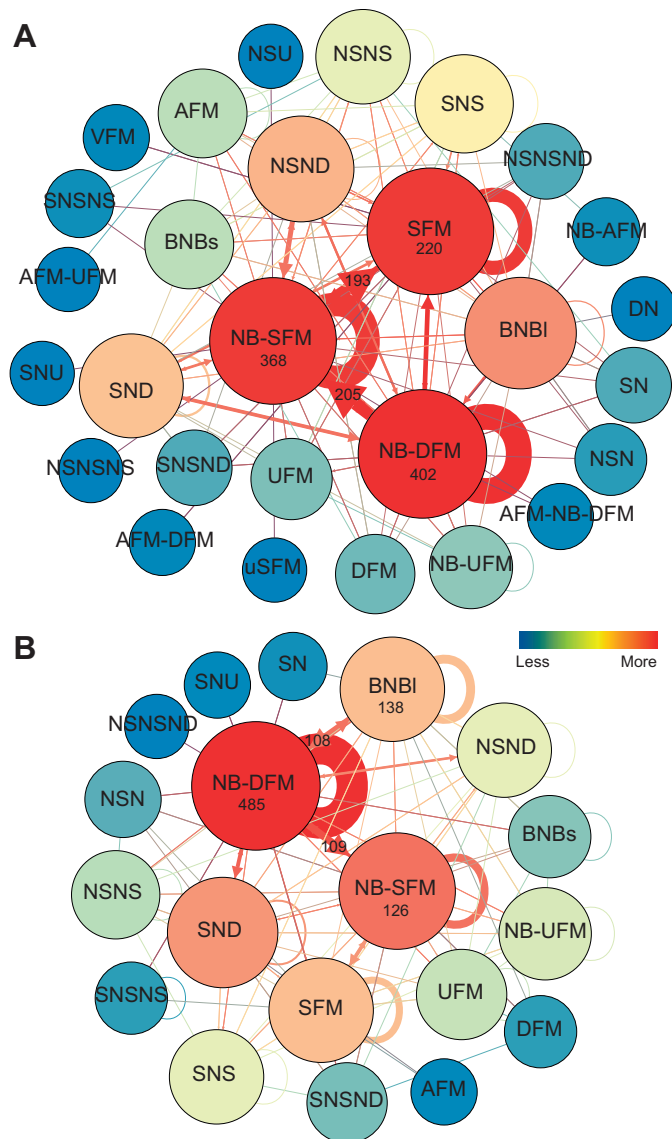


Fig. 5. Syllable transitions of communication vocalizations of greater horseshoe bats in aggressive and distress contexts. Circles represent syllables. Lines with arrows represent transition directions. Arcs at right represent self-transitions. The sizes of circles and arrows represent quantities of syllables and transitions. The numbers of the top five transitions are shown. (A) Aggressive context; (B) distress context.

preferences of syllable types, transitions, syllables occurring in different positions of a sequence and the differences in the characteristic of sequences highlighted the point that discrepancy in sequence structure existed between the two behavioural contexts. The good performance of the algorithms and the extracted important features indicated that machine learning algorithms could be a powerful tool for classifying call sequences of social vocalizations between different contexts. The method presented should enable data mining from large sound datasets in the initial step of studies on the syntax of social vocalizations in bats.

Machine learning methods could be an ideal choice for acoustic research owing to their good generalization to numerous studies. The methods have been used to solve complex problems that were previously intractable, such as dealing with large datasets and acoustic recognition in multi-species and complicated environments (Walters et al., 2012; Shamir et al., 2014; Priyadarshani et al., 2018).

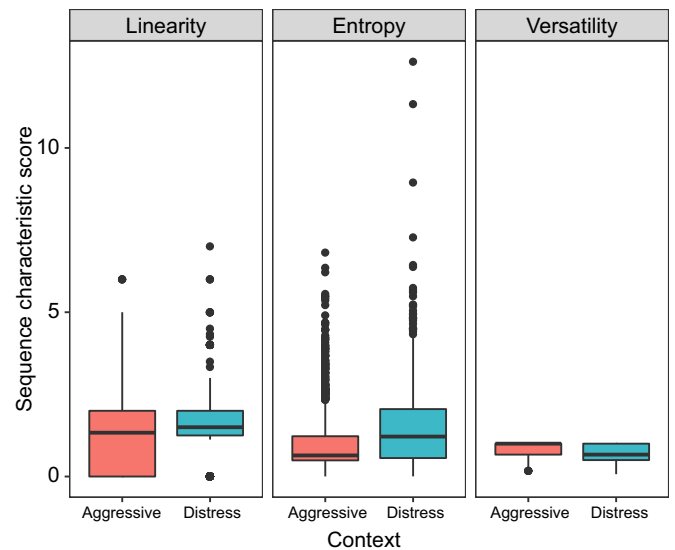


Fig. 6. Statistical comparisons of features related to characteristics of call sequences. Significant differences were found in feature linearity, entropy and versatility of call sequences between the two contexts. Scores of sequence characteristics were calculated according to the formula given in the Materials and Methods (see Feature selection).

Recently, a few studies have applied machine learning tools to acoustic detection and species classification in bats by analysing their echolocation calls (Skowronski and Harris, 2006; Armitage and Ober, 2010; Aodha et al., 2018). Our results indicated that machine learning could also be used to classify call sequences of social vocalizations in different behavioural contexts by integrating sound features of vocal subunits. Although discriminant analysis has also been frequently used for classification (Lachenbruch and Goldstein, 1979), it is more suitable for displaying the functions of fewer than three features (Mika et al., 1999; Huberty and Olejnik, 2005). Compared with discriminant analysis, the machine learning algorithms employed in the present study could clearly rank the importance of 12 features simultaneously. In addition, the logistic regression model has advantages over discriminant analysis and Hotelling's T^2 test in not needing normally distributed variables (Hoffman, 2019).

Because the aim of the present study was to compare context-dependent call sequences employing machine learning methods, our results showed successful classification of call sequences of two distinct kinds of behaviour, distress and aggression. Actually, machine learning methods also could deal with tasks of varying complexity. For example, four supervised machine learning methods were conducted on barks of domestic dogs, which obtained high percentages of correct classifications on sex (85.13%), age (80.25%), individual (67.63%) and context (55.50%) (Larrañaga et al., 2015). Prat et al. (2016) employed the Gaussian mixture model-universal background model algorithm for vocalization classifications of different aggressive contexts and different emitters in Egyptian fruit bats and obtained high balanced accuracy (different aggressive contexts for each emitter: 75%, emitters: 71%). Although the above studies focused on the call classification but not extraction of syntactic structures, they suggested the potential ability of machine learning methods for call classification of similar contexts. But further study is still invited to test the appropriate algorithms and their performances in comparing the structures of call sequences when the context is less distinct.

An important factor to consider is the ease of implementing a given method. All the algorithms we adopted could be utilized

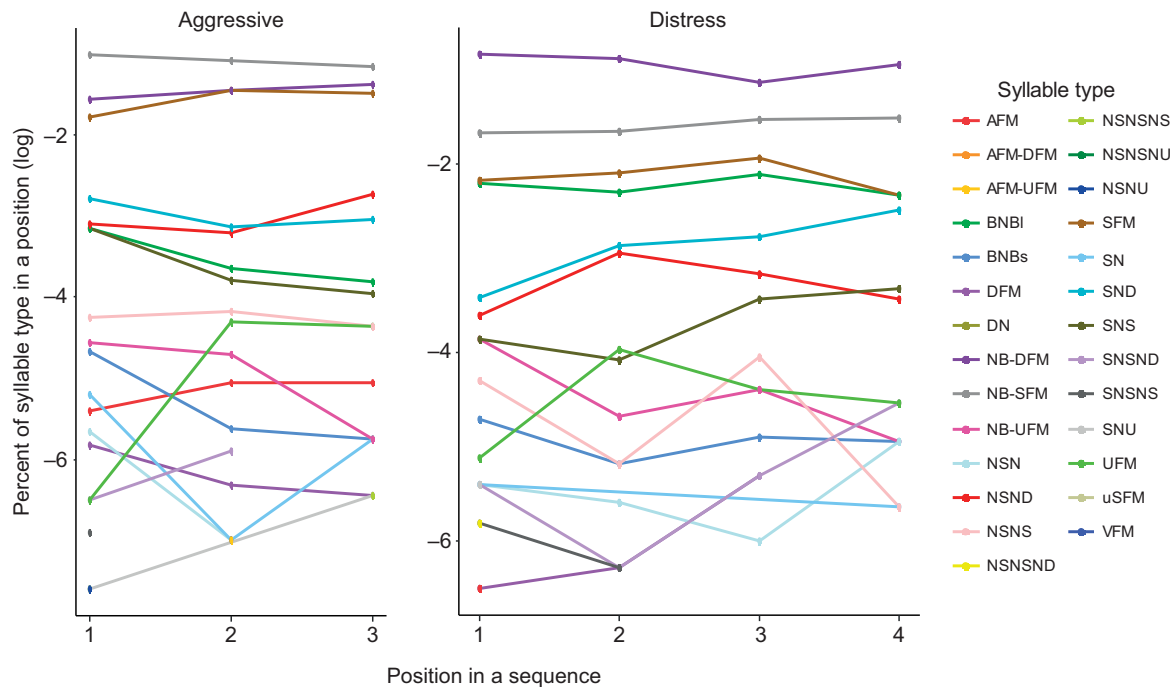


Fig. 7. Percentage of syllable types varied in different positions in aggressive versus distress calls. In each position, the frequencies of syllable type were tested with a chi-squared test, and all were significantly different.

easily and widely (Kotsiantis et al., 2007; Armitage and Ober, 2010). None need computing resources beyond an ordinary personal computer. Moreover, the features in our study were not species-specific and could be appropriate to other animal categories (Priyadarshani et al., 2018). The sound features were extracted based on syllables, which could be obtained from vocalizations of most animal species. In addition, the types of features integrated and estimated by the machine learning algorithms were variable, including not only features of sequence structure but also features of sound emitters such as individual label and gender. This demonstrates another advantage, that is, that machine learning can handle combinations of different parameters regardless of their units (Alice and Amanda, 2018). The proposed methods in our study were mainly suited for comparison and classification of call sequences under different conditions. Robust extraction and description of the structures of call sequences would require much improved processing.

The selective preference in the sound features concerning sequence structures suggested that greater horseshoe bats might order and arrange syllables according to certain rules, that is, syntactic structure in variable behavioural contexts. Using syntax composed of different structures of calls under specific situations may be a common phenomenon in many animal species. For example, syntax depending on social context was found in the 'chick-a-dee' calls of chickadees and in the songs of male mice (Clucas et al., 2004; Chabout et al., 2015). When confronted with conspecifics or predators, it was reasonable that greater horseshoe bats tended to have different reactions. Although a number of syllable types were found in all recording files, the sum of percentages for the five most frequently used types was greater than 80% in both the aggressive and distress contexts. This demonstrated that greater horseshoe bats in non-breeding periods could make good use of the order and arrangement of limited types instead of emitting complex composite syllables such as songs for mating (Davidson and Wilkinson, 2004; Bohn et al., 2009). Pioneering

research in non-human primates and birds has indicated that animal signals can be functionally referential (Townsend et al., 2013; Scarantino and Clay, 2015). The sequence compositions of Titi monkeys (*Callicebus nigrifrons*) and the specificity note combinations of discrete alarm calls of Japanese great tits (*Parus major minor*) can both be used to communicate predator type (Căsar et al., 2013; Suzuki, 2014).

In summary, our research reveals a tangible instance for employing machine learning methods to explore vocalization data. The results provided three useful and efficient models for analysing syntactic variation in bioacoustics. Using the power of machine learning, researchers can extract useful information from many vocalizations before they design behavioural experiments for further analysis. This study also demonstrated the presence of complex vocalization and potential syntactic structures of call sequences in non-breeding contexts of bats. Further experimentation, such as using playback of calls, is necessary to investigate the information encoded in different syntactic patterns.

Acknowledgements

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: K.K.Z.; Methodology: K.K.Z., W.M.; Validation: Y.L.; Formal analysis: K.K.Z., T.L.; Investigation: K.K.Z., T.L., M.X.L., A.Q.L., Y.H.X.; Data curation: K.K.Z., M.X.L., A.Q.L., Y.H.X.; Writing - original draft: K.K.Z.; Writing - review & editing: W.M., Y.L.; Visualization: K.K.Z., T.L.; Supervision: Y.L.; Funding acquisition: Y.L.

Funding

This work was funded by the National Natural Science Foundation of China (grants 31770429, 31670390), the Natural Science Foundation of Jilin (grant 20180101263JC), The Program for Introducing Talents to Universities (grant

B16011) and the National Program for '1000 Talent Plan for High-Level Foreign Experts' (grant WQ20142200259).

Data availability

All data and code have been deposited in GitHub: <https://github.com/zkkandrew/syntaxofbats>

References

- Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J. and Aide, T. M.** (2009). Automated classification of bird and amphibian calls using machine learning: a comparison of methods. *Ecol. Inform.* **4**, 206-214. doi:10.1016/j.ecoinf.2009.06.005
- Alice, Z. and Amanda, C.** (2018). *Feature Engineering for Machine Learning*. O'Reilly Media.
- Aodha, O. M., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., Harder, B., Kinsey, L., Mead, G. R. and Newson, S. E. et al.** (2018). Bat detective—Deep learning tools for bat acoustic signal detection. *PLoS Comput. Biol.* **14**, 156869. doi:10.1101/156869
- Armitage, D. W. and Ober, H. K.** (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecol. Inform.* **5**, 465-473. doi:10.1016/j.ecoinf.2010.08.001
- Berwick, R. C., Okanoya, K., Beckers, G. J. L. and Bolhuis, J. J.** (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* **15**, 113-121. doi:10.1016/j.tics.2011.01.002
- Bohn, K. M., Schmidt-French, B., Ma, S. T. and Pollak, G. D.** (2008). Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats. *J. Acoust. Soc. Am.* **124**, 1838-1848. doi:10.1121/1.2953314
- Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M. and Pollak, G. D.** (2009). Versatility and stereotypy of free-tailed bat songs. *PLoS ONE* **4**, e6746. doi:10.1371/journal.pone.0006746
- Bohn, K. M., Smarsh, G. C. and Smotherman, M.** (2013). Social context evokes rapid changes in bat song syntax. *Anim. Behav.* **85**, 1485-1491. doi:10.1016/j.anbehav.2013.04.002
- Bradbury, J. W. and Vehrencamp, S. L.** (2011). *Principles of Animal Communication*. Sinauer.
- Briefer, E. F., Rybak, F. and Aubin, T.** (2013). Does true syntax or simple auditory object support the role of skylark song dialect? *Anim. Behav.* **86**, 1131-1137. doi:10.1016/j.anbehav.2013.09.019
- Byers, B. E. and Kroodsmas, D. E.** (2009). Female mate choice and songbird song repertoires. *Anim. Behav.* **77**, 13-22. doi:10.1016/j.anbehav.2008.10.003
- Candiotti, A., Zuberbühler, K. and Lemasson, A.** (2012). Context-related call combinations in female Diana monkeys. *Anim. Cogn.* **15**, 327-339. doi:10.1007/s10071-011-0456-8
- Căsar, C., Zuberbühler, K., Young, R. J. and Byrne, R. W.** (2013). Titi monkey call sequences vary with predator location and type. *Biol. Lett.* **9**, 20130535. doi:10.1098/rsbl.2013.0535
- Catchpole, C. K. and Slater, P. J.** (2003). *Bird Song: Biological Themes and Variations*. Cambridge University Press.
- Chabout, J., Sarkar, A., Dunson, D. B. and Jarvis, E. D.** (2015). Male mice song syntax depends on social contexts and influences female preferences. *Front. Behav. Neurosci.* **9**, 76. doi:10.3389/fnbeh.2015.00076
- Chen, W.-P., Chen, S.-S., Lin, C.-C., Chen, Y.-Z. and Lin, W.-C.** (2012). Automatic recognition of frog calls using a multi-stage average spectrum. *Comput. Math. Appl.* **64**, 1270-1281. doi:10.1016/j.camwa.2012.03.071
- Clarke, E., Reichard, U. H. and Zuberbühler, K.** (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE* **1**, e73. doi:10.1371/journal.pone.0000073
- Clucas, B. A., Freeberg, T. M. and Lucas, J. R.** (2004). Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*). *Behav. Ecol. Sociobiol.* **57**, 187-196. doi:10.1007/s00265-004-0847-9
- Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B. and Townsend, S. W.** (2014). Language evolution: syntax before phonology? *Proc. Biol. Sci.* **281**, 20140263. doi:10.1098/rspb.2014.0263
- Cristianini, N. and Shawe-Taylor, J.** (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.
- Davidson, S. M. and Wilkinson, G. S.** (2004). Function of male song in the greater white-lined bat, *Saccopteryx bilineata*. *Anim. Behav.* **67**, 883-891. doi:10.1016/j.anbehav.2003.06.016
- Deslandes, V., Faria, L. R. R., Borges, M. E. and Pie, M. R.** (2014). The structure of an avian syllable syntax network. *Behav. Processes* **106**, 53-59. doi:10.1016/j.beproc.2014.04.010
- Duchesne, T., Fortin, D. and Courbin, N.** (2010). Mixed conditional logistic regression for habitat selection studies. *J. Anim. Ecol.* **79**, 548-555. doi:10.1111/j.1365-2656.2010.01670.x
- Eckenweber, M. and Knörnschild, M.** (2016). Responsiveness to conspecific distress calls is influenced by day-roost proximity in bats (*Saccopteryx bilineata*). *R. Soc. Open Sci.* **3**, 160151. doi:10.1098/rsos.160151
- Fagerlund, S.** (2007). Bird species recognition using support vector machines. *EURASIP J. Adv. Signal Process.* **2007**, 038637. doi:10.1155/2007/38637
- Fedurek, P. and Slocombe, K. E.** (2011). Primate vocal communication: a useful tool for understanding human speech and language evolution? *Hum. Biol.* **83**, 153-173. doi:10.3378/027.083.0202
- Fenton, M. B., Faure, P. A. and Ratcliffe, J. M.** (2012). Evolution of high duty cycle echolocation in bats. *J. Exp. Biol.* **215**, 2935-2944. doi:10.1242/jeb.073171
- Filatova, O. A., Deecke, V. B., Ford, J. K. B., Matkin, C. O., Barrett-Lennard, L. G., Guzeev, M. A., Burdin, A. M. and Hoyt, E.** (2012). Call diversity in the North Pacific killer whale populations: implications for dialect evolution and population history. *Anim. Behav.* **83**, 595-603. doi:10.1016/j.anbehav.2011.12.013
- Fitch, W. T.** (2010). *The Evolution of Language*. New Scientist.
- Gadziola, M. A., Grimsley, J. M. S., Faure, P. A. and Wenstrup, J. J.** (2012). Social vocalizations of big brown bats vary with behavioral context. *PLoS ONE* **7**, e44550. doi:10.1371/journal.pone.0044550
- Gillam, E. and Fenton, M. B.** (2016). Roles of acoustic social communication in the lives of bats. In *Bat Bioacoustics* (ed. R. R. Fay and A. N. Popper), pp. 117-139. Springer.
- Green, S. R., Mercado, E., Pack, A. A. and Herman, L. M.** (2011). Recurring patterns in the songs of humpback whales (*Megaptera novaeangliae*). *Behav. Processes* **86**, 284-294. doi:10.1016/j.beproc.2010.12.014
- Hailman, J. P., Ficken, M. S. and Ficken, R. W.** (1985). The 'chick-a-dee' calls of *Parus atricapillus*: a recombinant system of animal communication compared with written English. *Semiotica* **56**, 191-224. doi:10.1515/semi.1985.56.3-4.191
- Hoffman, J. I. E.** (2019). Logistic regression. In *Basic Biostatistics for Medical and Biomedical Practitioners*, 2nd edn (ed. J. I. E. Hoffman), pp. 581-589. Academic Press.
- Huang, C.-J., Yang, Y.-J., Yang, D.-X. and Chen, Y.-J.** (2009). Frog classification using machine learning techniques. *Expert Syst. Appl.* **36**, 3737-3743. doi:10.1016/j.eswa.2008.02.059
- Huang, X., Metzner, W., Zhang, K., Wang, Y., Luo, B., Sun, C., Jiang, T. and Feng, J.** (2018). Acoustic similarity elicits responses to heterospecific distress calls in bats (Mammalia: Chiroptera). *Anim. Behav.* **146**, 143-154. doi:10.1016/j.anbehav.2018.10.018
- Huberty, C. J. and Olejnik, S.** (2005). Discriminant analysis in research. In *Applied MANOVA and Discriminant Analysis* (ed. W. A. Shewhart and S. S. Wilks), pp. 5-12. Wiley Online Library.
- Jiang, T., Huang, X., Wu, H. and Feng, J.** (2017). Size and quality information in acoustic signals of *Rhinolophus ferrumequinum* in distress situations. *Physiol. Behav.* **173**, 252-257. doi:10.1016/j.physbeh.2017.02.025
- Jones, G. and Siemers, B. M.** (2011). The communicative potential of bat echolocation pulses. *J. Comp. Physiol. A* **197**, 447-457. doi:10.1007/s00359-010-0565-x
- Kanwal, J. S., Matsumura, S., Ohlemiller, K. and Suga, N.** (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *J. Acoust. Soc. Am.* **96**, 1229-1254. doi:10.1121/1.410273
- Kershenbaum, A., Ilany, A., Blaustein, L. and Geffen, E.** (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proc. Biol. Sci.* **279**, 2974-2981. doi:10.1098/rspb.2012.0322
- Knörnschild, M.** (2014). Vocal production learning in bats. *Curr. Opin. Neurobiol.* **28**, 80-85. doi:10.1016/j.conb.2014.06.014
- Kotsiantis, S. B., Zaharakis, I. D. and Pintelas, P. E.** (2007). Machine learning: a review of classification and combining techniques. *Artif. Intell. Rev.* **26**, 159-190. doi:10.1007/s10462-007-9052-3
- Lachenbruch, P. A. and Goldstein, M.** (1979). Discriminant analysis. *Biometrics* **35**, 69-85. doi:10.2307/2529937
- Larrañaga, A., Bielza, C., Pongrácz, P., Faragó, T., Bálint, A. and Larrañaga, P.** (2015). Comparing supervised learning methods for classifying sex, age, context and individual Mudi dogs from barking. *Anim. Cogn.* **18**, 405-421. doi:10.1007/s10071-014-0811-7
- Lin, H.-J., Kanwal, J. S., Jiang, T.-L., Liu, Y. and Feng, J.** (2015). Social and vocal behavior in adult greater tube-nosed bats (*Murina leucogaster*). *Zoology* **118**, 192-202. doi:10.1016/j.zool.2014.12.005
- Lin, A., Jiang, T., Feng, J. and Kanwal, J. S.** (2016a). Acoustically diverse vocalization repertoire in the Himalayan leaf-nosed bat, a widely distributed *Hipposideros* species. *J. Acoust. Soc. Am.* **140**, 3765. doi:10.1121/1.4966286
- Lin, A., Liu, H., Chang, Y., Lu, G. and Feng, J.** (2016b). Behavioural response of the greater horseshoe bat to geographical variation in echolocation calls. *Behav. Ecol. Sociobiol.* **70**, 1765-1776. doi:10.1007/s00265-016-2182-3
- Luo, B., Jiang, T., Liu, Y., Wang, J., Lin, A., Wei, X. and Feng, J.** (2013). Brevity is prevalent in bat short-range communication. *J. Comp. Physiol. A* **199**, 325-333. doi:10.1007/s00359-013-0793-y
- Ma, J., Kobayashi, K., Zhang, S. and Metzner, W.** (2006). Vocal communication in adult greater horseshoe bats, *Rhinolophus ferrumequinum*. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **192**, 535-550. doi:10.1007/s00359-006-0094-9
- Magrath, R. D., Haff, T. M., Fallow, P. M. and Radford, A. N.** (2015). Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. *Biol. Rev.* **90**, 560-586. doi:10.1111/brv.12122

- Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (2013). *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media.
- Mika, S., Ratsch, G., Weston, J., Scholkopf, B. and Mullers, K.-R. (1999). Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX*, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, pp. 41-48. IEEE.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825-2830.
- Pozzi, L., Gamba, M. and Giacoma, C. (2010). The use of artificial neural networks to classify primate vocalizations: a pilot study on black lemurs. *Am. J. Primatol.* **72**, 337-348.
- Prat, Y., Taub, M. and Yovel, Y. (2016). Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Sci. Rep.* **6**, 39419. doi:10.1038/srep39419
- Prat, Y., Azoulay, L., Dor, R. and Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biol.* **15**, e2002556. doi:10.1371/journal.pbio.2002556
- Priyadarshani, N., Marsland, S. and Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *J. Avian Biol.* **49**, jav-01447. doi:10.1111/jav.01447
- Prugh, L. R., Hodges, K. E., Sinclair, A. R. E. and Brashares, J. S. (2008). Effect of habitat area and isolation on fragmented animal populations. *Proc. Natl. Acad. Sci. USA* **105**, 20770. doi:10.1073/pnas.0806080105
- Ranjard, L. and Ross, H. A. (2008). Unsupervised bird song syllable classification using evolving neural networks. *J. Acoust. Soc. Am.* **123**, 4358-4368. doi:10.1121/1.2903861
- Raschka, S. and Mirjalili, V. (2017). *Python Machine Learning*, 2nd edn. Packt Publishing.
- Russ, J. M., Jones, G., Mackie, I. and Racey, P. A. (2004). Interspecific responses to distress calls in bats (Chiroptera: Vespertilionidae): a function for convergence in call design? *Anim. Behav.* **67**, 1005-1014. doi:10.1016/j.anbehav.2003.09.003
- Scarantino, A. and Clay, Z. (2015). Contextually variable signals can be functionally referential. *Anim. Behav.* **100**, e1-e8. doi:10.1016/j.anbehav.2014.08.017
- Scharff, C. and Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11**, 2896-2913. doi:10.1523/JNEUROSCI.11-09-02896.1991
- Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Căsar, C. and Keenan, S. et al. (2016). Formal monkey linguistics: the debate. *Theoretical Linguistics* **42**, 173-201. doi:10.1515/tl-2016-0010
- Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A. M., Tyack, P., Samarra, F., Miller, P. and Wallin, J. (2014). Classification of large acoustic datasets using machine learning and crowdsourcing: application to whale calls. *J. Acoust. Soc. Am.* **135**, 953-962. doi:10.1121/1.4861348
- Siemers, B. M., Wiegrebe, L. and Grothe, B. (2011). Ecology and neuroethology of bat echolocation: a tribute to Gerhard Neuweiler. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **197**, 399-402. doi:10.1007/s00359-011-0633-x
- Skowronski, M. D. and Harris, J. G. (2006). Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. *J. Acoust. Soc. Am.* **119**, 1817-1833. doi:10.1121/1.2166948
- Stathopoulos, V., Zamora-Gutierrez, V., Jones, K. E. and Girolami, M. (2018). Bat echolocation call identification for biodiversity monitoring: a probabilistic approach. *J. R. Stat. Soc. C* **67**, 165-183. doi:10.1111/rssc.12217
- Sun, C., Zhang, C., Gu, H., Jiang, T. and Feng, J. (2019). Self-assessment strategy during contest decisions between male great Himalayan leaf-nosed bats. *Behav. Ecol. Sociobiol.* **73**, 45. doi:10.1007/s00265-019-2657-0
- Suzuki, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim. Behav.* **87**, 59-65. doi:10.1016/j.anbehav.2013.10.009
- Suzuki, T. N., Wheatcroft, D. and Griesser, M. (2018). Call combinations in birds and the evolution of compositional syntax. *PLoS Biol.* **16**, e2006532. doi:10.1371/journal.pbio.2006532
- Townsend, S. W., Manser, M. B. and Hauber, M. (2013). Functionally referential communication in mammals: the past, present and the future. *Ethology* **119**, 1-11. doi:10.1111/eth.12015
- Turesson, H. K., Ribeiro, S., Pereira, D. R., Papa, J. P. and de Albuquerque, V. H. C. (2016). Machine learning algorithms for automatic classification of marmoset vocalizations. *PLoS ONE* **11**, e0163041. doi:10.1371/journal.pone.0163041
- Valletta, J. J., Torney, C., Kings, M., Thornton, A. and Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Anim. Behav.* **124**, 203-220. doi:10.1016/j.anbehav.2016.12.005
- Walters, C. L., Freeman, R., Collen, A., Dietz, C., Brock Fenton, M., Jones, G., Obrist, M. K., Puechmaille, S. J., Sattler, T., Siemers, B. M. et al. (2012). A continental-scale tool for acoustic identification of European bats. *J. Appl. Ecol.* **49**, 1064-1074. doi:10.1111/j.1365-2664.2012.02182.x
- Zhao, X., Jiang, T., Gu, H., Liu, H., Sun, C., Liu, Y. and Feng, J. (2018). Are aggressive vocalizations the honest signals of body size and quality in female Asian particoloured bats? *Behav. Ecol. Sociobiol.* **72**, 96. doi:10.1007/s00265-018-2510-x
- Zhao, X., Jiang, T., Liu, H., Wang, Y., Liu, Y. and Feng, J. (2019). Acoustic signalling of aggressive intent in the agonistic encounters of female Asian particoloured bats. *Anim. Behav.* **149**, 65-75. doi:10.1016/j.anbehav.2019.01.012