Bat Call Identification with Gaussian Process Multinomial Probit Regression and a Dynamic Time Warping Kernel

Vassilios Stathopoulos
Department of Statistics
University of Warwick

Veronica Zamora-Gutierrez
Department of Zoology
University of Cambridge

Kate E. Jones
Centre for Biodiversity and Environment Research,
Department of Genetics, Evolution and Environment
UCL

Mark Girolami
Department of Statistics
University of Warwick

Abstract

We study the problem of identifying bat species from echolocation calls in order to build automated bioacoustic monitoring algorithms. We employ the Dynamic Time Warping algorithm which has been successfully applied for bird flight calls identification and show that classification performance is superior to hand crafted call shape parameters used in previous research. This highlights that generic bioacoustic software with good classification rates can be constructed with little domain knowledge. We conduct a study with field data of 21 bat species from the north and central Mexico using a multinomial probit regression model with Gaussian process prior and a full EP approximation of the posterior of latent function values. Results indicate high classification accuracy across almost all classes while misclassification rate across families of species is low highlighting the common evolutionary path of echolocation in bats.

1 Introduction

In many tropical ecosystems, bats are keystone species as they act as important pollinators, seed dispersal agents and regulators of insect populations (Jones et al., 2009). Inspite of their importance, most bat population studies in the tropics have been short term and the lack of long term bat monitoring programs is a result of their inherent difficulty. Bats produce unique sounds at frequencies that usually do not overlap with other species and most bat species have evolved species-specific echolocation calls (Fenton and Bell, 1981; Jones and Teeling, 2006; Ahlen and Baage, 1999). However, their calls also show great interspecies variation and flexibility caused by habitat, geography, sex, age, etc. and in other cases there is a great overlap of call structures between species which makes species identification complicated (Obrist, 1995; Murray et al., 2001; Schnitzler et al., 2003). Developing automatic identification tools would therefore assist in creating long term acoustic monitoring programs for biodiversity.

This work is a first step towards this direction. Our aim here is not to do an exhaustive comparison of methods, but to show that using state of the art algorithms from the Machine Learning literature and with no significant tunning or heavily engineered feature extraction good identification rates can be achieved.

In this study we use data of 21 species collected in North and Central Mexico and treat bat call identification as a supervised classification problem. A representative set of bat calls is used to train a classification model which is then applied to classify novel instances of bat calls. We employ a Multinomial probit regression model with Gaussian process prior (Girolami and Rogers, 2006) which is a state of the art discriminative classification model achieving good generalization capabilities with moderate to low numbers of training data. We also utilize a kernel representation of the data that directly compares the calls' spectrograms and thus requires minor tunning.

The rest of the paper is organised as follows. In Section 2 we briefly review related work to our study. In Section 3 we describe our methods. The GP classification model is described in detail in Section 3.1 and how to

Appearing in Proceedings of the 17^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2014, Reykjavik, Iceland. JMLR: W&CP volume 33.

efficiently optimise parameters is discussed in Section 3.2. Different data representations are discussed in Sections 3.3-3.5. Section 4 discusses how data are collected and our experiments while Section 5 concludes the paper. The data used in this paper are available to download[1] as supplementary material with the aim to promote researchers in the Statistics, Artificial Intelligence and Machine Learning fields to conduct research in the area of bioacoustic monitoring for biodiversity programs.

2 Related Work

2.1 Bat Call Classification

Previous research on bat call identification has also approached the problem from a supervised learning perspective. The most studied methods employ a number of call parameters extracted from the calls spectrogram or FFT and then Discriminant Function Analysis and Artificial Neural Networks are employed for supervised classification (Walters et al., 2012; Parsons and Jones, 2000; Fenton and Bell, 1981). Although the call shape parameters encode important prior knowledge on bat call shapes, they are not flexible enough to model inter and intra species variations and capture the rich information encoded in the calls' spectrograms, such as harmonics. Moreover, the classification algorithms utilised in previous research impose constraints on the data representation and make integration of different sources of data and representation difficult. Finally, optimising the architecture of an Artificial Neural Network is not trivial and little guidance is available.

In this work we show that by comparing directly the spectrograms with the Dynamic Time Warping (DTW) algorithm we can obtain better classification accuracy than using call shape parameters. However we also show that augmenting the DTW representation with call shape parameters can further improve classification indicating that prior knowledge on call shapes is an important factor for bat call identification.

2.2 Dynamic Time Warping Kernels

Dynamic Time Warping (DTW) is a dynamic programming algorithm (Sakoe and Chiba, 1978) for comparing sequence-based and time-series data which can vary in time or speed. Given two sequences, $\boldsymbol{x}$ of length N and $\boldsymbol{y}$ of length M, it stretches or expands $\boldsymbol{x}$ in order to match $\boldsymbol{y}$ by minimizing the alignment cost based on an application specific function or some distant measure obtained from the warped paths.

[1] `http://www.engage-project.org/`

More formally, given two sequences $\boldsymbol{x}, \boldsymbol{y}$ of lengths N and M respectively and a local cost or dissimilarity matrix $\boldsymbol{D} \in \mathbb{R}^{N,M}$ for each pair of elements in $\boldsymbol{x}, \boldsymbol{y}$, an *alignment path* is a sequence $p = (p_1, \ldots, p_L)$ with $p_l = (n_l, m_l) \in [1:N] \times [1:M]$. The total cost of an *alignment path* is $c_p = \sum_{l=1}^{L} D(x_{n_l}, y_{m_l})$ while the optimal alignment score is defined as the alignment with the minimum total score, i.e. $c_{p^*} = \underset{p}{\operatorname{argmin}}\, c_p$.

With the success of kernel based classification methods such as the Support Vector Machines and Gaussian process classifiers several researches have investigated the use of DTW to construct positive definite kernel functions. The kernel proposed by Hansheng and Bingyu (2007) however is not guaranteed to be positive definite. Damoulas et al. (2010) proposed a *global* DTW kernel by constructing a vector representation utilising the optimal alignment costs with all training sequences. Their work has shown great classification accuracy of bird flight calls and we therefore adopt their method in our study.

2.3 Multi-Class GP Classification

Gaussian process regression (Rasmussen and Williams, 2005) has been a well known algorithm in the Machine Learning comunity, known for it superior generalisation capabilities with moderate to low numbers of training data. It is based on a kernel representation of the data, allowing for different types of data representations to be utilised, and is not restricting the use of vector representations with the same length, thus making it suitable for times-series.

Classification with GP models however is not amendable to analytical solutions and usually approximate inference methods are used. For binary classification, the Expectation Propagation (EP) algorithm has been shown to provide better approximation to the necessary integrals required for inference (Kuss and Rasmussen, 2005). However, for the multinomial case, where there are many mutually exclusive classes, the EP algorithm requires additional approximations of the *tilted* distributions (Girolami and Zhong, 2007). Recently Riihimaki et al. (2013) proposed the *nested* EP algorithm where the moments of the *tilted* distributions are obtained by an inner EP approximation. They show that a single iteration of the inner EP algorithm is enough for the algorithm to converge thus providing significant computational savings. In this paper we show that the same algorithm of Riihimaki et al. (2013) can be obtained by an augmentation and permutation of the latent function values and thus there is no need to interpret the algorithm as a *nested* EP.

3 Methodology

We approach bat call identification as a classification problem where the class response variables $y_n \in \{1, \dots C\}$ indicate the species id for the n^{th} call in the library and $\boldsymbol{x} \in \mathbb{R}^D$ is a D-dimensional vector representation of the call, e.g. features extracted from the call's spectrogram. We will discuss how such representation is generated for each call in Sections 3.3 - 3.4. Species' ids from all calls in the library are collected in a vector $\boldsymbol{y} = [y_1, \dots, y_N]$ and all call vector representations are collected in the matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_N]^T$ of size $N \times D$. In Section 3.1 we will define a probabilistic model for the conditional probability $p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta})$ where θ denotes a vector of unknown model parameters with an associated prior distribution $p(\boldsymbol{\theta})$. The id for a new call, y^*, with vector representation $\boldsymbol{x}^*$ is obtained by the class with highest probability from $p(y^*|\boldsymbol{x}^*, \boldsymbol{X}, \boldsymbol{y}, \hat{\boldsymbol{\theta}})$ where parameter estimates $\hat{\boldsymbol{\theta}}$ are obtained by maximizing the posterior distribution, i.e. $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{y})$.

3.1 Multinomial Probit Regression with GP prior

The probabilistic model assumes a *latent* function $\boldsymbol{f} : \mathbb{R}^D \rightarrow \mathbb{R}^C$ with *latent* values $\boldsymbol{f}(\boldsymbol{x}_n) = \boldsymbol{f}_n = [f_n^1, f_n^2, \dots, f_n^C]^T$ such that when transformed by a sigmoid-like function give the class probabilities $p(y_n|\boldsymbol{f}_n)$. Here we use a the multinomial probit function,

$$p(y_n|\boldsymbol{f}_n) = \int \mathcal{N}(u_n|0,1) \prod_{j=1, j\neq y_n}^{C} \Phi(u_n + f_n^{y_n} - f_n^j) du_n, \tag{1}$$

which is convenient for deriving the EP approximation and Gibbs sampling (Seeger et al., 2006; Girolami and Rogers, 2006). For the *latent* function values we assume independent zero-mean Gaussian process priors for each class similar to Rasmussen and Williams (2005). Collecting *latent* function values for all calls and classes in $\boldsymbol{f} = [f_1^1, \dots, f_N^1, f_1^2, \dots, f_N^2 \dots, f_1^C, \dots, f_N^C]^T$ the GP prior is

$$p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{0}, \boldsymbol{K}(\boldsymbol{\theta})), \tag{2}$$

where $\boldsymbol{K}(\boldsymbol{\theta})$ is a $CN \times CN$ block covariance matrix with block matrices $\boldsymbol{K}^1(\boldsymbol{\theta}), \dots, \boldsymbol{K}^C(\boldsymbol{\theta})$, each of size $N \times N$, on its diagonal. Elements $K_{i,j}^c$ define the prior covariance between the *latent* function values f_i^c, f_j^c governed by a covariance function $k(\boldsymbol{x}_i, \boldsymbol{x}_j|\boldsymbol{\theta})$ with unknown parameters $\boldsymbol{\theta}$. A common choice for the covariance function is the squared exponential defined as

$$k_{se}(\boldsymbol{x}_i, \boldsymbol{x}_j|\boldsymbol{\theta}) = \sigma^2 \exp\left(-\frac{1}{2}\lambda^{-2}\sum_{d=1}^{D}(x_{i,d} - x_{j,d})^2\right), \tag{3}$$

with parameters $\boldsymbol{\theta} = [\sigma^2, \lambda]^T$, but we will discuss other variations in Sections 3.3 - 3.5.

Optimising the unknown kernel parameters $\boldsymbol{\theta}$ involves computing and maximising the posterior

$$p(\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{y}) \propto p(\boldsymbol{\theta}) \int p(\boldsymbol{y}|\boldsymbol{f}) p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{\theta}) d\boldsymbol{f}. \tag{4}$$

Making predictions for a new call, y_*, $\boldsymbol{x}_*$, involves two steps. First computing the distribution of the *latent* function values for the new call

$$p(\boldsymbol{f}_*|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \int p(\boldsymbol{f}_*|\boldsymbol{x}_*, \boldsymbol{f}, \boldsymbol{X}, \hat{\boldsymbol{\theta}}) p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{y}, \hat{\boldsymbol{\theta}}) d\boldsymbol{f}, \tag{5}$$

and then computing the class probabilities using the multinomial probit function

$$p(y_*|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \int p(y_*|\boldsymbol{f}_*) p(\boldsymbol{f}_*|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}, \hat{\boldsymbol{\theta}}) d\boldsymbol{f}_*. \tag{6}$$

3.2 Full EP approximation

Unfortunately exact inference, i.e. computing the integrals, for Equations (4-6) is not possible and we have to either resort to numerical estimation through Markov Chain Monte Carlo or use approximate methods. Due to the large number of classes (21 species in our data) in this work we consider the latter approach and use Expectation Propagation (EP) (Minka, 2001) to approximate the posterior of the *latent* function values $p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta})$ in Equations (4) and (5) while for computing the integral in (6) we can again use the EP algorithm.

The EP method approximates the posterior using

$$q_{EP}(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}) \approx \frac{1}{Z_{EP}} p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{\theta}) \prod_{n=1}^{N} \widetilde{t}_n(\boldsymbol{f_n}|\widetilde{Z}_n, \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n), \tag{7}$$

where $\widetilde{t}_n(\boldsymbol{f_n}|\widetilde{Z}_n, \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n) = \widetilde{Z}_n \mathcal{N}\left(\boldsymbol{f_n}|\widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n\right)$ are local *likelihood* approximate terms with parameters $\widetilde{Z}_n, \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n$. The approximation parameters are updated by first computing the *cavity* distribution

$$q_{-n}(\boldsymbol{f}_n) = q_{EP}(\boldsymbol{f}_n|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}) \widetilde{t}_n(\boldsymbol{f_n}|\widetilde{Z}_n, \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n)^{-1} \tag{8}$$

and then matching $\widetilde{Z}_n, \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n$ with the zero, first and second moments of the corresponding *tilted* distribution

$$\hat{q}(\boldsymbol{f}_n) = \hat{Z}_n^{-1} q_{-n}(\boldsymbol{f}_n) p(y_n|\boldsymbol{f}_n). \tag{9}$$

Unlike the binary probit case, where the *tilted* distribution (9) is univariate and thus its moments are easy to compute, the *tilted* distribution for the multinomial probit model is C-dimensional. Previous work on EP approximations for the multinomial probit model (Girolami and Zhong, 2007) further approximated the moments of the *tilted* distribution using the Laplace approximation. This assumes that the distributions can be closely approximated my a multivariate normal.

In this work we show that a full EP algorithm can be derived by augmenting the *latent* function values $\boldsymbol{f}$ with the *auxiliary* variables u_n from Equation (1) and permuting both the augmented variables and the covariance matrix $\boldsymbol{K}(\boldsymbol{\theta})$. This results in the same algorithm as the *"nested"* EP approximation presented by Riihimaki et al. (2013). However this presentation clearly shows why a single iteration of the *inner* EP for the *tilted* distributions using the moments estimated from the previous iteration of the *outer* EP is enough for the algorithm to converge.

We introduce the new variables $\boldsymbol{w}$ which are formed by augmenting $\boldsymbol{f}$ with u_n and permuting such that $\boldsymbol{w} = [f_1^1, \ldots, f_1^C, u_1, f_2^1, \ldots, f_2^C, u_2, \ldots, f_N^1, \ldots, f_N^C, u_N]^T$. Similarly we augment the covariance matrix $\boldsymbol{K}(\boldsymbol{\theta})$ and permute accordingly such that the new covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ is a $(C+1)N \times (C+1)N$ block matrix with blocks $\boldsymbol{V}(\boldsymbol{\theta})_{i,j} = \text{diag}([K_{i,j}^1, \ldots, K_{i,j}^C, \delta_{i=j}])$, $i, j \in \{1, \ldots, N\}$ of size $C+1 \times C+1$ and $\delta_{i=j}$ is 1 if and only if $i = j$. Now we can write the posterior for $\boldsymbol{w}$ as

$$p(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \boldsymbol{V}) \prod_{n=1}^{N} \prod_{j=1, j\neq y_n}^{C} \Phi(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}), \tag{10}$$

where $\boldsymbol{w}_n = [f_n^1, \ldots, f_n^C, u_n]^T$ and $\boldsymbol{b}_{n,j} = [(\boldsymbol{e}_{y_n} - \boldsymbol{e}_j), 1]^T$ with $\boldsymbol{e}_j$ a C-dimensional vector of zeros and the j^{th} element set to 1.

The EP approximate posterior for $\boldsymbol{w}$ follows as

$$q_{ep}(\boldsymbol{w}) = Z_{ep}^{-1} \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \boldsymbol{V}) \prod_{n=1}^{N} \prod_{j=1, j\neq y_n}^{C} \widetilde{t}_{n,j}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}), \tag{11}$$

where $\widetilde{t}_{n,j}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}) = \widetilde{Z}_{n,j}^{-1} \mathcal{N}\left(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j} | \widetilde{\beta}_{n,j}, \widetilde{\alpha}_{n,j}\right)$ are the local approximate terms with parameters $\widetilde{Z}_{n,j}, \widetilde{\beta}_{n,j}, \widetilde{\alpha}_{n,j}$. This corresponds to an approximate posterior with $N(C-1)$ local approximation terms which have to be updated by matching their moments with the corresponding *tilted* distributions

$$\hat{q}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}) = \hat{Z}_{n,j}^{-1} q_{-n,j}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}) \Phi(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}), \tag{12}$$

where $q_{-n,j}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}) = q_{ep}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j}) \widetilde{t}_{n,j}(\boldsymbol{w}_n^T \boldsymbol{b}_{n,j})^{-1}$ are the cavity distributions. Calculating the moments for the *tilted* distribution can now be done analytically as Equation (12) resembles the *tilted* distribution of the probit model (Rasmussen and Williams, 2005; Riihimaki et al., 2013).

3.3 Spectrogram Features

The vector representation $\boldsymbol{x}_n$ for each call is constructed by extracting call shape parameters from the call's spectrogram similar to Walters et al. (2012). The spectrogram of a call is calculated by using a hamming window of size 256 with 95% overlap and an FFT length of 512. The frequency range of the spectrogram is thresholded by removing frequencies below 5kHz and above 210kHz. An example of a call's spectrogram is illustrated in Figure 1. In total 32 parameters are calculated including the call's duration in miliseconds, the highest and lowest frequencies of the call, its total frequency spread, the frequency with maximum amplitude, the frequencies at the start and end of the call etc. We do not give a full list of the call parameters here as they are available to download in the supplementary material[2] accompanying this paper.

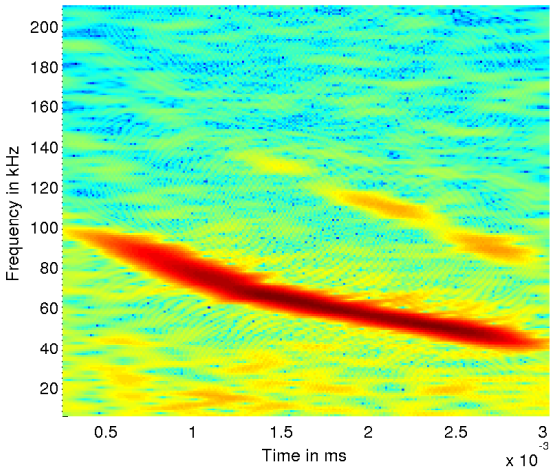

Figure 1: Example of a call's spectrogram. See text for details on spectrogram computation.

All 32 call parameters are concatenated in the vector $\boldsymbol{x}_n$ and a squared exponential kernel

$$k_{sei}(\boldsymbol{x}_i, \boldsymbol{x}_j|\boldsymbol{\theta}) = \sigma^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D} \lambda_d^{-2}(x_{i,d} - x_{j,d})^2\right), \tag{13}$$

with individual length scales for each parameter is used for the GP classifier.

[2] http://www.engage-project.org/

3.4 Dynamic Time Warping Kernel

Although extracting call shape parameters from the spectrogram of a call captures some of the call's characteristics and shape, there is still a lot of information that is discarded, e.g. harmonics. An alternative to characterising a call using predefined parameters is to directly utilise its spectrogram. However due to the differences in call duration the spectrograms will need to be normalised in order to have the same length using some form of interpolation. In this work we borrow ideas from speech recognition (Sakoe and Chiba, 1978) and previous work on bird call classification (Damoulas et al., 2010) and employ the Dynamic Time Warping (DTW) kernel to directly compare two calls' spectrograms.

Given two calls i, j from the library and their spectrograms $\boldsymbol{S}_i, \boldsymbol{S}_j$, where $\boldsymbol{S}_i \in \mathbb{C}^{F\times W}$ with F being the number of frequency bands and W the number of windows, the dissimilarity matrix $\boldsymbol{D}^{i,j} \in \mathbb{R}^{W\times W}$ is constructed such that

$$\boldsymbol{D}^{i,j}(w,v) = 1 - \frac{\boldsymbol{S}_i(:,w)^T \boldsymbol{S}_j(:,v)}{\sqrt{\boldsymbol{S}_i(:,w)^T \boldsymbol{S}_i(:,w) \boldsymbol{S}_j(:,v)^T \boldsymbol{S}_j(:,v)^T}}. \tag{14}$$

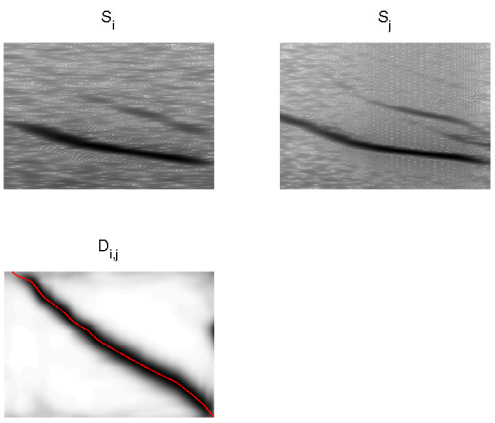

Figure 2: Example of DTW optimal warping path for 2 call spectrograms from the same species. Top row, call spectrograms; bottom row, dissimilarity matrix and optimal warping path.

DTW uses the dissimilarity matrix in order to stretch or expand spectrogram $\boldsymbol{S}_i$ over time in order to match $\boldsymbol{S}_j$ by calculating the optimal warping path with the smallest alignment cost, $c_{i,j}$, using dynamic programming. Figure 2 illustrates the optimal warping path for two calls in the library.

For each call we construct a vector representation $\boldsymbol{x}_n$ by computing the optimal warping paths with all N calls from the library and concatenating the alignment costs such that $\boldsymbol{x}_n = [c_{n,1}, \ldots, c_{n,N}]$. We then use the squared exponential covariance function in Equation (3) for the covariance matrix of the GP classifier.

3.5 Multiple Kernel GP

GP classifiers allow for integrating information from different sources or different representations of the data by combining covariance functions. Although both representations discussed in the previous sections are extracted from a call's spectrogram, some of the call parameters used in Section 3.3 involve non-linear and complex transformations of the spectrograms by utilising prior knowledge of bat call shapes. Since such knowledge is important for bat call identification and is not present in the DTW representation we combine both kernels by a weighted sum and treating the weights as unknown parameters.

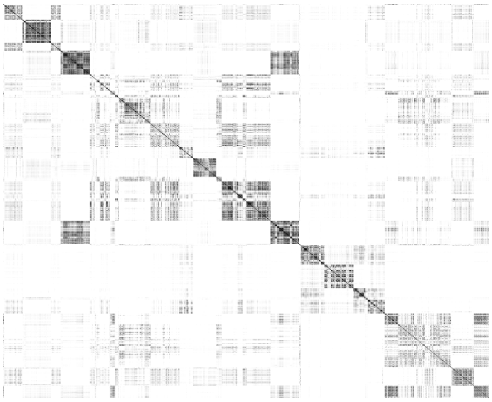

Figure 3: Class sorted kernel matrix for the weighted average of the DTW and call shape parameters representations.

Denoting the vector representation computed by extracting call parameters from the spectrogram by $\boldsymbol{x}_n^{(vect)}$ and the DTW by $\boldsymbol{x}_n^{(dtw)}$, the new covariance function is

$$w_1 k_{sei}(\boldsymbol{x}_i^{(vect)}, \boldsymbol{x}_j^{(vect)}) + w_2 k_{se}(\boldsymbol{x}_i^{(dtw)}, \boldsymbol{x}_j^{(dtw)}),$$

where we restrict w_1 and w_2 to sum to 1. Figure 3 illustrates the kernel matrix obtained on the training set of calls in our library with optimised parameters.

4 Experimental Setup and Data

4.1 Data

Bat echolocation calls were recorded across North and Central Mexico from June to November 2012 and from

Table 1: Dataset statistics

Species	Samples	Calls
Family: *Emballonuridae*		
1 *Balantiopteryx plicata*	16	384
Family: *Molossidae*		
2 *Nyctinomops femorosaccus*	16	311
3 *Tadarida brasiliensis*	49	580
Family: *Mormoopidae*		
4 *Mormoops megalophylla*	10	135
5 *Pteronotus davyi*	8	106
6 *Pteronotus parnellii*	23	313
7 *Pteronotu personatus*	7	51
Family: *Phyllostomidae*		
8 *Artibeus jamaicensis*	11	82
9 *Desmodus rotundus*	6	38
10 *Leptonycteris yerbabuenae*	26	392
11 *Macrotus californicus*	6	53
12 *Sturnira ludovici*	12	71
Family: *Vespertilionidae*		
13 *Antrozous pallidus*	58	1937
14 *Eptesicus fuscus*	74	1589
15 *Idionycteris phyllotis*	6	177
16 *Lasiurus blossevillii*	10	90
17 *Lasiurus cinereus*	5	42
18 *Lasiurus xanthinus*	8	204
19 *Myotis volans*	8	140
20 *Myotis yumanensis*	5	89
21 *Pipistrellus hesperus*	85	2445

February to May 2013. We used 10 mist nets of 12, 9 and 6 m and 36 mm knot-to-knot mesh size erected at ground level (0-3 m) to capture bats. Nets were set 30 minutes before sunset and closed 30 minutes after sunrise to cover two of the most important bat activity peaks thus increasing capture success. Live-trapped bats were measured and identified to species level using field keys (Medellín et al., 2008; Ceballos and Oliva, 2005) and bat taxonomy followed Simmons (2008). We constructed an echolocation call library by recording the calls of captured individuals using two different techniques: 1) bats were recorded while released from the hand about 6 to 10 m from the bat detector in open areas and away from vegetation, 2) bats were tied to a zip-line and recorded while flying along the zip flight path. Echolocation calls were recorded with a Pettersson 1000x bat detector (Pettersson Elektronik AB, Uppsala, Sweden) and stored on a Sandisk 8 GB Extreme CF Compact Flash Card. The bat detector was set to manually record calls in real time, full spectrum at 500 KHz. Each recording consists of multiple calls from a single individual bat.

In total our dataset consists of 21 species, 449 individual bats and 8429 calls. Table 1 gives a summary of the dataset. Care must be taken when spliting the data to training and test sets during cross-validation in order to ensure that calls from the same individual bat recording are not in both sets. For that we split our dataset using recordings instead of calls. For species with less than 100 recordings we include as many calls as possible up to a maximum of 100 calls per species. The raw data as well as the post processed and 5-fold cross validation sets are available to download[3] as a supplementary material for this paper.

4.2 Experiments

We compare the classification accuracy of the multinomial probit regression with Gaussian process prior classifier using the three representations discussed in Sections 3.3-3.3. The values of the call shape parameters are normalised to have zero mean and standard deviation equal to one by subtracting the mean and dividing by the standard deviation of the call shape parameters in the training set. For the 33 covariance function parameters, σ^2 and $\lambda_1, \ldots, \lambda_{32}$ we use independent Gamma priors with shape parameter 1.5 and scale parameter 10. For the DTW representation each call vector of optimal alignment costs is normalised to unit length and independent Gamma (1.5, 10) priors are used for the magnitude and length-scale covariance function parameters. The weights for the linear combination of the DTW and call shape kernel functions in Equation (3.5) are restricted to be positive and sum to 1 and a flat Dirichlet prior is used.

As a baseline we also compare with a multi-class Support Vector Machine (SVM) classifier using the LibSVM software library (Chang and Lin, 2011). For the SVM we use the call shape parameters and the DTW representations with a squared exponential kernel. The kernel lengthscale parameter, γ, and the relaxation parameter, C, for the SVM where optimised using a held out validation data set and a grid search with values $C \in \{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{13}, 2^{3}\}$. Combining both kernel representations using a weighted sum would require optimising 4 parameters simultaneously with cross validation which is not straight-forward using grid search therefore we do not compare against this option.

[3] `http://www.engage-project.org/`

4.3 Results

Table 2 compares the misclassification rate of the three methods. Results are averages of a 5-fold cross validation. We can see that the DTW representation is significantly better for characterising the species variations achieving a better classification accuracy. However, results can be improved by also considering information from the call shape parameters. Moreover, the optimised weights for the kernel combination significantly favor the DTW covariance function with a weight of ≈ 0.8 in contrast to the call shape parameters with weight ≈ 0.2. If we fix the weight parameters to equal values we obtain a classification error rate of 0.22 ± 0.031 highlighting the importance of the DTW kernel matrix.

The independent length scales allow us also to interpret the discriminatory power of the call shape parameters. In our experiments, the frequency at the center of the duration of a call, the characteristic call frequency (Determined by finding the point in the final 40% of the call having the lowest slope or exhibiting the end of the main trend of the body of the call), as well as the start and end frequencies of the call have consistently obtained a small lengthscale parameter value indicating their importance in species discrimination. This coincides with expert knowledge on bat call shapes where these call shape parameters are extensively used for identifying species.

Table 2: Classification results, smaller values are better.

Method	Error rate	Std.
SVM shape parameters	0.26	$\pm$0.064
SVM DTW	0.25	$\pm$0.035
GP shape parameters	0.24	$\pm$0.052
GP DTW	0.21	$\pm$0.026
GP DTW + shape param.	**0.20**	**$\pm$0.037**

In Figure 4 the confusion matrix from the best classification results, 15% misclassification rate, are shown. There is an overall high accuracy for all classes with the exception of species *Lasiurus xanthinus*, class 18, which is often misclassified as *Antrozous pallidus*, class 13, which needs to be further investigated. In contrast, the very similar call shapes of the *Myotis* species are easily discriminated. Finally, misclassification rates are higher to within family species compared to species from other families indicating a common evolutionary path of bat echolocation.

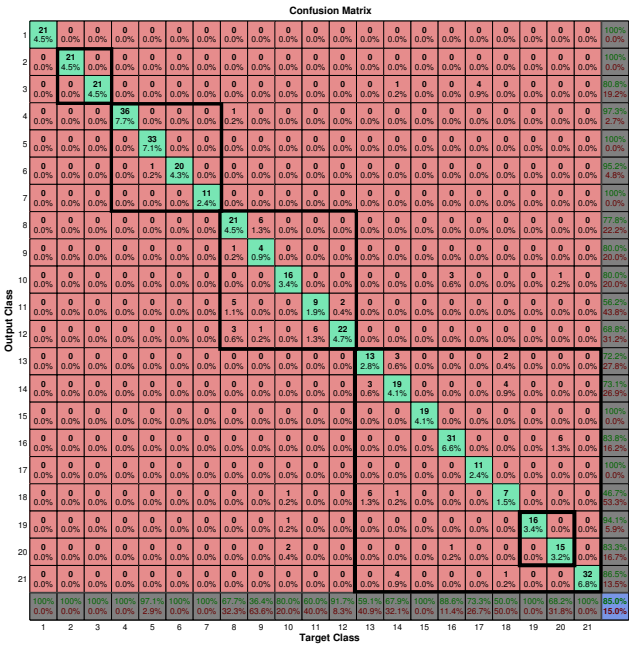

Figure 4: Confusion matrix of the best classification. Classes are in the same order and grouped as in Table 1

5 Conclusions and Future Work

Previous works highlight the complexity to discriminate species from the Phyllostomidae family, while others recognized *Myotis* species hard to classify as well. The high accuracy obtained in this study to separate species in the Phyllostomidae family from other families and its ability to discriminate between *Myotis* species sets the ground for a further development of an automatic identification tool for Mexican bats. Although only a small set of Mexican bat species was used in this study, it suggests promising applications to a bigger set of species. Despite these limitations, the development of a national call library of full-spectrum calls together with the echolocation classification tool will set the foundations to establish a long-term National Bat Acoustic Monitoring Program. This is a feasible alternative for developing countries to create biodiversity monitoring programs and develop volunteer networks since they are easier and less costly to implement at broad scales and on a long term basis as compared to other monitoring techniques.

Acknowledgments

VS, MAG, and KEJ are grateful to the UK Engineering and Physical Sciences Research Council (EPSRC) for supporting this research through project grant EP/K015664/1. MAG is grateful for support by an EPSRC Established Research Fellowship

EP/J016934/1 and a Royal Society Wolfson Research Merit Award.

VZG would like to thank all the people that kindly provided field and logistical support. Financial support was provided by the CONACYT (No. 310731), Cambridge Commonwealth European and International Trust (No. 301879989), The Rufford Small Grants Foundation (No. 12059-1), American Society of Mammalogists, Bat Conservation International, Idea Wild and The Whitmore Trust. A collecting permit was granted by SEMARNAT, Mexico (No. 03374).

References

I. Ahlen and H. Baage. Use of ultrasound detectors for bat studies in Europe : experiences from field identification , surveys , and monitoring. *Acta Chiropterologica*, 1:137–150, 1999.

G. Ceballos and G. Oliva. *Los mamíferos silvestres de México*. CONABIO UNAM Fondo de Cultura Econmica, Mexico, DF, 2005.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2: 27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

T. Damoulas, S. Henry, A. Farnsworth, M. Lanzone, and C. Gomes. Bayesian classification of flight calls with a novel dynamic time warping kernel. In *Proceedings of the 2010 Ninth International Conference on Machine Learning and Applications*, ICMLA '10, pages 424–429, Washington, DC, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4300-0. doi: 10.1109/ICMLA.2010.69.

M. B. Fenton and G. P. Bell. Recognition of Species of Insectivorous Bats by their Echolocation Calls. *Journal of Mammology*, 62(2):233–242, 1981.

M. Girolami and M. Zhong. Data integration for classification problems employing gaussian process priors. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 465–472. MIT Press, Cambridge, MA, 2007.

Mark Girolami and S. Rogers. Variational bayesian multinomial probit regression with gaussian process priors. *Neural Computation*, 18(8):1790–1817, 2006. Probit regression; gaussian process, variational bayes, multi-class classification.

l. Hansheng and S. Bingyu. A study on the dynamic time warping in kernel machines. In *Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on*, pages 839–845, 2007. doi: 10.1109/SITIS.2007.112.

G. Jones and E. Teeling. The Evolution of echolocation in bats. *Trends in Ecology and Evolution*, 21:149–156, 2006.

G. Jones, D. Jacobs, T. Kunz, M. Willig, and P. Racey. Carpe noctem: the importance of bats as bioindicators. *Endangered Species Research*, 8:93–115, 2009.

M. Kuss and C. E. Rasmussen. Assessing approximate inference for binary gaussian process classification. *Journal of Machine Learning Research*, pages 1679–1704, 2005.

R. A. Medellín, H. Arita, and O. Sánchez. *Identificación de los murciélagos de México: Clave de campo*. Publicaciones Especiales. Asociación Mexicana de Mastozoología A. C., Mexico, DF, 2008.

T. Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)*, pages 362–369, San Francisco, CA, 2001. Morgan Kaufmann.

K. Murray, E. Britzke, and L. Robbins. Variation in search phase calls of bats. *Journal of Mammalogy*, 82:728–737, 2001.

M. K Obrist. Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. *Behavioral Ecology and Sociobiology*, 36: 207–219, 1995.

S. Parsons and G. Jones. Acoustic identication of twelve species of echolocating bat by discriminant function analysis and artifact neural networks. *Journal of Experimental Biology*, 203:2641–2656, 2000.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

J. Riihimaki, P. Jylanki, and A. Vehtari. Nested expectation propagation for gaussian process classification with a multinomial probit likelihood. *Journal of Machine Learning Research*, 14:75–109, 2013.

H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEE Transactions on Acoustics, Speech and Signal Processing*, 26:43–49, 1978.

H. U. Schnitzler, C.F Moss, and A. Denzinger. From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology and Evolution*, 18:386–394, 2003.

M. Seeger, D. L. Neil, and R. Herbrich. Efficient nonparametric bayesian modelling with sparse gaussian process approximations. Technical report, Max Planck Institute for Biological

Cybernetics, Tubingen, Germany, 2006. http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.101.3964.

N. B. Simmons. Order Chiroptera. In *Wilson, D.E., Reeder, D.M. (eds.), Mammal Species of the World: A Taxonomic and Geographic Reference, third ed.*, pages 312–529. Johns Hopkins University Press, 2008.

C. L. Walters, R. Freeman, A. Collen, c. Dietz, M. Brock Fenton, G. Jones, M. K. Obrist, S. J. Puechmaille, T. Sattler, B. M. Siemers, S. Parsons, and K. E. Jones. A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, 49(5):1064–1074, 2012. ISSN 00218901.