Bioacoustics: The International Journal of Animal Sound and its Recording
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbio20

A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features

Jinkui Cheng [a], Bengui Xie [a], Congtian Lin [a] & Liqiang Ji [a]
[a] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, 1-5 Beichenxi Road, Beijing, 100101, China
Version of record first published: 26 Mar 2012.

To cite this article: Jinkui Cheng , Bengui Xie , Congtian Lin & Liqiang Ji (2012): A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features, Bioacoustics: The International Journal of Animal Sound and its Recording, 21:2, 157-171

To link to this article: http://dx.doi.org/10.1080/09524622.2012.669664

Bioacoustics
Vol. 21, No. 2, June 2012, 157–171

Taylor & Francis
Taylor & Francis Group

A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features

Jinkui Cheng, Bengui Xie, Congtian Lin and Liqiang Ji*

Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, 1-5 Beichenxi Road, Beijing 100101, China

(*Received 6 August 2011; final version received 5 February 2012*)

Species- and individual-specific animal calls can be used in identification as verified in playback experiments and analyses of features extracted from these signals. The use of machine-learning methods and acoustic features borrowed from human speech recognition to identify animals at the species and individual level has increased recently. To date there have been few studies comparing the performances of these methods and features used for call-type-independent species and individual identification. We compared the performance of four machine-learning classifiers in the identification of ten passerine species, and individual identification for three passerines using two acoustic features. The methods did not require us to pre-categorize the component syllables in call-type-independent species and individual identification systems. The results of our experiment indicated that support vector machines (SVM) performed best generally, regardless of which acoustic feature was used, linear predictive coefficients (LPCs) increased the recognition accuracies of hidden Markov models (HMM) greatly, and the most appropriate classifiers for LPCs and Mel-frequency cepstral coefficients (MFCCs) were HMM and SVM respectively. This study will assist researchers in selecting classifiers and features to use in future species and individual recognition studies.

Keywords: call-type-independent; species identification; individual recognition; machine-learning; Mel-frequency cepstral coefficients; linear predictive coefficients; passerine

Introduction

Species and individual identification based on acoustic features of calls is a useful tool for the study and monitoring of animals, especially for species that cannot easily be marked using traditional methods (Terry et al. 2005). It can improve census estimates and offer unique information that can expand our knowledge of the population ecology of some species (Peake and McGregor 2001; Laiolo et al. 2007). It also helps researchers to identify new species and estimate the taxonomic position of species, especially those for which morphological and molecular data are scant (Tietze et al. 2008; Packert et al. 2009). Thus, interest in this field is on the rise.

Animal vocalizations have evolved to be species-specific. This has been demonstrated by many playback experiments (Nietsch and Kopp 1998; De Kort and Ten Cate 2001; Charrier and Sturdy 2005). Many animals use calls that contain information about their

*Corresponding author. Email: ji@ioz.ac.cn

ISSN 0952-4622 print/ISSN 2165-0586 online

http://dx.doi.org/10.1080/09524622.2012.669664
http://www.tandfonline.com

motivation, sex, age, emotion, and even identity when communicating with conspecifics (Soltis et al. 2005). Vocal recognition at the individual level has been verified by playback experiments in some species (Palleroni et al. 2006; Gasser et al. 2009; Wilson and Mennill 2010; Xia et al. 2010). Species or individual recognition can be determined for some species based upon the acoustic features of their calls.

The majority of studies in species and individual identification have used traditional acoustic features, including fundamental frequency, maximum frequency, minimum frequency, syllable energy, zero-crossing rate, pulse rate, signal bandwidth, mean syllable duration and mean interval duration (Sousa-Lima et al. 2002; Molnar et al. 2008). Examples of species recognition include the identification of two closely related African fish (Crawford et al. 1997), and the identification of nine dolphin species based on 12 variables found in their calls (Oswald et al. 2003). Individual identification has also been studied using traditional acoustic features in many species, including frogs (Bee et al. 2001; Friedl and Klump 2002), marmots (Blumstein and Munos 2005), skuas (Charrier et al. 2001), foxes (Darden et al. 2003), owls (Galeotti and Sacchi 2001; Lengagne 2001) and kittiwakes (Aubin et al. 2007). Recently, researchers applied acoustic features borrowed from the field of human speech recognition to identify species and individuals, which can be easily used for automatic species or individual identification systems. For example, this approach was applied to species groups in the passerine family and used mel-frequency cepstral coefficients (MFCCs) (Somervuo et al. 2006). MFCCs have also been used to identify frog species and cricket species (Lee et al. 2006) and individual African elephants (*Loxodonta africana*) (Clemins et al. 2005). Another approach is the use of linear predictive coefficients (LPCs) to distinguish species and individuals and has been applied to birds (Juang and Chen 2007) and pigs (*Sus scrofa*) (Schon et al. 2001). In general, MFCCs and LPCs are the most popular acoustic features for species and individual identification.

In early research on species and individual identification using acoustic features, discriminant function analysis was the dominant method for the classification and recognition of features and was applied to many taxa including insects (Lee et al. 2006), amphibians (Bee et al. 2001), birds (Lengagne 2001) and mammals (Darden et al. 2003; Oswald et al. 2003; Blumstein and Munos 2005). Interest in this field has risen and many machine-learning methods borrowed from the field of pattern recognition have been used to deal with the problem of species and individual identification. For example, Parsons (2001) used artificial neural networks (ANN) to identify bats from echolocation calls and Chesmore (2004) applied a similar tool when developing an automated identification system for insects. Support vector machines (SVM) have been applied to species recognition in frogs and birds (Fagerlund 2007; Acevedo et al. 2009), as SVM can provide equal or even better performance than traditional methods (Guo and Li 2003; Huang et al. 2009). Other research, focused on hidden Markov models (HMM) for species and individual identification, obtained good results (Clemins et al. 2005; Trawicki et al. 2005; Somervuo et al. 2006; Trifa et al. 2008). Gaussian mixture models (GMM) are also an important machine-learning method and have been used to recognize 28 species of bird and four dolphin species based on cepstral features extracted from calls (Roch et al. 2007; Lee et al. 2008). In general, neural networks (NN), SVM, HMM and GMM are the most common machine-learning methods used in species and individual recognition. Studies using machine-learning methods obtained the greatest recognition accuracy. For species and individual recognition, only a few studies have combined machine-learning methods and the two acoustic features but have achieved very high identification rates (Clemins et al. 2005; Trawicki et al. 2005; Fox et al. 2008; Trifa et al. 2008).

Bird song is typically divided into four hierarchical levels: notes, syllables, phrases and song. The majority of current research is based on the similarity between syllable-type-specific features (Fox 2008). This is referred to as call-type-dependent recognition; however, given changeability in animal vocalizations between and within species and individuals, automatic call-type-independent identification is of greater use. Call-type-independent means that the models for bird species or individuals can be trained using any types of syllables and tested using the same or different types. The identification models are insensitive to the type of syllables. Although very few call-type-independent studies have been completed, accuracies have been very high (Fox et al. 2008; Cheng et al. 2010). Further, little research has been done on comparing the performance of different machine-learning methods using acoustic features and comparing the ability of acoustic features in call-type-independent species and individual identification. Here, we aim to compare the performance of machine-learning methods and the suitability of using acoustic features for call-type-independent species and individual identification. To achieve this we chose 10 passerine species and four machine-learning methods: radial basis function networks (RBFN, a special kind of ANN), SVM, HMM and GMM; and two acoustic features: MFCCs and LPCs. Our approach is call-type-independent and aimed at determining the optimum method-feature pair.

Materials

Call-type-independent individual identification was carried out in three passerines: Chinese leaf warbler (*Phylloscopus yunnanensis*), Hume's warbler (*Phylloscopus humei*) and Chinese bulbul (*Pycnonotus sinensis*). Species identification was conducted using 10 passerine species: Chinese leaf warbler, Hume's warbler, Chinese bulbul, Gansu leaf warbler (*Phylloscopus kansuensis*), Black bulbul (*Hypsipetes leucocephalus*), Red-whiskered bulbul (*Pycnonotus jocosus*), Brown-breasted bulbul (*Pycnonotus xanthorrhous*), Chestnut bulbul (*Hemixos castanonotus*), Collared finchbill (*Spizixos semitorques*) and Mountain bulbul (*Hypsipetes mcclellandii*).

Chinese leaf warblers were recorded from Taibaishan National Nature Reserve (33°49′30″ ∼ 34°05′35″N, 107°22′25″ ∼ 107°51′30″E), Gansu leaf warblers and Hume's warblers were recorded from Lianhuashan National Nature Reserve (34°56′ ∼ 34°58′N, 103°44′ ∼ 103°48′), and Chinese bulbuls were recorded from several locations across Fujian, Guangxi and Jiangsu provinces, China. Data from the remaining six Pycnonotidae species were recorded from different provinces within China. All recordings were conducted during the breeding seasons of the species using a WM-D6C professional recorder (Sony Corporation, Tokyo, Japan) with a MKH 416-P48 directional microphone (Sennheiser, Wedemark, Germany) placed 2–8 m from a singing bird. The recordings contained a mixture of bird songs and calls. Recordings were converted to a digital medium at 22.05 kHz sampling frequency and saved in 16-bit wave format using Batsound v3.10 (Pettersson Elektronik AB, Uppsala, Sweden).

Recordings for individual recognition of one species were all recorded from different times and locations, thus we identified individuals easily. In order to avoid the classifiers picking up the differences in acoustic background at different recording sites, we did our best to record and select clean recordings with little noise. Recordings for training and testing datasets were more than 15 seconds and 2 seconds in length, respectively. The records of each bird species contain 1–4 syllable types which were all used for species or individual identification in our experiment. The syllable types were sampled equally for the training and testing data sets (see Tables 1 and 2), but the individual birds and recording locations in the training and testing sets were different. The spectrogram of the

Table 1. The segmentation results of sounds of ten species for species recognition.

Species	Data set	Number of syllable types	Number of sound samples
Chinese leaf warbler	Training	2	27
	Testing	2	12
Hume's warbler	Training	1	26
	Testing	1	13
Gansu leaf warbler	Training	1	9
	Testing	1	6
Chinese bulbul	Training	4	24
	Testing	4	12
Black bulbul	Training	2	4
	Testing	2	4
Red-whiskered bulbul	Training	2	4
	Testing	2	4
Brown-breasted bulbul	Training	1	6
	Testing	1	4
Chestnut bulbul	Training	1	4
	Testing	1	2
Collared finchbill	Training	1	4
	Testing	1	2
Mountain bulbul	Training	2	6
	Testing	2	4

Table 2. The segmentation results of each individual's sounds for individual recognition.

Species	Data set	Number of individuals	Number of syllable types	Sound samples for each individual
Chinese leaf warbler	Training	15	2	3
	Testing	15	2	2
Hume's warbler	Training	26	1	2
	Testing	26	1	1
Chinese bulbul	Training	14	4	3
	Testing	14	4	2

syllable types of each species is shown in Figure 1. Because it is difficult to divide bird songs of different species into syllables with a single standard, the syllable types shown in Figure 1 sometimes represent phrases that can be divided into syllables further.

Methods

The architecture of our acoustic-driven call-type-independent species and individual recognition system for birds was divided into three modules: signal preprocessing, feature extraction and classification (see Figure 2). In the signal preprocessing stage we segmented the signals using an energy threshold and emphasized the high frequency of the signal using a digital filter.

Features

Although bird calls are produced by the syrinx, unique to birds (King 1989), the source-filter model, which is popular in speech and speaker recognition is also apt for explaining the

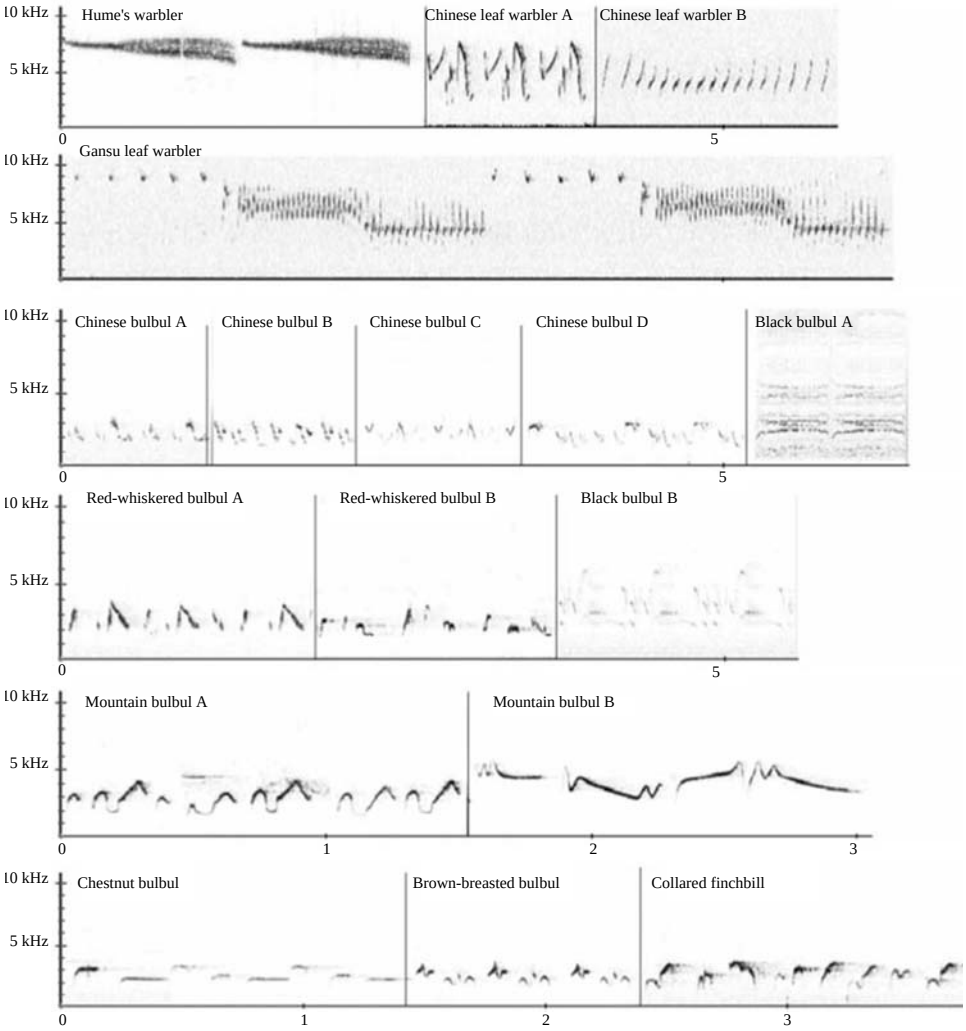

Figure 1. Spectrogram of different syllable types of each species with time (s) and frequency axes.

production of bird calls. Thus, we can use this model and corresponding acoustic features such as MFCCs and LPCs borrowed from speech recognition to identify species and individuals in birds. Moreover, MFCCs and LPCs are the most popular features of existing speaker identification systems.

When attempting call-type-independent identification, we did not need to classify the syllables of a song before extracting feature vectors. However, before acoustic features were extracted, songs were segmented to remove blank and noisy segments. On the assumption that every sound record began with blank or background noisy frames, we calculated mean energy E of the first 10 frames of one recording and used $2E$ as the energy threshold to decide which frame remaining should be removed. Once segmented, sound signals (now consisting of syllables) were divided into two sets to train and test the classifiers. The segmented results for species identification in ten passerine species and for individual identification in three passerine species are shown in Tables 1 and 2 respectively. Then the sounds were

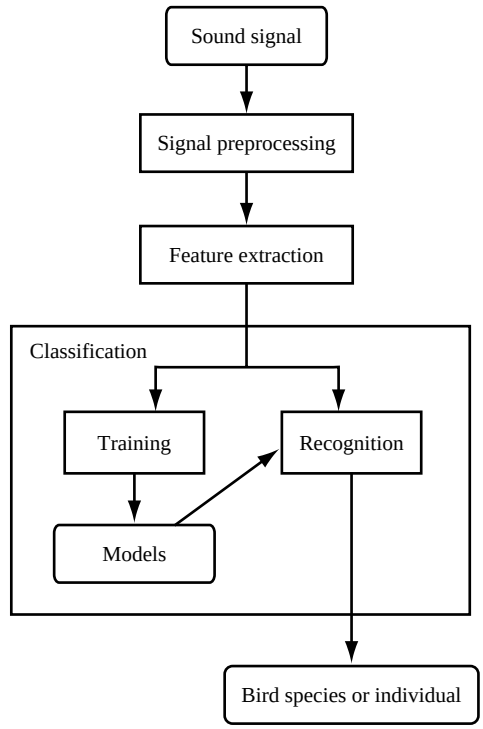

Figure 2. Architecture of our call-type-independent species and individual recognition system.

pre-emphasized to emphasize the high frequency of the signal, using a digital filter described by the formula

$$H(Z) = 1 - \mu z^{-1} \tag{1}$$

where μ is 0.95.

The sounds were divided into a set of overlapping frames, windowed using the Hamming window function, and finally the acoustic features were extracted and the sound signals were represented by feature vectors (see Figure 3). For each frame one feature vector was extracted; the values of this vector constitute an independent observation for the classifiers. We show how an exemplar recording was processed from the original sound signal to input vectors for the classifiers in Figure 4. In this paper the dimension of LPCs was optimized simply by applying a series of numbers from 4 to 30; we finally chose 13 for the best performance. In the same way we chose 24 as the dimension of MFCCs. The entire feature vector was used as the training or testing vector. MFCCs and LPCs were extracted from frames of the signal and the length of the frame was an important factor that may affect recognition performance. Figure 5 shows individual identification results with different window lengths using MFCCs and SVM in Chinese bulbuls. It also shows the results of different extraction times of LPCs in Hume's warbler using HMM. The best length of the extraction time for the two features should be 35 ms (Figure 5). In speech recognition the most popular window length of short-time features such as LPCs and MFCCs is 20–40 ms and is consistent with our tests on bird calls. Thus, the window length used in our experiment was 35 ms and the overlapping between successive frames was 17 ms, which was approximately half of the window length.

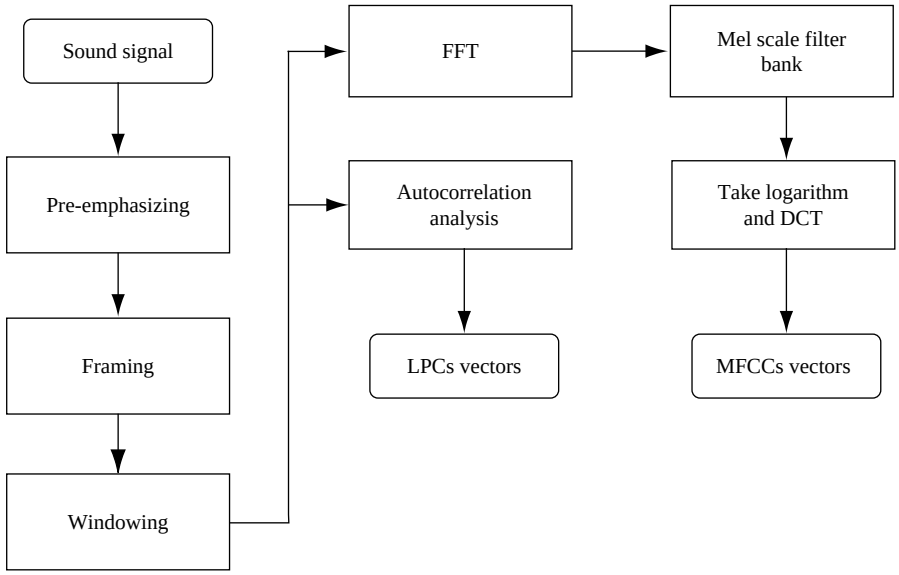

Figure 3. Acoustic feature extraction.

LPCs

The LPC model is an all-pole model of the spectrum which emphasizes the spectrum peaks and is called an autoregressive (AR) model. Usually, there exists a correlation between samples in a segment of acoustic data. LPCs attempt to encode an acoustic segment as a set of coefficients in a given equation and predict the value of a sample by a linear combination of values of previous samples:

$$\hat{S}(n) = \sum_{i=1}^{p} a_i s(n-i) \tag{2}$$

where $\hat{S}(n)$ is the predicted value of sample n, $s(n-i)$ is the value of sample $n-i$ in a frame, a_i is coefficient and p is the order of LPCs. Here p is 13.

The coefficients are estimated by minimizing the mean square error between the actual values of the samples and the values predicted by the equation above, which is described by formula (3) (Trifa et al. 2008).

$$\varepsilon^2(n) = \left[s(n) - \sum_{i=1}^{p} a_i s(n-i) \right]^2 \tag{3}$$

Here, $\varepsilon^2(n)$ is square error, $s(n)$ is the actual value of the sample that was predicted.

LPCs provide a good approximation of the vocal tract spectral envelope, and can be found by autocorrelation analysis according the procedure in Schon et al. (2001).

MFCCs

MFCC models are not dependent on the AR model. MFCC models the frequency spectrum and the vibration of the air column in the vocal tract. In the bird syrinx it is also the turbulent

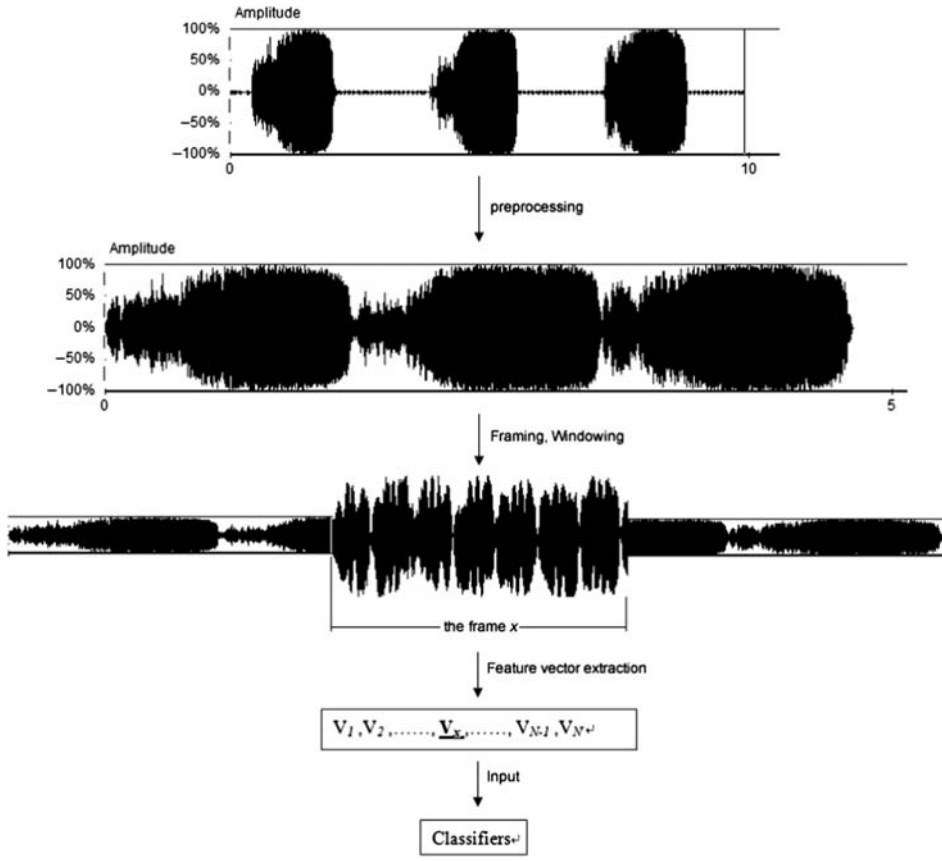

Figure 4. An exemplar process, from original sound signal to input vectors for the classifiers. V_x is the feature vector extracted from the xth frame.

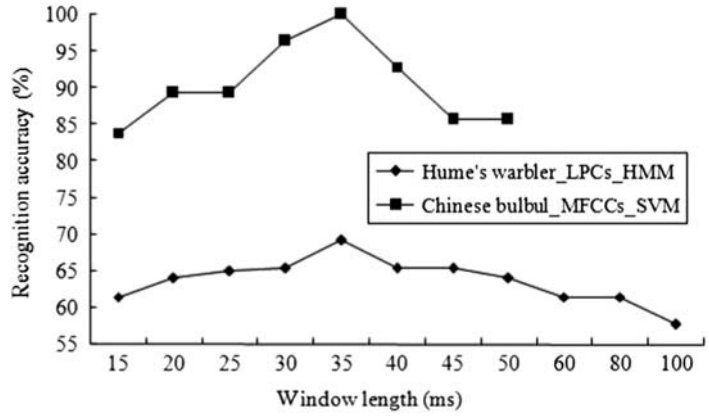

Figure 5. Individual recognition results with different window lengths using LPCs and HMM in Hume's warbler and using MFCCs and SVM in Chinese bulbul.

flow of air that produces the sound, thus we can use MFCCs to model bird vocalizations. MFCCs are popular because they tend to be uncorrelated and are computationally efficient. MFCCs, like psychoacoustical parameters, incorporate human perceptual information and have some resilience to noise (Clemins et al. 2005). The division of critical frequency band and the width of the critical frequency band used for calculating MFCCs also parallel psychoacoustical parameters. MFCCs can be computed by taking a discrete cosine transformation (DCT) of the logarithmic spectrum obtained by warping the signal to the Mel scale using a set of filters (Figure 3). Procedural detail can be found elsewhere (Cheng et al. 2010).

Classifiers

The most popular machine-learning classifiers are GMM, HMM, SVM and NN and these have performed well in species and individual identification. When features are extracted, a classifier should be trained to distinguish the feature sets and classify them into classes, then the trained classifier can classify a testing feature vector to one of the classes or a new class by comparing it with the stored reference templates for each class. All training and testing data sets for each model compared in this paper were the same. For individual identification we used training datasets of different individuals of one species to train the classifiers and testing datasets to test the performance of the classifiers. We repeated the procedure for each of the three species. For species identification we prepared training and testing datasets for each species, and then we trained and tested the classifiers using datasets of different species. Recordings used for individual recognition were also used for species; however, these recordings were split into different datasets for the two tasks. In the training stage we cut each training dataset into five smaller sets stochastically. Four sets were used to train the classifiers and the other one was used to test the classification performance of the trained classifiers. We repeated the procedure to optimize the parameters of classifiers and avoid the problem of over-fitting, which can be called '5-fold cross-validation'.

GMM

GMM is well known for modelling arbitrarily complex distributions with multiple components and is an effective classifier for many tasks (Roch et al. 2007). The feature distribution can be represented using a GMM based on a weighted sum of component Gaussian densities. Here, we assumed that the feature space was characterized by a set of broad acoustic classes and within each class the acoustic feature distribution was modelled by a component Gaussian density of a GMM (Tsai and Chang 2002). The parameters of the GMM were estimated using the expectation-maximization (EM) algorithm that guarantees a monotonic increase in likelihood (Biernacki 2007). In our experiment, we followed the procedure described in Reynolds and Rose (1995) to construct the GMMs, calculate the parameters of each GMM and test the GMMs. The number of Gaussian components was optimized simply by applying a series of numbers from 4 to 64. We found that the relationship between recognition accuracies and number of components can be represented by a bell-curve. Thus, we chose 16 Gaussian components in our experiment for the best performance.

HMM

The HMM can model a discrete-time dynamical system described by a Markov process with unknown parameters. It has two levels: (1) the hidden level, which consists of a finite number of states and transitions among the states with transition probabilities; and (2) the

level of observations, which are represented by a sequence of feature vectors that are assumed to be emitted from hidden states of some probability density (Kogan and Margoliash 1998). The feature vectors extracted from the frames of the signal at a discrete time result in a time series. HMM is trained to model the temporal evolution of the features of each class, and recognition can be done by looking for which HMM is the most likely to produce a given sequence of feature vectors (Trifa et al. 2008). Usually, the HMM topology is determined first and the training is performed so that the state label sequence corresponds to the observations. The commonly used Baum-Welch training utilizes the EM algorithm principle so that the state indices are considered as latent variables and the model parameters are trained using all possible state alignments weighted by their probabilities. HMM can be found and trained following the process described in Trifa et al. (2008). In this paper we used HMM to model the transformation of feature vectors extracted from successive signal frames. We used a HMM toolbox for MATLAB to train and test our models using successive feature vectors as the inputs to HMM (Murphy 2005). The number of hidden states was eight and the number of output values was N, the same as the dimension of feature vectors.

SVM

The SVM is a sophisticated kernel-based machine-learning classifier introduced by Vapnik (1999) that has attracted much attention as a new classification technique with good generalization ability (Cristianini and Shawe-Taylor 2000). The idea of SVM is to map input vectors into a high-dimensional feature space and linearly separate feature vectors with an optimal hyper-plane considering both the structural and empirical risks (Huang et al. 2009). The placement of the hyper-plane is based on the location of support vectors, which are the marginal samples. Because linear separation of call classes was not possible, we used a radial basis function (RBF) kernel to transform the feature space to enable the fitting of a maximum-margin hyper-plane. In our experiment we used the LIBSVM software, which is a professional and free toolbox for SVM training and testing (Chang and Lin 2001). Following the LIBSVM software guide, we trained the SVM with the default initial parameters and optimized the key parameters gamma (g) and cost (c) of each SVM using five-fold cross validation. In different recognition tasks, key parameters g and c were optimized to different values.

RBFN

RBFN is a special kind of ANN that contains only one hidden layer and has been widely used for function approximation and pattern recognition. The structure of RBFN was developed by several researchers (Broomhead and Lowe 1988; Moody and Darken 1989) and consists of three layers: input, hidden and output layers. The nodes of the hidden layer contain several radial basis functions; we used the Gaussian basis function. The most important parameters of RBFN are the centre of the radial basis function in the hidden layer and weights connecting the hidden layer with the output layer. The structure and the procedure of parameter learning can be seen in Choi et al. (2003). In our experiment, we used the neural network toolbox for MATLAB to train and optimize RBFN. The toolbox has many powerful functions to construct, train, optimize and test a RBFN easily. Each multi-dimension feature vector and the target class identifier of the vector were used as the inputs to the RBFN. In the training stage, the number of hidden neurons and weights connecting the hidden neurons with output neurons were optimized. Different tasks

obtained different parameters; for species recognition in this paper the number of hidden neurons was 2150 and the weights was a 2150 (the number of hidden neurons) by 10 (the number of target classes) matrix.

Results

We conducted call-type-independent species identification in ten passerine species and individual identification in three passerines. The song records were supplied by other research groups; they studied these passerine species for many years and recorded several songs of these species. The recognition results are shown in Tables 3 and 4, where the accuracy rate and classification accuracy were both the rate of correctly classified test samples. Overall, for species and individual identification, SVM performed best regardless of which feature was used. Only when the LPCs were used to identify the individuals in Chinese leaf warbler did HMM perform better than SVM. The accuracy rates for SVM were between 76.9% and 100%. RBFN obtained higher classification accuracies than GMM for all species and individual recognition tasks. The recognition accuracies for GMM ranged between 56.9% and 85.7% and the accuracies for RBFN were between 57.7% and 96.4%. Both GMM and RBFN performed better than HMM when MFCCs were used. But HMM obtained higher accuracy rates than GMM and RBFN when LPCs were used for individual recognition in Hume's warbler and Chinese leaf warbler. The performance of HMM was improved greatly when LPCs were used. Accuracy rates for HMM were between 23.1% and 40.0% when MFCCs were used, but the accuracy rates increased to 69.2% and 93.3% when LPCs were used. LPCs had a strong temporal relationship with each other but the MFCCs did not and HMM can model the temporal evolution of the features. We believe this explains the different recognition results of HMM across the two features.

From Tables 3 and 4, we also can see that for the classifiers SVM, GMM and RBFN MFCCs performed better than LPCs in all call-type-independent species and individual

Table 3. Call-type-independent species and individual recognition results of the four classifiers using the feature MFCCs.

		Individual recognition		
Model	Species recognition	Chinese bulbul	Hume's warbler	Chinese leaf warbler
SVM	87.3%	100.0%	80.8%	90.0%
GMM	56.9%	85.7%	57.7%	83.3%
RBFNN	70.6%	96.4%	65.4%	90.0%
HMM	35.3%	35.7%	23.1%	40.0%

Table 4. Call-type-independent species and individual recognition results of the four classifiers using the feature LPCs.

		Individual recognition		
Model	Species recognition	Chinese bulbul	Hume's warbler	Chinese leaf warbler
SVM	82.4%	100.0%	76.9%	90.0%
GMM	74.5%	57.1%	57.7%	70.0%
RBFNN	78.4%	75.0%	57.7%	73.3%
HMM	70.6%	75.0%	69.2%	93.3%

Table 5. Recognition results for each species in species recognition using MFCCs and SVM.

Species	Number of training samples	Number of testing samples	Correct classified samples	Correct rate
Chinese leaf warbler	27	12	11	91.7%
Hume's warbler	26	13	12	92.3%
Gansu leaf warbler	9	6	4	66.7%
Chinese bulbul	24	12	12	100.0%
Black bulbul	4	4	4	100.0%
Red-whiskered bulbul	4	4	2	50.0%
Brown-breasted bulbul	6	4	3	75.0%
Chestnut bulbul	4	2	2	100.0%
Collared finchbill	4	2	1	50.0%
Mountain bulbul	6	4	4	100.0%

recognition tasks, but for HMM LPCs performed better than MFCCs. When we used HMM combining LPCs, sometimes we obtained the best results, just as with the individual identification of the Chinese leaf warbler.

In order to see the effect of sample size on recognition results, we compared the correct rate of each species in species recognition using MFCCs and SVM (see Table 5). We found no evident relationship between sample sizes and recognition results. From Tables 2, 3 and 4 we can see that the sample sizes of Hume's warbler were equal or larger than the Chinese leaf warbler and Chinese bulbul; however, individual recognition accuracies in Hume's warbler were lower than the other two species. Only one syllable type was contained in recordings of Hume's warbler, but more than two types were used for individual recognition in the Chinese leaf warbler and Chinese bulbul (see Figure 1). Thus, the variability of songs between individuals in Hume's warbler was smaller than the Chinese leaf warbler and Chinese bulbul. Based on our results we think that the variability of songs between individuals may play a more important role than sample size in the recognition tasks.

Discussion

In this paper we compared the performances of four machine-learning methods for the call-type-independent identification of ten passerine species and individual identification of three passerine species using two acoustic features. SVM showed the highest identification accuracy in almost all species and individual identification tasks, which demonstrated the potential power of SVM in pattern recognition again, but when LPCs were used to identify the individuals in the Chinese leaf warbler, HMM showed the best performance. Ober and Armitage (2010) compared four supervised learning methods in the classification of bat echolocation calls using traditional acoustic features and found that SVM performed second best in general and best for some classification tasks. In the comparison of three machine learning techniques for automated classification of bird and amphibian calls, SVM obtained the highest average true positive rate and lowest average false positive rate (Acevedo et al. 2009). The recognition results of our experiment were basically in accordance with the results of these previous studies. Our results indicate that the most appropriate classifier for MFCCs was SVM; however, GMM and RBFN also obtained good recognition accuracy for both call-type-independent species and individuals. MFCCs were better than LPCs for SVM, GMM and RBFN according to the recognition results, which was in accord with the conclusion of a study of call-type-dependent automated species identification of antbirds in a Mexican rainforest (Trifa et al. 2008). But for HMM, LPCs were better than MFCCs and

they greatly improved the recognition accuracy of HMM. This is because LPCs have a strong temporal relationship but the MFCCs do not and HMM can model the temporal evolution of the LPCs. Thus, for HMM the most appropriate acoustic features are LPCs.

The works of Fox et al. (2008) were among the first attempts to identify individuals of bird species based on call-type-independent acoustic features. They used a multilayer perceptions classifier and MFCCs and achieved accuracies of 69.3% to 97.1% in three passerine species (Fox et al. 2008). The best call-type-independent individual identification results of our experiment in three different passerine species were between 80.8% and 100%. Although the syllable types were sampled equally for the training and testing data sets, we did not classify the syllables into different types and extract syllable-type-specific features, nor did we use the similarities of syllable-type-specific features among different bird species and individuals to identify bird species and individuals. Thus, we think the results obtained in our experiments are call-type-independent. Call-type-independent species or individual identification, although sometimes resulting in slightly lower accuracy than call-type-dependent identifications, has the huge advantage of being applicable to all species regardless of the amount of shared vocalizations or temporal changes in vocal repertoires (Fox et al. 2008).

The type of classifier and feature used will of course depend on the task but the results presented here will help researchers to determine which classifier and feature should be used in their research and will aid the future development of species and individual recognition systems.

Acknowledgements

We thank the Avian Ecology Group, Key Laboratory of Animal Ecology and Conservation Biology, Ornithological Research Group, Key Laboratory of Zoological Systematics and Evolution (all at the Institute of Zoology, Chinese Academy of Sciences) for recordings of birds. Thanks to Jinlin Li, Nan Lu and Xiaoying Xing. We are also grateful to members of the Biodiversity Informatics Group for their support, especially Jiangning Wang and Huijie Qiao.

References

Acevedo MA, Corrada-Bravo CJ, Corrada-Bravo H, Villanueva-Rivera LJ, Aide TM. 2009. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecological Informatics 4:206–214.

Aubin T, Mathevon N, Staszewski V, Boulinier T. 2007. Acoustic communication in the Kittiwake *Rissa tridactyla*: potential cues for sexual and individual signatures in long calls. Polar Biology 30:1027–1033.

Bee MA, Kozich CE, Blackwell KJ, Gerhardt HC. 2001. Individual variation in advertisement calls of territorial male green frogs, *Rana clamitans*: Implications for individual discrimination. Ethology 107:65–84.

Biernacki C. 2007. Degeneracy in the maximum likelihood estimation of univariate Gaussian mixtures for grouped data and behaviour of the EM algorithm. Scandinavian Journal of Statistics 34:569–586.

Blumstein DT, Munos O. 2005. Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls. Animal Behaviour 69:353–361.

Broomhead DS, Lowe D. 1988. Multivariable functional interpolation and adaptive networks. Complex Systems 2:321–355.

Chang C-C, Lin C-J. 2001. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:1–27.

Charrier I, Jouventin P, Mathevon N, Aubin T. 2001. Individual identity coding depends on call type in the South Polar skua *Catharacta maccormicki*. Polar Biology 24:378–382.

Charrier I, Sturdy CB. 2005. Call-based species recognition in black-capped chickadees. Behavioural Processes 70:271–281.

Cheng JK, Sun YH, Ji LQ. 2010. A call-independent and automatic acoustic system for the individual recognition of animals: a novel model using four passerines. Pattern Recognition 43:3846–3852.
Chesmore D. 2004. Automated bioacoustic identification of species. Anais Da Academia Brasileira De Ciencias 76:435–440.
Choi SW, Lee DW, Park JH, Lee IB. 2003. Nonlinear regression using RBFN with linear submodels. Chemometrics and Intelligent Laboratory Systems 65:191–208.
Clemins PJ, Johnson MT, Leong KM, Savage A. 2005. Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. Journal of the Acoustical Society of America 117:956–963.
Crawford JD, Cook AP, Heberlein AS. 1997. Bioacoustic behavior of African fishes (Mormyridae): Potential cues for species and individual recognition in Pollimyrus. Journal of the Acoustical Society of America 102:1200–1212.
Cristianini N, Shawe-Taylor J. 2000. The learning methodology. In: Cristianini N, Shawe-Taylor J, editors. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press. p. 1–8.
Darden SK, Dabelsteen T, Pedersen SB. 2003. A potential tool for swift fox (*Vulpes velox*) conservation: Individuality of long-range barking sequences. Journal of Mammalogy 84:1417–1427.
De Kort SR, Ten Cate C. 2001. Response to interspecific vocalizations is affected by degree of phylogenetic relatedness in Streptopelia doves. Animal Behaviour 61:239–247.
Fagerlund S. 2007. Bird species recognition using support vector machines. Eurasip Journal on Advances in Signal Processing Process 2007:1–8.
Fox EJ. S. 2008. A new perspective on acoustic individual recognition in animals with limited call sharing or changing repertoires. Animal Behaviour 75:1187–1194.
Fox EJ. S, Roberts JD, Bennamoun M. 2008. Call-independent individual identification in birds. Bioacoustics 18:51–67.
Friedl TW. P, Klump GM. 2002. The vocal behaviour of male European treefrogs (*Hyla arborea*): Implications for inter- and intrasexual selection. Behaviour 139:113–136.
Galeotti P, Sacchi R. 2001. Turnover of territorial Scops Owls *Otus scops* as estimated by spectrographic analyses of male hoots. Journal of Avian Biology 32:256–262.
Gasser H, Amezquita A, Hodl W. 2009. Who is calling? Intraspecific call variation in the aromobatid frog *Allobates femoralis*. Ethology 115:596–607.
Guo GD, Li SZ. 2003. Content-based audio classification and retrieval by support vector machines. IEEE Transactions on Neural Networks 14:209–215.
Huang CJ, Yang YJ, Yang DX, Chen YJ. 2009. Frog classification using machine learning techniques. Expert Systems with Applications 36:3737–3743.
Juang CF, Chen TM. 2007. Birdsong recognition using prediction-based recurrent neural fuzzy networks. Neurocomputing 71:121–130.
King AS. 1989. Functional anatomy of the syrinx. In: King AS, McLelland J, editors. Form and function in birds. Vol. 4. London: Academic Press. p. 105–192.
Kogan JA, Margoliash D. 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. Journal of the Acoustical Society of America A 103:2185–2196.
Laiolo P, Vogeli M, Serrano D, Tella JL. 2007. Testing acoustic versus physical marking: two complementary methods for individual-based monitoring of elusive species. Journal of Avian Biology 38:672–681.
Lee CH, Chou CH, Han CC, Huang RZ. 2006. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. Pattern Recognition Letters 27:93–101.
Lee CH, Han CC, Chuang CC. 2008. Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. IEEE Transactions on Audio Speech and Language Processing Process 16:1541–1550.
Lengagne T. 2001. Temporal stability in the individual features in the calls of eagle owls (*Bubo bubo*). Behaviour. 138:1407–1419.
Molnar C, Kaplan F, Roy P, Pachet F, Pongracz P, Doka A, Miklosi A. 2008. Classification of dog barks: a machine learning approach. Animal Cognition 11:389–400.
Moody TJ, Darken CJ. 1989. Fast learning in networks of locally tuned processing units. Neural Computation 1:281–294.

Murphy K. 2005. Hidden Markov Model (HMM) Toolbox for Matlab. http://www.cs.ubc.ca/ ~ murphyk/Software/HMM/hmm.html.
Nietsch A, Kopp ML. 1998. Role of vocalization in species differentiation of Sulawesi tarsiers. Folia Primatologica 69:371–378.
Ober HK, Armitage DW. 2010. A comparison of supervised learning techniques in the classification of bat echolocation calls. Ecological Informatics 5:465–473.
Oswald JN, Barlow J, Norris TF. 2003. Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean. Marine Mammal Science 19:20–37.
Packert M, Martens J, Severinghaus LL. 2009. The Taiwan Firecrest (*Regulus goodfellowi*) belongs to the Goldcrest assemblage (*Regulus regulus* s. l.): evidence from mitochondrial DNA and the territorial song of the Regulidae. Journal of Ornithology 150:205–220.
Palleroni A, Sproul C, Hauser MD. 2006. Cottontop tamarin, *Saguinus oedipus*, alarm calls contain sufficient information for recognition of individual identity. Animal Behaviour 72:1379–1385.
Parsons S. 2001. Identification of New Zealand bats (*Chalinolobus tuberculatus* and *Mystacina tuberculata*) in flight from analysis of echolocation calls by artificial neural networks. Journal of Zoology 253:447–456.
Peake TM, McGregor PK. 2001. Corncrake *Crex crex* census estimates: a conservation application of vocal individuality. Animal Biodiversity and Conservation 24:81–90.
Reynolds DA, Rose RC. 1995. Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing 3:72–83.
Roch MA, Soldevilla MS, Burtenshaw JC, Henderson EE, Hildebrand JA. 2007. Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. Journal of the Acoustical Society of America 121:1737–1748.
Schon PC, Puppe B, Manteuffel G. 2001. Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). Journal of the Acoustical Society of America 110:1425–1431.
Soltis J, Leong K, Savage A. 2005. African elephant vocal communication II: rumble variation reflects the individual identity and emotional state of callers. Animal Behaviour 70:589–599.
Somervuo P, Harma A, Fagerlund S. 2006. Parametric representations of bird sounds for automatic species recognition. IEEE Transactions on Audio Speech and Language Processing Process 14:2252–2263.
Sousa-Lima RS, Paglia AP, Da Fonseca GA. B. 2002. Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia). Animal Behaviour 63:301–310.
Terry AM. R, Peake TM, McGregor PK. 2005. The role of vocal individuality in conservation. Frontiers in Zoology 2:10.
Tietze DT, Martens J, Sun YH, Packert M. 2008. Evolutionary history of treecreeper vocalisations (Aves: Certhia). Organisms Diversity & Evolution 8:305–324.
Trawicki MB, Johnson MT, Osiejuk TS. 2005. Automatic song-type classification and speaker identification of Norwegian Ortolan Bunting (*Emberiza hortulana*) vocalizations. 2005 IEEE Workshop on Machine Learning for Signal Processing (MLSP). p. 277–282.
Trifa VM, Kirschel AN. G, Taylor CE, Vallejo EE. 2008. Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. Journal of the Acoustical Society of America 123:2424–2431.
Tsai WH, Chang WW. 2002. Discriminative training of Gaussian mixture bigram models with application to Chinese dialect identification. Speech Communication 36:317–326.
Vapnik V. 1999. Methods of pattern recognition. In: Vapnik V, editor. The nature of statistical learning theory. New York: Springer-Verlag. p. 123–167.
Wilson DR, Mennill DJ. 2010. Black-capped chickadees, *Poecile atricapillus*, can use individually distinctive songs to discriminate among conspecifics. Animal Behaviour 79:1267–1275.
Xia CW, Xiao H, Zhang YY. 2010. Individual variation in brownish-flanked bush warbler songs. Condor 112:591–595.