Ecological Informatics 35 (2016) 61–70

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Identification of European woodpecker species in audio recordings from their drumming rolls

Juliette Florentin [a,*], Thierry Dutoit [b], Olivier Verlinden [a]

[a] Department of Theoretical Mechanics, Dynamics and Vibrations, University of Mons, Mons, Belgium
[b] Department of Circuit Theory and Signal Processing, University of Mons, Mons, Belgium

ARTICLE INFO

Article history:
Received 10 May 2016
Received in revised form 23 August 2016
Accepted 28 August 2016
Available online 30 August 2016

Keywords:
Woodpeckers
Drumming
Species identification
Classification
t-SNE dimensionality reduction
Acoustic features

ABSTRACT

Drumming sounds are substantial clues when searching audio recordings for the presence of woodpeckers. Woodpeckers use drumming for territory defence and mate attraction to such an extent that some species have no vocalisations for these functions. This implies that drumming bears species markers. This hypothesis stands at the root of our project to develop an autonomous program for the identification of drumming species. To proceed, we assembled a database of 361 recordings from open-access bird sound archives. The recordings were for nine drumming species found on the European continent. Focusing on the signal below 1500 Hz, we reviewed all audio files and extracted 2665 drumming rolls. For recordings where multiple birds were present, the drumming rolls were attributed to individual birds. This allowed keeping track of the time interval between successive rolls. The characteristic traits of drumming are decidedly temporal. Consequently, the spectral features that have been successful in other recent bird identification studies were not applicable in our case. We developed specialized drumming parameters and automated their calculation. We then performed a t-SNE dimensionality reduction to visualise the dataset and to demonstrate that our parameters detached the different classes properly. Eventually, a k-NN algorithm accurately labelled 87.2% of the submitted test samples. The time structure within the drumming rolls (speed, acceleration) provided the critical features. The duration of the rolls followed in importance. The results compare well to existing literature and attest to the feasibility of monitoring European woodpecker species by tracking drumming.

1. Introduction

The continuous acoustic monitoring of wildlife and birds in particular has generated terabytes of data (Aide et al., 2013; Jahn et al., 2013; Towsey et al., 2014; Florentin et al., 2015), of which a large part is still awaiting exhaustive processing. The dormant information pertains to the presence or absence of species within certain areas, the evolution of bird communities over seasons and the potential degradation of habitats (Farina et al., 2011). Hence, in recent years, researchers have steered towards the development of robust algorithms that would be capable of identifying all species and most notably all bird species captured on audio recordings (Blumstein et al., 2011). Such algorithms are required to perform well on the two critical functions of 1) detecting bird sounds in the audio stream and 2) identifying the species emitting these sounds. The work we present in the present paper primarily addresses the second function, species identification, in a context where short extracts of bird sounds have already been made available.

It must be noted however that birdsong detection techniques sometimes inherently include the species identification step. This is the case for spectrogram cross-correlation. This well-established technique allows searching long audio recordings for one specific sound. It is well-suited to sounds that produce repeatable spectrogram patterns, such as stereotyped songs (Ulloa et al., 2016). The concept is to have a template of the target spectrogram image sliding over a continuous audio stream until a maximum in cross-correlation is reached. Swiston and Mennill (2009) used it to detect double-knocks from two species of woodpeckers (*Campephilus guatemalensis* and *Campephilus principalis*). There was no expectation that potential species traits in double-knocks would play a role and the same template was used for both species. Mean detection rates of 24% and 8% respectively were achieved, while the number of false positives was 77 times greater than with human observers.

If species identification is to be addressed separately, then its object is to automate the classification of short sound extracts lasting between a few seconds and a few minutes. Recent works in this area engaged in the identification of hundreds of bird species in collections of up to ten thousand audio files (Potamitis, 2014; Stowell and Plumbley, 2014; Lasseck, 2015). For this type of task, the choice of classifier evolved from template comparisons (Kogan and Margoliash, 1998) towards

* Corresponding author at: Department of Theoretical Mechanics, Dynamics and Vibrations, University of Mons, Bd Dolez 31, 7000 Mons, Belgium.
E-mail address: juliette.florentin@umons.ac.be (J. Florentin).

http://dx.doi.org/10.1016/j.ecoinf.2016.08.006
1574-9541/

hidden Markov models (Somervuo et al., 2006; Brandes, 2008), artificial neural networks (Fox et al., 2008; Ranjard and Ross, 2008) and finally single- or multi-label random forests (Potamitis, 2014; Stowell and Plumbley, 2014). The most common acoustic features are the Short-Time Fourier Transforms (STFT), either in full (Stowell and Plumbley, 2014) or compressed in the form of Mel-Frequency Cepstral Coefficients (MFCC) (Kogan and Margoliash, 1998; Somervuo et al., 2006; Fox et al., 2008; Ranjard and Ross, 2008; Lee et al., 2013). The time dimension is then handled by calculating the mean and standard deviation of the acoustic features over the duration of the sound extract (Stowell and Plumbley, 2014) or other statistics such as moments and regressions (Lasseck, 2015). Spectrogram images are an alternative possible basis for acoustic features (Lee et al., 2013; Potamitis, 2014; Lasseck, 2015). Potamitis (2014) and Lasseck (2015) derived their features from cross-correlation scores with a set of template images. Lasseck (2015) showed that these outperform spectral features in the classification of syllables and elements of songs. Multiple authors reported difficulties pertaining to the "variability" of songs (Kogan and Margoliash, 1998; Somervuo et al., 2006; Potamitis, 2014). The following species were offered as difficult cases: canaries (Kogan and Margoliash, 1998), Finnish blackcaps and pied flycatchers (Somervuo et al., 2006) and European tits (Potamitis, 2014). All are passerines, i.e. the order with the most elaborate songs. A typical percentage of correct identifications for passerines is 70%–80% (Somervuo et al., 2006; Brandes, 2008; Fox et al., 2008).

Overall, the success of the above classification studies is contingent on the existence of generic acoustic features that can grasp the species traits in the vocalisations of any species. On the other hand, Kogan and Margoliash (1998) suggested that their classification results would have been improved by the use of biological features. These are the features that birds themselves use to recognize their conspecifics. They are specific to species or to a group of species (Catchpole and Slater, 2008). Lasseck (2015) reselected a subset of features independently for each species to optimize the classification score. Bardeli et al. (2010) advocated redefining the features for each new species to improve the recognition rate. Ulloa et al. (2016) observed that spectrogram cross-correlation was not appropriate for all vocalisations and Lasseck (2015) that it did not render temporal structures and repetition rates in bird songs. All these comments reflect a need to fall back on differentiated features. The consequence is that it is more realistic to mine audio streams for the presence of a predetermined limited set of species than to aim at an exhaustive analysis. This sets the philosophy behind the present work: classification algorithms need to be built from the bottom-up, starting from sub-groups of species. We chose to target European woodpeckers.

2. Materials

2.1. European woodpeckers

Woodpeckers are members of the *Picidae* family within the order of the Piciformes. A list of the eleven species present on the European continent is found in Table 1.

Dendrocopos minor and *Jynx torquilla* are the smallest specimens, 14–16 cm and 16–19 cm in length, respectively. *D. martius* is the largest with a length of 45–50 cm. Hybrids exist between *D. major* and *D. syriacus*, between *Picus canus* and *P. viridis* and between *P. sharpei* and *P. viridis*. Until 2012, *P. sharpei* was a sub-species of *P. viridis* (Gorman, 2014). The reasons that make woodpeckers an interesting target for acoustic monitoring are plenty. They are valued as ecosystem keystones (Gorman, 2014) and indicators of forest health (Mikusinski and Angelstam, 1998). Some species are targeted by regional conservation programs, e.g. *D. medius* in Sweden (Pettersson, 1985) or *P. canus* in Belgium.

Woodpeckers have relatively simple vocalisations (Gorman, 2014). Their peculiarity is their use of drumming for territory marking and mate attraction (Zabka, 1980; Tremain et al., 2008). Some species altogether forego the use of vocal signals for these functions (*D. major*, *D. syriacus*, *D. leucotos*, *Picoides tridactylus*) (Table 1). It follows that drumming sounds have been thought to carry the species and individual information (Zabka, 1980; Dodenhoff et al., 2001). Only *J. torquilla* does not use drumming. For the other species, both sexes drum. Male-female pairs have synchronized drumming duets during the mating season. Drumming contests also occur between neighbours and males of different species. Drumming Rolls (DRs) are easily recognizable in spectrograms. The example in Fig. 1 shows the time parameters of drumming: the time between DRs, the time between strokes and the DR duration. These traits are straightforward features to classify DRs. Woodpeckers also produce isolated strokes (while foraging or digging holes) and demonstrative tapping. The latter consists in constant-speed rolls which are shorter, slower and quieter than drumming rolls (Zabka, 1980). Demonstrative tapping is hypothesized to achieve nearby communication (Zabka, 1980; Tremain et al., 2008) whereas drumming would primarily be for long-distance communication (Zabka, 1980; Stark et al., 1998). Our analysis focuses on drumming only.

Zabka (1980) ran a previous study on woodpecker drumming using 240 DRs from the same nine species as in the present work, i.e. all but *J. torquilla* and *P. sharpei*. *D. major* made up almost half of his collection (Table 8). The time intervals between strokes and the duration of rolls were measured manually on spectrograms and signal envelopes. This author rejected the time between rolls as a viable acoustic feature because of excessive variability. His findings were that *D. major* had the shortest DR, and *D. leucotos* and *D. martius* the longest. The time structure (evolution of the time between strokes) followed either a linear law or a decreasing exponential law and was a critical species trait. Some woodpeckers accelerated (e.g. *D. major*); others decelerated (e.g. *D. minor*). There was a significant difference in DR duration between *D. major* males and females, and between *D. major* male neighbours. Stark et al. (1998) performed another statistical analysis on drumming parameters for 11 woodpecker species occurring in California (3347 DRs). DR duration, number of strokes per DR, average time interval between strokes and cadence (strokes per second) were considered. Cadence was found to be the best indicator for species differentiation. 78% of all samples were correctly reclassified using their

Table 1
Woodpecker species and drumming database composition.

Index	Species	Drumming	Advertising call	Original files	Drumming rolls
1	*Dendrocopos leucotos*	Yes	No	43	248
2	*Dendrocopos major*	Yes	No	115	818
3	*Dryocopus martius*	Yes	Yes	27	84
4	*Dendrocopos medius*	Rare	Yes	3	8
5	*Dendrocopos minor*	Yes	Yes	67	832
6	*Dendrocopos syriacus*	Yes	No	3	8
7	*Jynx torquilla*	No	Yes	0	0
8	*Picus canus*	Yes	Yes	29	104
9	*Picus sharpei*	Rare	Yes	0	0
10	*Picoides tridactylus*	Yes	No	68	547
11	*Picus viridis*	Rare	Yes	6	16
			Total	361	2665

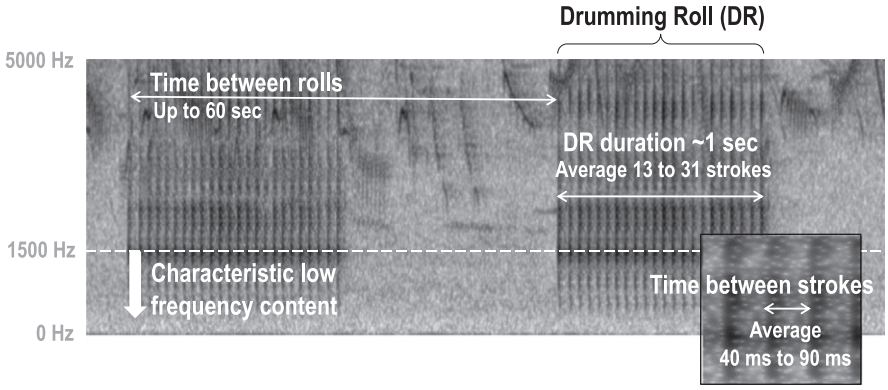

Fig. 1. Spectrogram with drumming rolls (XC173903; see footnote 1).

parameter set; 91% when considering only sympatric species (i.e. with overlapping geographical ranges). The authors argued that only species susceptible to live in the same areas needed to differentiate their acoustic signals. A few drumming parameters could never be enough to fully differentiate the 239 woodpecker species that inhabit the planet (as counted by Gorman, 2014), unless the breeding range is factored in. Our approach, considering only the European species, is consistent with this observation.

2.2. Drumming database

For the present study, 267 original files were retrieved from Xeno-Canto (XC)[i] and 94 from Tierstimmen (TS)[ii] for nine species (see Table 1). *J. torquilla* does not drum and there was no data for *P. sharpei*. The files represent 2 h and 20 min of recordings (4 GB) with a minimum of one DR per file. The dataset is skewed towards the most common species (*D. major*, *D. minor*) and species without an advertising call (*D. leucotos*, *P. tridactylus*). *P. viridis* and *D. medius* rarely drum and thus are poorly represented. Finally, *D. syriacus* is present in Eastern Europe where fewer recordings are available. Both archives use the mp3 format, with a sampling frequency of either 44.1 kHz or 48 kHz. The files can last up to several minutes. Some were low-pass filtered or edited by the recordists. We resampled all files at 11,025 Hz. For all Fourier transforms, we used frames of 23.2 ms with 50% overlap.

3. Methods

The full process flow, from the extraction of drumming rolls from audio records to the identification of woodpecker species, is sketched in Fig. 2. 324 DRs were manually extracted from the TS data, and 2341 from XC. The DRs were saved in individual files including 150 ms leading and trailing silences. The different birds issuing the DRs were identified in recordings where more than one bird was drumming. This was used to calculate the time interval between rolls; see Section 3.1.1. Subsequent sections address the parameterization of individual DRs (Sections 3.1.2 and 3.1.3) and their classification (Section 3.2).

3.1. Acoustic features for drumming

Zabka (1980) and Stark et al. (1998) both determined that the temporal parameters of drumming were the diagnostic ones. In both studies the time between strokes was the primary variable for the separation of species. This means that features based on the STFT have little chance of successfully clustering woodpecker drums. If the STFT is averaged over time as in Stowell and Plumbley (2014), then the rhythm is lost. The same goes for the MFCC. Spectrogram images as in Potamitis (2014) and Lasseck (2015) preserve the time dimension, but the time step is imposed by the STFT processing parameters, namely the frame duration and overlap. Numbers in Zabka (1980) indicate that a precision to the millisecond is required (Table 8). Alas, for frames computed every 23.2 ms with a 50% overlap, the precision in time is limited to 11.6 ms. Adapting these numbers is detrimental to the frequency resolution and to the computational effort. Again, spectrogram images do not render temporal structures and repetition rates in bird songs well (Lasseck, 2015). Our targets are the temporal parameters indicated in Fig. 1 and they require a different approach.

This being said, the spectral information is not entirely without merit. Woodpeckers choose the tree on which they drum, potentially because they like its sound. This preferred sound might be a species trait. Stark et al. (1998) supported this hypothesis. Some birds might also have the strength or skill to excite higher trunk harmonics. The STFT or MFCC would reflect all this information. Therefore, we included spectral parameters in the feature set. The final eight features are documented in Table 2, including mathematical formulations for the temporal features. All features were normalized prior to classification.

3.1.1. Time interval between rolls

The time between the start of two successive drumming rolls (Fig. 1) was estimated from the DR file start times. For isolated DRs (only one DR in the original audio record), it could not be computed. We identified series of DRs separated by at most 60 s. Hypothetically, when the interval between DRs exceeds 60 s, the bird is resting or doing something else. The time interval between rolls for one given DR was calculated by averaging the time intervals to the previous and to the following DR in the series. This is feature f_8 in Table 2.

Multiple woodpeckers drumming in succession were a complication for the analysis (e.g. XC169038 contains DRs from six confirmed individuals, XC89410 from four). DRs in a sequence had to be assigned to distinct individuals. We grouped DRs associated to one bird and one tree using k-means unsupervised clustering. Fortunately, drumming on different trees produces different nuances of sound, which is directly reflected in the spectral content. We thus used the stroke spectrum as basis for our feature vector. This is appropriate because all DRs from one bird hitting on the same tree spot have a stable spectral content; the sorting algorithm does not need to cope with shifts in frequency. The stroke frames were identified as those whose spectral energy was greater than the median frame energy for the DR. The stroke spectrum is then the average of the spectra of all these frames. Eventually, the feature vector we used was the derivative with respect to frequency of the normalized stroke spectrum. This proved more resistant to variations in signal amplitude. The feature matrix was then complemented with a training set of 76 manually pre-selected vectors. These were triplets of DRs issued by the same bird hitting the same tree. Within the triplets the DRs were thus similar in spectrum, and different triplets had distinct

[i] The Xeno-Canto Foundation, http://www.xeno-canto.org/. Audio file XCn is accessible at http://www.xeno-canto.org/n.
[ii] Museum für Naturkunde Berlin, http://www.animalsoundarchive.org/.

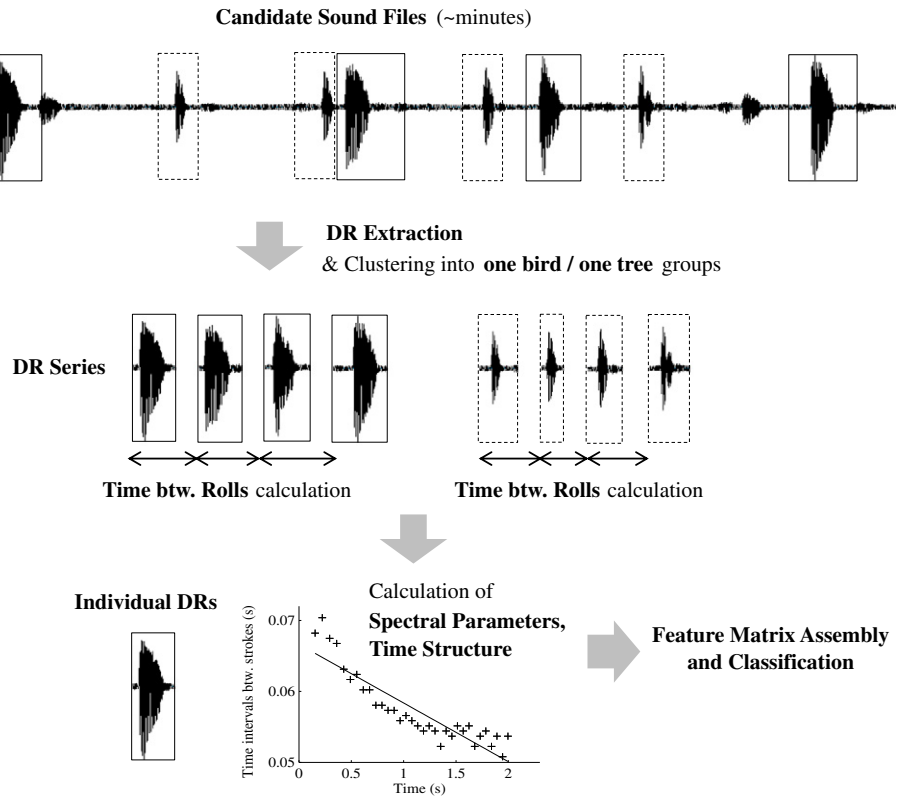

Fig. 2. Drumming species identification process flow (example: XC89410).

timbres. We selected three triplets per species if available (9 files). This added training set was meant to help k-means gauge the level of discrepancy that makes two spectra unalike. For the initial cluster centres, we picked a set of vectors which were different from each other, yet similar to other vectors, i.e. not isolated cases. As the number of bird/tree combinations (clusters) was a priori unknown, a loop was implemented where k-means was run for an increasing number of clusters. As soon as the cluster centres became similar (the dot product of the corresponding vectors yielded an angle which was less than 10°), the loop was stopped. We enforced that the final number of clusters had to remain below 6 to keep the worst mathematical artefacts at bay.

3.1.2. Refined frequency range and spectrum centroid

The default bandpass filter in our analyses was 300 Hz–1500 Hz. With a low bound of 300 Hz, common background noise is circumvented. With a high bound of 1500 Hz, the characteristic low frequency content of drumming is encompassed (see Fig. 1) and yet interferences from other birds are minimized. If drumming sounds must travel long distances in the high absorption environment of a forest, then the lower frequency content is the essential part of the signal. In some cases however, the maximum spectral peak lied outside the 300 Hz–1500 Hz range (see Fig. 5b) and the upper boundary had to be extended. Variations in the spectra of the frames containing the strokes were used to that end. Drumming produces a stable stroke spectrum, whereas the passerine songs which populate the higher frequencies are inherently frivolous. The distinction can be made visible using the standard deviation of the stroke spectrum (dash-dot line in the top right plot in Fig. 3). Above a threshold in the standard deviation, the signal corresponds either to passerines (5000 Hz peak on Fig. 3) or to the upper tail of the strokes that can be discarded for our analysis (2200 Hz peak).

Once a frequency range was adopted, the mean stroke spectrum was used to determine the spectrum centroid and peak frequency (solid curve and markers in the top right plot in Fig. 3; features f_6 and f_7 in Table 2). The Fig. 3 plot is representative of the typical drumming spectrum. There is one large peak and sometimes a small harmonic. This

Table 2
Acoustic features for drumming, expressed for n strokes with (t_i,y_i) coordinates in the envelope curve.

Numbering	Features	Mathematical formulations	
f_1	First interval	$(t_{i+1}-t_i)=f_2\cdot(i-1)+f_1$	for $i=1\cdots n-1$
f_2	Delta interval		
f_3	DR duration	$f_3=t_n-t_1+\varepsilon$	ε takes into account the width of the first and last peaks;
f_4	Number of strokes	$f_4=n$	
f_5	Amplitude slope	$\frac{y_i}{y_{max}}=f_5\cdot t_i+\eta$	for $i=5\cdots n$ η not used;
f_6	Spectrum centroid		
f_7	Spectrum peak		
f_8	Time between rolls	for the pth DR in a series of P DRs:	
		$f_8=(t_1)_2-(t_1)_1$	for $p=1$
		$f_8=\frac{1}{2}((t_1)_{p+1}-(t_1)_{p-1})$	for $p=2\cdots(P-1)$
		$f_8=(t_1)_P-(t_1)_{P-1}$	for $p=P$

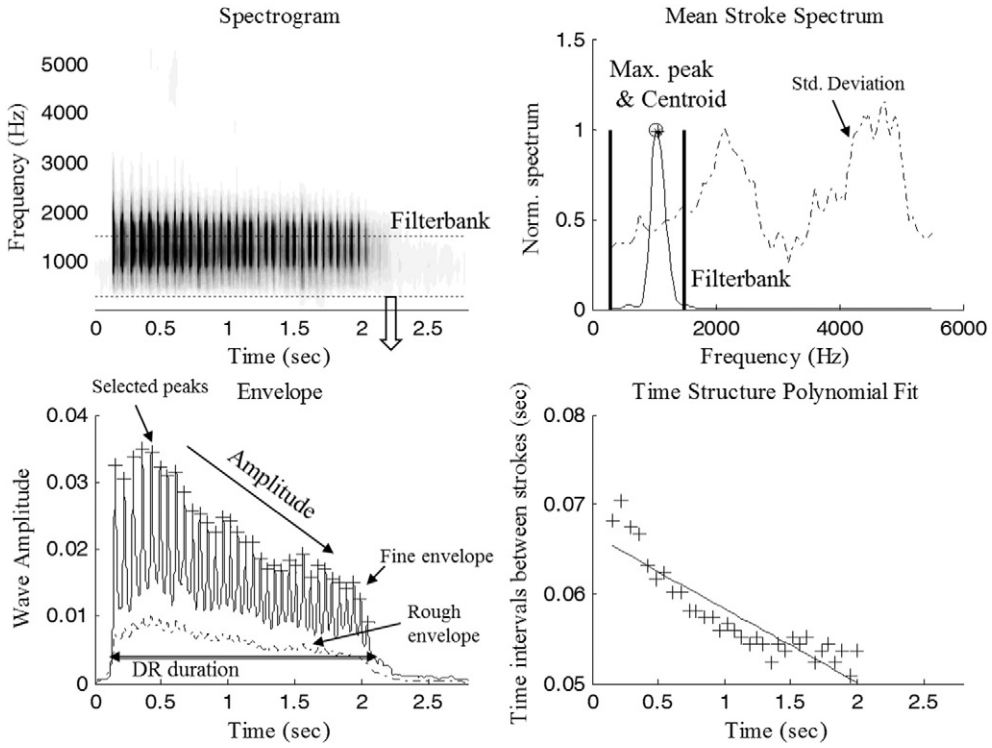

Fig. 3. Acoustic parameters calculation from spectrum and envelope (XC83280).

explains why our spectral description is so bare. The peak frequency captures the large peak. Then, in the presence of harmonics, the centroid shifts to higher frequencies. Note however that the argument for tracking harmonics remains weak because the forest absorbs higher frequencies as they propagate.

3.1.3. Time structure and DR duration

The determination of the time structure began with picking the time positions of the strokes in the DR. We used the peaks of the waveform envelope. Envelopes are generated by low-pass filtering the absolute value of the signal (Lartillot and Toiviainen, 2007). Their time resolution is inherited from the original signal. In our case, with a sampling frequency of 11,025 Hz, the time step was 0.09 ms. We calculated two envelopes. The first one had a high time constant for the low-pass filter and captured only rough energy bursts in the signal. It was used to select a time interval where the curve was above a threshold and then deduce the DR duration (feature f_3, Table 2). The second envelope had a short time constant and retained more details from the original waveform. The individual strokes were visible and well detached. This envelope was used for peak picking and the determination of the time structure (Fig. 3, lower left image).

Peaks located within the DR duration interval were selected. The fringes were checked for additional peaks that approximately matched the time structure. The number of strokes in the DR was set as the number of selected peaks (feature f_4, Table 2).

The time structure is a first-degree polynomial fit through the $(t,\delta t)$ data where t is the time position of a stroke and δt the time interval between this stroke and the next one (Fig. 3, lower right image). The final polynomial was modified to match the form in Zabka (1980), i.e. time interval versus stroke number. The two polynomial coefficients were saved as acoustic parameters for the classification. They are the initial time interval (feature f_1, Table 2) and the delta interval (feature f_2, Table 2). The delta interval is the slope of the polynomial or the difference in the duration of two successive intervals. In mechanical terms, it relates to acceleration, while the initial time interval is the inverse of speed. Zabka used both first-degree polynomials and decreasing exponentials to describe time structures. We did not find in our data the necessity for the exponential form. His exponential functions were converted back to polynomials to allow the comparison with our results (Table 8).

A second polynomial fit was produced on the amplitude of the peaks (normalized to the maximum value). This was to capture the fact that some birds drum with an increasing intensity, others with a stable or decreasing intensity. Only the slope coefficient was saved (feature f_5, Table 2).

Fig. 3 summarizes the calculation of features f_1 through f_7 for a typical DR extracted from XC83280. The top left image is the spectrogram of the segment; the boundaries of the band-pass filter are showed with dashed lines. The bottom left image contains the envelopes and the peak selection. The DR duration and amplitude trend are annotated. The bottom right image shows the time intervals involved in the determination of the time structure polynomial.

3.2. Visualisation and classification

We used two techniques: k-Nearest Neighbour (k-NN) for the classification, and t-Distributed Stochastic Neighbour Embedding (t-SNE) for a preliminary visualisation of the database and a health check on

Table 3
Training set composition.

	Total files	Training set composition	
		(in number of files)	(in %)
D. leucotos	248	83	33.5
D. major	818	113	13.8
D. martius	84	43	51.2
D. medius	8	5	62.5
D. minor	832	113	13.6
D. syriacus	8	5	62.5
P. canus	104	50	48.1
P. tridactylus	547	112	20.5
P. viridis	16	10	62.5

Table 4
Mean parameter values.

	Number of files	First interval (ms)	Delta interval (ms)	DR duration (s)	Number of strokes	Amplitude slope	Spectrum centroid (Hz)	Spectrum peak (Hz)	Time between rolls (s)
D. leucotos	248	71.3	−1.6	1.60	30.1	−0.45	1206	1041	18.6
D. major	818	60.1	−2.4	0.65	13.0	−0.68	1481	1323	11.7
D. martius	84	68.7	−0.7	1.86	31.0	−0.22	801	683	13.2
D. medius	8	42.5	−0.6	1.11	25.5	0.18	1387	624	4.3
D. minor	832	53.6	0.1	1.16	21.1	0.07	1708	1581	5.7
D. syriacus	8	58.6	−1.6	0.97	19.4	−0.58	1486	1330	20.1
P. canus	104	48.3	0.1	1.57	31.5	−0.03	1116	965	19.1
P. tridactylus	547	77.6	−0.5	1.36	18.9	0.07	1231	1041	16.2
P. viridis	16	50.6	0.3	1.27	23.0	−0.02	1150	918	11.5

our features. The t-SNE unsupervised scheme (van der Maaten and Hinton, 2008) produces 2D coordinates based on feature vectors of any dimension. Provided the acoustic features have any merit, the different classes form separate clusters in the resulting scatterplot. This is similar to Linear Discriminant Analysis (LDA). However, the dimensionality reduction in t-SNE is a non-linear operation and the final 2D coordinates have no physical interpretation. This is because t-SNE does not represent distances between feature vectors. It replaces them with conditional probabilities between pairs of points both in the high and in the low dimensional space. The probability that point A_j is similar to point A_i derives from a distribution centred on A_i. The probability is high for points close to A_i and low at greater distances. The distribution is Gaussian in the high-dimensional space and Student-t in the low-dimensional space. The Student-t distribution has heavier tails, i.e. there is more margin to position points away from the distribution centre. Thus, a modest distance in the high dimension produces a greater distance in the low dimension. This avoids crowding certain regions of the map. The variance of the distributions is chosen so as to keep 5–50 points in the bulk of the distribution close to A_i. This number of points is called the perplexity and is left to the preference of the user. Finally, t-SNE optimizes the coordinates in the low-dimensional space so as to match the pairwise probabilities of both spaces. The t-SNE visualisation typically offers more separation of the classes than LDA because there is more freedom to spread out the points. Another advantage is that two dimensions are sufficient by design. t-SNE has enjoyed a widespread success in the machine learning community (Dupont et al., 2013) due

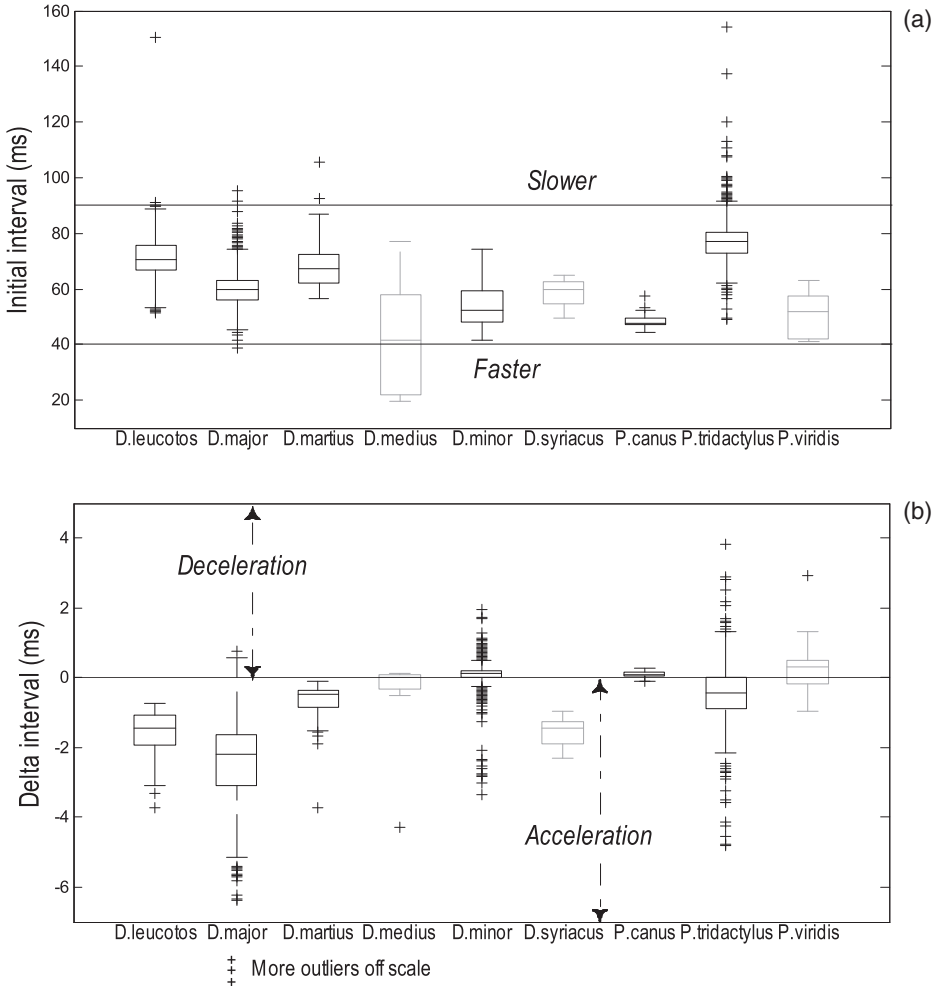

Fig. 4. Parameter distributions per species for (a) initial interval (b) delta interval; in grey, classes with less than 20 samples.

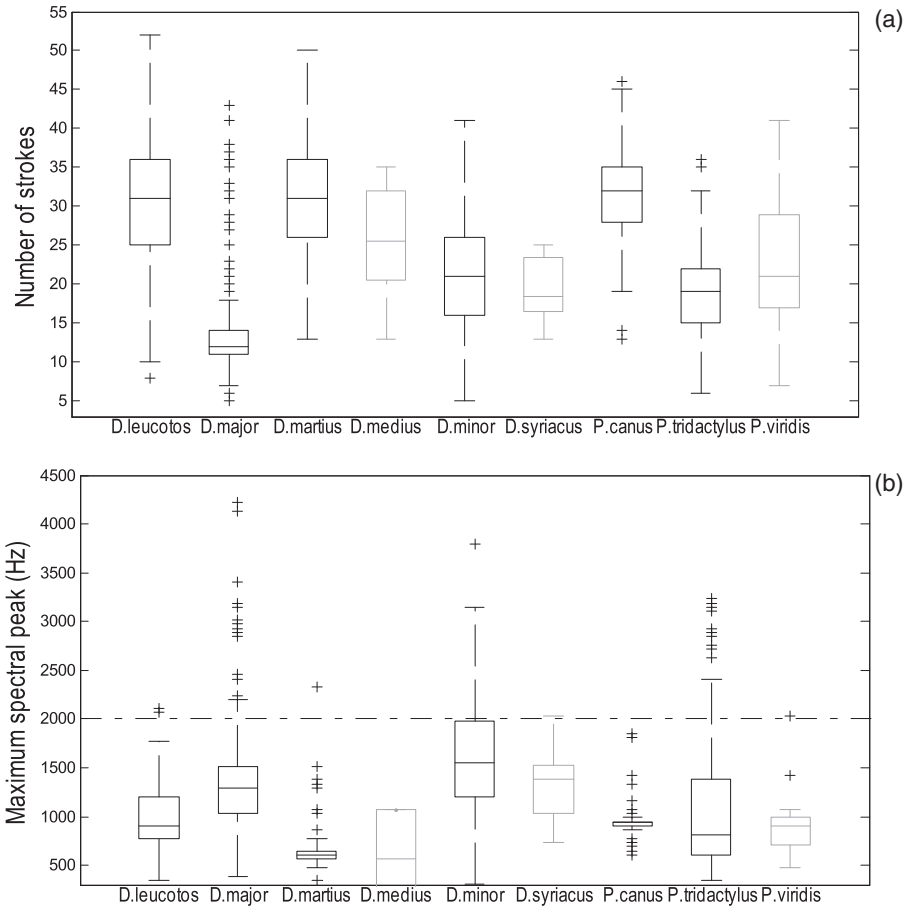

Fig. 5. Parameter distributions per species for (a) number of strokes (b) maximum spectral peak; in grey, classes with less than 20 samples.

to its unique ability to translate the qualities of large and high-dimensional databases into simple maps. In our case, producing a t-SNE map provides qualitative information about the adequacy of our acoustic features and about the classes that are insufficiently characterized. Author van der Maaten readily supplies a Matlab implementation of t-SNE[iii] where a Principal Component Analysis (PCA) is performed on the features before the actual dimensionality reduction.

For the classification itself, we chose k-NN with $k = 5$ (Mitchell, 1997). Test samples were thus matched to their five nearest neighbours in the training base. The class that received the majority of the votes was assigned. In this way, the test samples were not matched to anomalies in the training base. Values of k greater than 5 were found to diminish the accuracy; they require larger training sets. More trial runs showed that a training set comprising 20% of the total database (i.e. 534 files out of 2665) was adequate. The accuracy of the classification plateaued for greater percentages. We also observed that correct identifications increased with the class size. We thus adjusted the composition of the training set (Table 3) to have the small classes (*D. medius*, *D. syriacus* and *P. viridis*) contribute approximately 60% of their sample count and the larger classes a reduced percentage. In absolute terms 60% only amounts to 5–10 files for the small classes. This remains insufficient. On the other hand, we were able to set aside close to 50 training samples for the intermediate classes *D. martius* and *P. canus*.

The k-NN classification was run two hundred times, each time with a different randomly selected training set. Fewer runs would have worked as well, as eventually the standard deviation in the accuracy results was only 0.8%. The sample selection for the training set was restricted to samples having a value defined for the time interval between rolls (feature f_8). The test samples with and without f_8 values were classified separately, using the full training set each time.

4. Results

4.1. Individual birds

The detection of individual bird/tree combinations was tested on 16 *D. major* files for which the XC or TS annotations report several birds.[iv] These files typically contain 20 to 40 DRs and two to four birds. The solutions were perfect for nine files; three more had two wrongly classified DRs; only one had poor results (5 misclassified DRs out of 15). Further checks on other species confirmed that the bird/tree assignments are realistic. For example, for *D. minor* XC171084 where 109 DRs were identified in a 35-minute file, long series of identical DRs (same bird/tree) were detected, followed by transitional phases. This complies with the ornithologist's notes, which indicate that the bird kept going back and forth between two trees. The bird/tree assignments are also in agreement with the audio. In the full drumming database, 42% of the recordings (153 out of 361) were found to have more than one bird/tree combination. Eventually, it was possible to assign a value of time interval between rolls to 92% of the 2665 DRs. Amongst the remaining 8%, 2% correspond to files where only one DR was recorded and 6% (171 DRs) are DRs that were isolated by the clustering (i.e., associated with no other DR).

[iii] http://lvdmaaten.github.io/tsne/.

[iv] TS14, TS29, XC45165, XC71449, XC84275, XC89410, XC92455, XC94435, XC94436, XC125675, XC128132, XC130203, XC138947, XC169038, XC169486, XC 170195.

Table 5
Parameter correlations; values greater than 0.8 in bold.

	Delta int.	First int.	DR dur.	N. strokes	Sp. cent.	Sp. peak	Amplitude	Time btw. rolls
Delta interval	**1,00**	−0,26	0,46	0,39	0,11	0,11	0,63	−0,10
First interval	−0,26	**1,00**	0,25	−0,05	−0,26	−0,24	0,02	0,24
DR duration	0,46	0,25	**1,00**	**0,91**	−0,15	−0,15	0,35	0,16
Number of strokes	0,39	−0,05	**0,91**	**1,00**	−0,12	−0,11	0,23	0,14
Spectrum centroid	0,11	−0,26	−0,15	−0,12	**1,00**	**0,83**	0,07	−0,22
Spectrum peak	0,11	−0,24	−0,15	−0,11	**0,83**	**1,00**	0,06	−0,19
Amplitude slope	0,63	0,02	0,35	0,23	0,07	0,06	**1,00**	−0,08
Time btw. rolls	−0,10	0,24	0,16	0,14	−0,22	−0,19	−0,08	**1,00**

4.2. Acoustic parameter values

Parameter mean values are documented in Table 4. Figs. 4 and 5 show box plots[v] for the time structure, the number of strokes and the maximum spectral peak. The light grey boxes are used to differentiate the small classes (*D. medius*, *D. syriacus* and *P. viridis*), for which the results should be treated with caution. As expected, there is a strong correlation between the number of peaks and the DR duration (0.91), and between the spectrum centroid and the maximum spectral peak (0.83); see Table 5. All outliers were checked manually for the time structure and the number of strokes. Because of the short DR duration for *D. major*, there were fewer peaks available for a robust polynomial fit and ultimately more algorithm failures for this taxon. Table 4 and Fig. 5(a) show that *D. major* consistently has the lowest number of strokes per DR, while *D. leucotos*, *D. martius* and *P. canus* have the greatest numbers of strokes.

Most species accelerate while drumming (negative polynomial slope in Table 4 and Fig. 4(b); the time interval between strokes diminishes), with *D. major* exhibiting the largest gradients. *D. minor* and *P. canus* are the two significant exceptions amongst the species with large datasets; here the values indicate slight deceleration or constant speed drumming. This is coherent with the fact that *D. minor* and *P. canus* are the fastest drummers, as reflected by their short initial intervals (Table 4 and Fig. 4(a)). Starting at a high speed, they tend to decelerate. *P. tridactylus* is the overall slowest bird, with multiple initial interval values far above the likely range of 40 ms–90 ms inferred from Zabka (1980) and marked on Fig. 4(a). For example, intervals of 120 ms are measured in XC153234.

Three species produce frequencies above 2000 Hz (*D. major*, *D. minor*, *P. tridactylus*) as seen in the extent of the whiskers on Fig. 5(b). Overall, the data justify our assumption that the low frequency content below 1500 Hz is characteristic for drumming. *D. minor* uses the highest pitch (and is the smallest bird) and *D. martius* the lowest (and is the largest bird). *D. major*, *D. syriacus*, *D. leucotos* and to a lesser extent *D. martius* display trends of decreasing peak amplitudes, while other species appear to drum with a constant intensity (Table 4). We note that pairs of twin species (*D. major* and *D. syriacus*; *P. canus* and *P. viridis*) have similar time structures but differ in the duration of the DR or in the number of strokes.

Table 6 shows Fisher's discriminating power for all parameters and the variation for the ones with positive values. Fisher's discriminating power D_k (Eq. (1)) is calculated to calibrate expectations regarding the acoustic indicator k. In Eq. (1), C is the number of classes, n_i the number of samples of class i, μ_{ik} the mean of the kth indicator over class i samples, μ_k the mean of the kth indicator over all samples and σ_{ik}^2 the variance of the kth indicator over class i samples. Following Zabka (1980), the variation V_{ik} for the kth indicator over class i samples is computed as the standard variation over the mean (Eq. (2)). Table 6 documents its average over all classes. We obtain 16% for the first interval and around 30% for DR duration, number of strokes and spectrum centroid.

[v] The boxes in Fig. 4 are bounded by the first and third quartiles of the data. The line within the box is the median. The whiskers extend to the values whose distance to the box is at most 1.5 times the height of the box (Tukey boxplot). Crosses are outliers.

With a variation of 66%, we confirm that the time interval between rolls is a volatile parameter. The discriminating powers are in agreement with these numbers: the time structure parameters are the critical ones, with DR duration in second line and other parameters (spectral parameters, amplitude, time between rolls) being less promising.

$$D_k = \frac{\sum_{i=1}^{C} n_i.(\mu_{ik}-\mu_k)^2}{\sum_{i=1}^{C} n_i.{\sigma_{ik}}^2} \quad (1)$$

$$V_{ik} = \frac{\sigma_{ik}}{\mu_{ik}} \quad (2)$$

4.3. Visualisation and classification

The t-SNE visualisation is in Fig. 6. The data groups are clustered together in separate regions of the map, which indicates that the parameter set achieves discrimination of the classes. This is more efficient for the largest classes (*D. major*, *D. minor*, *D. leucotos* and *P. tridactylus*), while there is some degree of overlapping for the medium-size classes (*D. martius* and *P. canus*). Again, the point coordinates in t-SNE maps have no physical meaning. Yet, the optimisation seemingly exploited DR duration in the horizontal direction and drumming speed in the vertical direction. *D. major* with the shortest DRs is to the left, whereas *D. leucotos*, *D. martius* and *P. canus* with DRs longer than 1.5 s are to the right. The fast *D. minor* and *P. canus* are at the top and the slow *P. tridactylus* at the bottom. As it seems, the parameters with the most discriminating power were prioritized.

The accuracy and the confusion matrix of the k-NN classification are documented in Table 7. The accuracy is the percentage of test samples that are correctly identified per class. The scores follow the class size and agree with the levels of clustering observed on the t-SNE map. The confusion matrix also agrees with the t-SNE map. For example, some *D. martius* samples are misclassified as *D. leucotos* and *P. tridactylus*. These three species are neighbours on the t-SNE maps. The accuracy for the full test dataset (i.e. excluding training set) is 87.2%. Inevitably, the small classes are poorly classified.

Following the work in Dupont et al. (2013), we also attempted the k-NN classification on the low-dimensional space, i.e. the coordinates of the points in the t-SNE map in Fig. 6. We correctly predicted an average

Table 6
Parameter prospective quality.

	Fisher discriminating power	Variation (%)
First interval	1.56	16%
Delta interval	1.23	
DR duration	1.01	29%
Number of strokes	0.85	29%
Amplitude slope	0.76	
Spectrum centroid	0.36	30%
Spectrum peak	0.27	41%
Time between rolls	0.40	66%

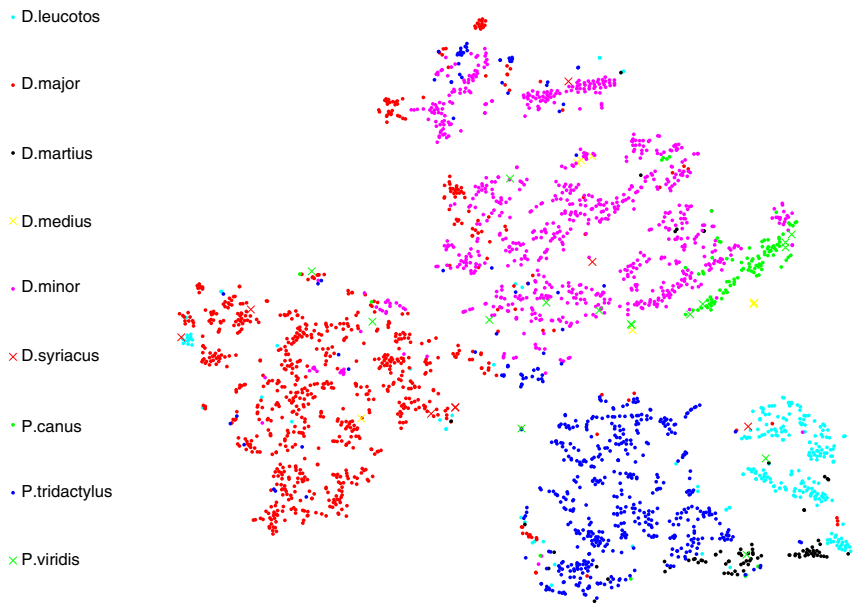

Fig. 6. t-SNE map with 2665 drumming roll samples.

of 86.1% of the test samples in 200 trials, with a standard deviation of 1.2%. Therefore in our case, t-SNE does not extract attributes from the database that were not already in evidence with our initial features. The technique provides us with visualisation but does not improve classification.

5. Discussion

The parameter values for the time structure and DR duration are remarkably in line with Zabka's numbers for the classes that are well represented in both studies (this study: Table 4; Zabka: Table 8). The associated variation coefficients also compare well. The delta interval value is particularly stable. This speaks to the independence of these parameters from the content and the size of the database. Zabka's sound files were not recovered for the present study and the ratio in database size is 11 to 1.

The time interval between rolls drove an important programming effort including the recognition of different individuals but proved inefficient in separating the species. For all well-represented taxa, the values essentially spanned the entire allowed range of 0–60 s. A few estimates were terminally unreliable as some recordists posting on XC edit silence out of their files (e.g. XC171084). No evidence was found in the data that the time interval between rolls depends on the bird drumming alone or with other birds.

The present study shows that time parameters are essential in characterizing woodpecker drumming. With this in mind, it follows that a parameterization that focuses on spectral content is inadequate to identify woodpecker species from their drumming sounds. Neither are spectrograms a good base to derive parameters because their limited time resolution (11.6 ms) blurs the differences between the taxa. The variations in drumming speed are subtler than that. For example, *D. martius* loses 0.7 ms with every stroke (Table 4). Once adequate acoustic parameters are obtained, the power of the classification algorithm is secondary. A simple scheme like k-NN is sufficient to obtain acceptable classification scores. If the acoustic features were deficient, there would be no significant recovery to be hoped from an advanced classifier such as random forest.

Finally, we join the praise for t-SNE which provided us with a remarkable visualisation of our database and its dynamics. The map in Fig. 6 is convincing evidence of the discriminating power of our feature set. We used a default perplexity of 30, which seemed to aggregate the largest classes well. Perplexity relates to the size of the typical core group of samples that can be considered as close neighbours. For the classes with 100 samples or less (*D. martius* and *P. canus*), disconnected sub-clusters suggest gaps in the dataset.

In a few instances, our program failed to calculate the drumming parameters. This may explain a few dubious placements in the t-SNE map in Fig. 6. The failures occurred when the program could not produce an envelope with clearly detached peaks. The two problematic situations are a) when the drumming roll is distant and is heard only faintly on the recording and b) when the drumming roll is superposed with another vocalisation. Option a) is hard to recover from. It means that recording stations can only process drumming within a certain distance. Option b) is not that common because the frequency range of

Table 7
k-NN confusion matrix and accuracy.

	Predicted classes: mean number of samples over 200 trials									
Actual classes ↓	*D. leuc.*	*D. maj.*	*D. mart.*	*D. med.*	*D. min.*	*D. syr.*	*P. can.*	*P. tri.*	*P. vir.*	Accuracy
D. leucotos	138.9	12.6	3.3	0.0	0.6	0.4	0.3	8.9	0.0	84.2%
D. major	15.3	625.6	0.9	0.0	38.7	0.9	1.1	22.2	0.2	88.7%
D. martius	5.7	0.8	26.4	0.0	2.2	0.2	1.5	4.3	0.0	64.4%
D. medius	0.0	0.9	0.0	0.6	1.1	0.0	0.4	0.0	0.0	19.8%
D. minor	0.8	27.4	1.3	1.7	647.1	0.2	24.2	15.9	0.4	90.0%
D. syriacus	0.8	1.7	0.0	0.0	0.0	0.4	0.0	0.1	0.0	12.3%
P. canus	0.2	1.2	0.1	0.0	6.3	0.1	44.7	1.1	0.4	82.8%
P. tridactylus	6.2	16.3	6.5	0.2	26.0	0.5	4.2	374.6	0.4	86.1%
P. viridis	0.1	0.3	1.0	0.0	3.0	0.0	1.2	0.2	0.1	1.2%

Table 8
Parameter mean values and variation published in Zabka (1980).

		Delta interval	First interval		DR duration		Number of strokes	
		Mean	Mean	Variation	Mean	Variation	Mean	Variation
	Number of files	(ms)	(ms)	(%)	(s)	(%)		(%)
D. leucotos	17	−1.2[a]	80.2[a]	8[b]	1.64	11.3	34.4	13.8
D. major	104	−2.0[a]	61.5[a]	8[b]	0.56	28.7	13.1	27.2
D. martius	21	−0.6[a]	72.0[a]	12[b]	1.61	31.3	29.2	27.2
D. medius	17	0.2	57.1	4[b]	1.29	10.9	23.0	26.4
D. minor	40	0.0	48.9	12[b]	1.19	24.1	24.6	22.9
D. syriacus	11	−2.0[a]	72.9[a]	8[b]	0.89	16.8	21.6	20.6
P. canus	16	0.1	52.3	13[b]	1.37	21.0	26.4	21.2
P. tridactylus	9	−0.5	76.1	3[b]	1.34	25.4	20.8	19.1
P. viridis	5	0.6	40.4	–	1.15	24.3	25.8	27.5

[a] Transformed from exponential form.
[b] Inferred from illustrations.

drumming is depleted in European forests. In any case, failure to process sporadic drumming rolls is not a road block because birds produce them in large quantities. There will be enough clean drumming rolls to make an identification.

6. Conclusion

The present study demonstrates the feasibility of identifying European woodpecker species from their drumming. In support of this, the calculation of drumming features was automated and k-NN classification was demonstrated to achieve an accuracy greater than 82.8% when sufficient training data was available. For the taxa that are less reliant on drumming, a characterization of the most remarkable vocalisations is still a necessity.

Robustness of the methodology is supported by the size of the database on which the algorithms were tested (2665 drumming rolls) and by the agreement in the parameter values, including their variation, with the previous study of Zabka (1980). The work would benefit from access to additional samples for *D. martius* and *D. syriacus*. The latter is seldom recorded. Other small classes correspond to rare drummers for which data cannot be abundant (*D. medius*, *P. viridis* and *P. sharpei*).

Last, we reiterate that our methodology is based on the assumption that the best characterizing parameters are species-specific (or genus-specific, or family-specific). This creates limitations on the number of taxa under consideration and is well-suited to mining problems. Incidentally, future work includes the development of a detection algorithm that could autonomously search long audio recordings for drumming rolls.

References

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. Peer J. 1, e103.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.H., Frommolt, K.H., 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recogn. Lett. 31, 1524–1534.

Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Particelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., McCowan, B., Ali, A.M., Kirschel, A.N.G., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. J. Appl. Ecol. 48, 758–767.

Brandes, T.S., 2008. Feature vector selection and use with hidden Markov models to identify frequency-modulated bioacoustic signals amidst noise. IEEE Trans. Audio Speech 16 (6), 1173–1180.

Catchpole, C.K., Slater, P.J.B., 2008. Bird Song: Biological Themes and Variations. Second edition. Cambridge University Press.

Dodenhoff, D.J., Stark, R.S., Johnson, E.V., 2001. Do woodpecker drums encode information for species recognition? Condor 103 (1), 143–150.

Dupont, S., Ravet, T., Picard-Limpens, C., Frisson, C., 2013. Nonlinear dimensionality reduction approaches applied to music and textural sounds. Proceedings of the 2013 IEEE International Conference on Multimedia and Expo (ICME); San Jose, USA.

Farina, A., Lattanzi, E., Malavasi, R., Pieretti, N., Piccioli, L., 2011. Avian soundscapes and cognitive landscapes: theory, application and ecological perspectives. Landsc. Ecol. 26, 1257–1267.

Florentin, J., Fauville, B., Gérard, M., Moiny, F., Rasmont, P., Kouroussis, G., Verlinden, O., 2015. Soundscape analysis and wildlife presence in the vicinity of a wind turbine. Proceedings of Euronoise 2015; Maastricht, The Netherlands.

Fox, E.J.S., Roberts, J.D., Bennamoun, M., 2008. Call-independent individual identification in birds. Bioacoustics 18, 51–67.

Gorman, G., 2014. Woodpeckers of the World: The Complete Guide. Bloomsbury Publishing.

Jahn, O., Mporas, I., Potamitis, I., Kotinas, I., Tsimpouris, C., Dimitrou, V., Kocsis, O., Riede, K., Fakotakis, N., 2013. The AmiBio project: automating the acoustic monitoring of biodiversity [poster]. 24th International Bioacoustics Congress; Pirenopolis, Brazil.

Kogan, J.A., Margoliash, D., 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. J. Acoust. Soc. Am. 103 (4), 2185–2196.

Lartillot, O., Toiviainen, P., 2007. A Matlab toolbox for musical feature extraction from audio. International Conference on Digital Audio Effects; Bordeaux, France.

Lasseck, M., 2015. Towards automatic large-scale identification of birds in audio recordings. Lect. Notes Comput. Sci 9283, 364–375.

Lee, C.H., Hsu, S.B., Shih, J.L., Chou, C.H., 2013. Continuous birdsong recognition using Gaussian mixture modeling of image shape features. IEEE Trans. Multimedia 15 (2), 454–464.

Mikusinski, G., Angelstam, P., 1998. Economic geography, forest distribution, and woodpecker diversity in Central Europe. Conserv. Biol. 12, 200–208.

Mitchell, T., 1997. Machine Learning. McGraw-Hill.

Pettersson, B., 1985. Extinction of an isolated population of the middle spotted woodpecker *Dendrocopos medius* (L.) in Sweden and its relation to general theories on extinction. Biol. Conserv. 32 (4), 335–353.

Potamitis, I., 2014. Automatic classification of a taxon-rich community recorded in the wilD. PLoS One 9 (5), e96936.

Ranjard, L., Ross, H.A., 2008. Unsupervised bird song syllable classification using evolving neural networks. J. Acoust. Soc. Am. 123 (6), 4358–4368.

Somervuo, P., Harma, A., Fagerlund, S., 2006. Parametric representations of bird sounds for automatic species recognition. IEEE Trans. Audio Speech 14 (6), 2252–2263.

Stark, R.D., Dodenhoff, D.J., Johnson, E.V., 1998. A quantitative analysis of woodpecker drumming. Condor 100 (2), 350–356.

Stowell, D., Plumbley, M.D., 2014. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. Peer J. 2, e488.

Swiston, K.A., Mennill, D.J., 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. J. Field Ornithol. 80 (1), 42–50.

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., Roe, P., 2014. Visualization of long-duration acoustic recordings of the environment. Procedia Comput. Sci. 29, 703–712.

Tremain, S.B., Swiston, K.A., Mennill, D.J., 2008. Seasonal variation in acoustic signals of pileated woodpeckers. Wilson J. Ornithol. 120 (3), 499–504.

Ulloa, J.S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J., 2016. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. Ecol. Inform. 31, 91–99.

van der Maaten, L.J.P., Hinton, G.E., 2008. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Zabka, H., 1980. Zur funktionellen Bedeutung der Instrumentallaute europäischer Spechte unter besonderer Berücksichtigung von *Dendrocopos major* (L.) und *D. minor* (L.) (Functional significance of European woodpecker instrument sounds with special consideration of *Dendrocopos major* and *D. minor*). Mitt Zool Mus Berlin 56 (Suppl.: Ann Ornithol. 4), 51–76 German.