Furnariidae species classification using extreme learning machines and spectral information

E. M. Albornoz[1], L. D. Vignolo[1], J. A. Sarquis[2], and C. E. Martínez[1]

[1] Instituto sinc(*i*)
Universidad Nacional del Litoral - CONICET, Argentina
{emalbornoz,ldvignolo,cmartinez}@sinc.unl.edu.ar
[2] Instituto Nacional de Limnología
Universidad Nacional del Litoral - CONICET, Argentina
juandres.sarquis@gmail.com

Abstract. Automatic bird species classification and identification are issues that have aroused interest in recent years. The main goals involve more exhaustive environmental monitoring and natural resources managing. One of the more relevant characteristics of calling birds is the vocalisation because this allows to recognise species or identify new ones, to know its natural history and macro-systematic relations, among others. In this work, some spectral-based features and extreme learning machines (ELM) are used to perform bird species classification. The experiments were carried on using 25 species of the family Furnariidae that inhabit the Paranaense Littoral region of Argentina (South America) and were validated in a cross-validation scheme. The results show that ELM classifier obtains high classification rates, more than 90% in accuracy, and the proposed features overperform the baseline features.

Keywords: Birds classification · spectral information · auditory representation · extreme learning machines.

1 Introduction

The presence of avian species is usually perceived through vocalisations, one the most noticeable characteristics of calling birds [34]. The census of bird species allows to estimate the biodiversity in a habitat due to they respond quickly to changes, are relatively easy to detect and may reflect changes at lower trophic levels (e.g. insects, plants) [9,28]. With the improvement of technological devices, more and more birds data can be collected in almost any habitat. Nevertheless, some problems arise as poor sample representation in remote regions, observer bias [26], defective monitoring [7], and high costs of sampling on large spatial and temporal scales, among others.

Bird vocalisations field has influenced the ethology [20], taxonomy [37] and evolutionary biology [31]. In addition, ecosystems monitoring is benefits from vocalisation identification because it allows registering and processing the recordings, and improving the data collection in the field [41]. Gather data in disjoint or large areas is essential for conducting reliable studies.

Fig. 1. Paranaense Littoral region (Argentina). Taken from [4].

Passeriformes produce complex songs and can adapt their content over time: depending on the audience [10] or to match it with that of their neighbours [32]. Even, they can take possession of new songs or syllables during their lifetime [29]. Specifically, the Furnariidae family has several songs and some species show similar structures in their songs, manifested in introductory syllables or in the trill format. More complexity is added because the environmental conditions (humidity, wind, temperature, etc.) may alter the recording process, modifying the features that are present in the structure of songs and in the calls (e.g. frequency, duration, amplitude, etc.) [19,47]. Consequently, researchers use recordings from known databases, in order to avoid errors and distortions in analyses and results. As the scientific community validates these registrations (attributes, labels, etc.), they are more credible than "home-made" records despite these can be also affected by environmental conditions. Some works describe vocalisation changes in certain Furnariidae species [5, 35, 46], however, it is novel to evaluates several vocalisations of Furnariidae species from South America simultaneously [4]. In this work, the analysed Furnariidae species inhabit the Paranaense Littoral region (see Figure 1). Many recent studies on bird vocalisations report that this region has become in an interesting place for this task [5].

The classification scheme can be defined as a pipeline of three steps: preprocessing, feature extraction and classification. The first one depends strongly on the recording process and involves filtering, segmentation and enhancement of audio signals. Regarding feature extraction, time- and frequency-based information was employed [25, 34]. In addition, characteristics originally developed for speech analysis were used for bird call recognition: mel frequency cepstral coefficients (MFCCs) [16] and standard functionals (mean, standard deviation, kurtosis, etc.) computed over these [13, 36]. Various techniques have been applied to bird call classification: Gaussian mixture model (GMM) [38], support vector machines (SVM) [4], random forest (RF) [8], among others. An interesting strategy based on the pairwise similarity measurements, computed on bird-call spectrograms, was evaluated in [25], where the authors used different classifiers to recognise four species. In [13], thirty-five species were classified using a SVM classifier and six functionals were obtained from each MFCC. A different approach was proposed in [42], where a classifier based on hidden Markov models (HMMs) was used to recognise bird calls through their temporal dynamics. Previous works developing full-automatic methods for vocalisation recognition can be examined in [17, 23, 40], and the current relevance of this topic is shown in some recent works [16, 36]. However, to address the vocalisation recognition of species belonging to the Furnariidae family is novel.

This study proposes to use Extreme Learning Machines classifier with spectral-based parameterisations for Furnariidae species classification. The model needs to be able to perform properly using data from three different databases. The main contributions of this work are the compilation of an interesting set of songs for 25 species of the Furnariidae family, to address the complex classification of species taken from the same family, a novel use of spectral-based and auditory inspired features for this task, and a novel approach using the ELM network.

The following section introduces the data, the features extraction process and the classifier. Section 3 deals with the experimental setup, presents the results an its discussions. Finally, conclusions are summarised and future work is proposed in the last section.

2 Materials and methods

This section resumes the speech database, the baseline systems on the task and our approach to feature extraction.

2.1 Bird call corpus and baseline system

To obtain a suitable number of vocalisations for training the classifiers and evaluating the performance, records from three well-known databases were selected and processed to obtain 751 recordings of Furnariidae species. Some of these were selected from the *Xeno-canto*[3] database [24, 33, 34], others were taken from the *Birds of Argentina & Uruguay: A Field Guide Total Edition* corpus [12, 27, 30], and finally, several recordings were taken from *The Internet Bird Collection*[4] [1].

This set of audio signals, obtained from different data sources, involves an additional complexity that the model should be able to handle.

[3] `http://www.xeno-canto.org/`
[4] `http://ibc.lynxeds.com/`

Similar to [4], the state-of-art (for speech signals) feature sets are obtained from the recordings using the *openSMILE* toolkit [14]. It calculates 6373 acoustic features using diverse functionals over low-level descriptor (LLD) contours, and with these we computed the three feature set used as baseline sets. A full description can be found in [43].

2.2 Mean of log-spectrum

The Mean of Log-Spectrum (MLS) coefficients is a set of features calculated from spectral data for different frequency bands. They were defined to extract relevant information from speech signals and were firstly used in the analysis and classification of spoken emotions (in clean and noisy conditions [2, 3]). The MLS coefficients are defined as the average of the signal spectrogram

$$S(k) = \frac{1}{N} \sum_{n=1}^{N} \log |v(n,k)|, \tag{1}$$

where k is a frequency band, N is the number of frames in the signal and $v(n,k)$ is the discrete Fourier transform of the signal in the frame n. For the computation, the spectrograms were obtained with Hamming windows of 25 ms.

2.3 Mean of the log-auditory spectrum

In the same way as previously, we propose to analyze the recordings by means of a related set of features based on the auditory spectrogram. The representation of the sound signal at the cochlear level and auditory cortical areas has been studied as an alternative to classical analysis methods, given its intrinsic selective tuning to relevant natural sound [44]. In [45], a model based on neurophysiological investigations at various stages of the auditory system was proposed. This model has two consecutive stages: an early auditory spectrogram with the activity of auditory nerve fibres (Figure 2.3), and a model of the primary auditory cortex used to process the spectrogram and find the spectro-temporal receptive fields. The first stage uses a bank of 128 cochlear (bandpass) filters in the range $[0 - 4000]$ Hz, with the central frequency of the filter at location x on the logarithmic frequency axis (in octaves) is defined as $f_x = f_0 2^x$(Hz), where f_0 is a reference frequency of 1 kHz. This frequency distribution proved to be satisfactory for the discrimination of acoustic clues in speech and further reconstruction of the signals [11]. Using the first stage output, a set of features is built using the mean of the log auditory spectrogram (MLSa) [3], as

$$S_a(k) = \frac{1}{N} \sum_{n=1}^{N} \log |a(n,k)|, \tag{2}$$

where k is a frequency band, N is the number of frames in the utterance and $a(n,k)$ is the k-th coefficient obtained by applying the auditory filter bank to the signal in the frame n.

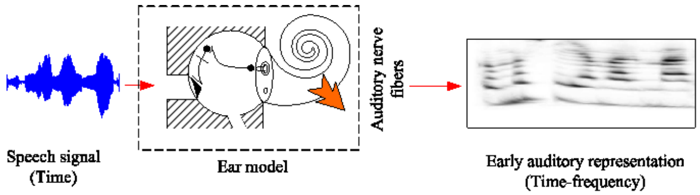

Fig. 2. Scheme of the used auditory model.

2.4 Extreme learning machines

The ELM is a kind of artificial neural network with one hidden layer [22] and its main peculiarity respecting to classical models is the training algorithm. It does not need parameter tuning and the hidden neurons are randomly initialised. Consequently, the training time is significantly reduced with respect to other training methods that use complex optimisation techniques.

Formally, let be J hidden units with F inputs and P output units. The hidden layer output is given by

$$h_j = \Phi(\mathbf{v}_j^T \mathbf{x} + b_j), \tag{3}$$

with Φ as a non-linear activation function, $\mathbf{v}_j$ the input weights and b_j the bias for the j-th hidden unit. The hidden-layer output, also known as projected features, can be expressed as $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]^T$. Rewriting the equation in a matrix form, with $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_P]$, $\mathbf{w}_p \in \mathbb{R}^J$ and $p = 1, \ldots, P$ as the output layer weights, the ELM output is

$$\tilde{\mathbf{Y}} = \mathbf{HW}. \tag{4}$$

If the function Φ satisfy certain properties (infinitely differentiability and random hidden weights) it can be shown that for any pair of inputs $(\mathbf{X}, \mathbf{Y})$ exists a number $J < N$ such $||\tilde{\mathbf{Y}} - \mathbf{Y}|| < \epsilon$ for any small ϵ [22]. This means that the ELM can approximate the target $\mathbf{Y}$ as much as we want by adjusting only the number of hidden units and the output weights. The optimisation problem for $\mathbf{W}$ can be written as

$$\underset{\mathbf{W}}{\text{minimise}} \quad ||\mathbf{HW} - \mathbf{Y}||_2, \tag{5}$$

which is a least square optimisation problem. The smallest norm solution is given by

$$\hat{\mathbf{W}} = \mathbf{H}^\dagger \mathbf{Y}, \tag{6}$$

where $\mathbf{H}^\dagger$ is the Moore-Penrose pseudoinverse [6]. This solution for the optimisation problem is greatly fast comparing with the classical classifiers as SVM or backpropagation multi-layer perceptrons. More mathematical details of the ELM algorithm and several comparison with other neural nets can be seen in [21, 22].

Table 1. Summary of best results taken from [4], (* MLP1, ** MLP2)

	RF100	SVM	MLP
Feature vector	Accuracy		
Baseline	80.10	84.95	86.89**
MFCC+Fun	83.01	85.92	**89.32***
Full-Set	80.10	83.50	74.27*
	UAR		
Baseline	67.00	74.07	79.21**
MFCC+Fun	70.43	75.18	**80.85****
Full-Set	65.24	72.46	61.90*

3 Results and discussions

In this section the experiments are presented and discussed. At first, a directly comparable work is introduced. Then, the experimental scheme and the results using ELM are showed.

In a previous work [4], this Furnariidae set (25 species) was classified using 206 records, all the experiments were performed in a cross-validation scheme. Three features sets were evaluated:

- *Baseline*: a set of means and variances computed for the first 17 MFCCs, their deltas and acceleration over the entire song (102 features),
- *MFCC+Fun*: a set of functionals computed only from the MFCCs (531 features),
- *Full-Set*: a set of 6373 state-of-the-art features defined for speech processing (INTERSPEECH 2013 ComParE Challenge [39]).

For the determination of performance, two figures of merit are used: the Accuracy is calculated as the mean recognition rate and the Unweighted Average Recall (UAR) is obtained as the average of the class-specific recalls achieved by the system. The baseline for the bird song identification task was defined based on previous works [13, 15]. In Table 1 the best results are showed, using Random Forest (RF) using 100 trees, support vector machines (SVM) and the two best MLP architectures: one hidden layer with a number of neurons set as $(Num.\ of\ inputs + Num.\ of\ outputs)/2$ (MLP1) and one hidden layer with a number of neurons set equal to the number of inputs (MLP2). It is important to remark that the best UAR result (82.21) was reached using the linear forward selection (LFS) [18] on the *Full-Set* and MLP1, however, it can not be exactly reproduced here for comparison (see [4]).

In the present work, 25 Furnariidae species (751 records) are classified using ELM. The three sets of features presented previously are evaluated using ELM, and compared with MLS+ set (328 features) and MLS++ set (346 features). The MLS+ includes 200 MLS and 128 MLSa coefficients, while the MLS++ added 18 speech-related features (13 MFCCs, pitch, energy, zero-crossing rate, short-term energy entropy and short-term spectral entropy). For all the experiments, a 6-fold cross-validation scheme was performed and for each case, the data were normalised using the percentile 5 and 95 from training partition. The average results for the best configurations are presented in tables 2 and 3. Table 2 shows that the spectral features (MLS and MLSa) improved the accuracy reached by the previos parameterisations, although the UAR is similar as with baseline set. As the MLS+ and MLS++ obtain similar results, it is possible to say that

Table 2. Summary of best results using ELM classifiers.

Feature vector	Accuracy [%]	UAR [%]
MLS+	93.22	84.89
MLS++	**93.61**	84.25
Baseline	91.48	**84.97**
MFCC+Fun	84.55	73.43
Full-Set	82.59	65.56

MFCC and the other prosodic values are not too much useful here. Obviously, both spectral representations are very useful for this issue and it is necessary to explore them more to find the optimal set of characteristics, making feature selection (e.g. with LBS) or incorporating new ones.

As the set of records is notably unbalanced, we propose an experiment to evaluate the balance by repetition. Its utilisation is very extended [39] and it consists in augmenting the minority classes repeating their records. Table 3 shows the results using the balanced data, where the *MLS+* reaches the same percentage as in previous experiment. While it is not possible to conclude that this balance is useful for the system, it is possible to say that it increases the computational cost for the training.

Table 3. Summary of best results using ELM classifiers, balanced.

Feature vector	Accuracy [%]	UAR [%]
MLS+	**93.22**	**84.89**
MLS++	92.13	81.69
Baseline	85.60	78.83
MFCC+Fun	86.43	75.94
Full-Set	86.66	71.75

4 Conclusions

Monitoring bird species allows to surveillance the environmental due to it reflects important ecosystem processes and human activities. This work addresses the bird call classification problem using spectral-based features, and compares the performance with previous proposals. Species from the Furnariidae family which inhabit the Paranaense Littoral region were analysed, and although they are well-known in the community, to focus the study on a big group from the same family is very novel.

Results shows that spectral information seems to be really useful to reach a high performance in this application, even considering diverse recordings sources what hinders the task. The sets of features which obtain the best rates here, were previously defined for speech-related tasks, consequently it would be interesting to define more specific features.

This approach could be improved for developing autonomous tools that allow ornithologists to know which species are present in particular areas in order to do ecological monitoring and management. Specifically, it could to help the labelling of Furnariidae recordings, while it could be used for remote and simultaneous monitoring in different areas.

In future research, these spectral parameterisations will be studied further with other classification schemes as deep neural networks that exploit also the local variability along the records. In this sense, an interesting approach is to use the spectral information as images to feed well-known nets as the AlexNet (to perform transfer learning) or to train a new net using spectral information from several bird families. Also, a more in-depth analysis of specific filter banks is needed to process bird song records an to obtain more useful information.

Acknowledgements

The authors wish to thank to *Agencia Nacional de Promoción Científica y Tecnológica* (ANPCyT)(with PICT-2015-977), *Universidad Nacional del Litoral* (with CAID-PJ-50020150100055LI and CAID-PJ-50020150100059LI) and *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET), from Argentina, for their support.

References

1. The internet bird collection. Reference Reviews **26**(8), 42–43 (2012)
2. Albornoz, E.M., Milone, D.H., Rufiner, H.L.: Spoken emotion recognition using hierarchical classifiers. Computer Speech & Language **25**(3), 556–570 (2011)
3. Albornoz, E.M., Milone, D.H., Rufiner, H.L.: Feature extraction based on bio-inspired model for robust emotion recognition. Soft Computing pp. 1–14 (2016)
4. Albornoz, E.M., Vignolo, L.D., Sarquis, J.A., Leon, E.: Automatic classification of furnariidae species from the paranaense littoral region using speech-related features and machine learning. Ecological Informatics **38**, 39–49 (2017)
5. Areta, J.I., Pearman, M.: Species limits and clinal variation in a widespread high andean furnariid: The buff-breasted earthcreeper (upucerthia validirostris). The Condor **115**(1), 131–142 (2013)
6. Ben-Israel, A., Greville, T.N.E.: Generalized inverses: theory and applications. Springer, 2 edn. (2001)

7. Betts, M., Mitchell, D., Diamond, A., Bêty, J.: Uneven rates of landscape change as a source of bias in roadside wildlife surveys. The Journal of Wildlife Management **71**(7), 2266–2273 (2007)
8. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001)
9. Burkart, R., Bárbaro, N., Sánchez, R., Gómez, D.: Eco-Regiones de la Argentina. Administración de Parques Nacionales (APN). Secretaría de Recursos Naturales y Desarrollo Sostenible, Presidencia de la Nación Argentina (1999)
10. Byers, B.E.: Geographic variation of song form within and among chestnut-sided warbler populations. The Auk pp. 288–299 (1996)
11. Chi, T., Ru, P., Shamma, S.A.: Multiresolution spectrotemporal analysis of complex sounds. Journal of the Acoustical Society of America **118**(2), 887–906 (2005)
12. Contreras, J.R., Agnolin, F., Davies, Y.E., Godoy, I., Giacchino, A., Ríos., E.E.: Atlas ornitogeográfico de la provincia de Formosa. Vazquez Mazzini (2014)
13. Dufour, O., Artieres, T., Glotin, H., Giraudet, P.: Soundscape Semiotics - Localization and Categorization, chap. Clusterized Mel Filter Cepstral Coefficients and Support Vector Machines for Bird Song Identification. InTech Open Book (2014)
14. Eyben, F., Weninger, F., Gross, F., Schuller, B.: Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. In: 21st ACM International Conference on Multimedia. pp. 835–838. Barcelona, Spain (October 2013)
15. Fagerlund, S.: Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing **2007**(1), 64–64 (2007)
16. Ganchev, T.D., Jahn, O., Marques, M.I., de Figueiredo, J.M., Schuchmann, K.L.: Automated acoustic detection of vanellus chilensis lampronotus. Expert Systems with Applications **42**(15-16), 6098 –6111 (2015)
17. Giannoulis, D., Benetos, E., Stowell, D., Rossignol, M., Lagrange, M., Plumbley, M.D.: Detection and classification of acoustic scenes and events: an ieee aasp challenge. In: Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)
18. Gütlein, M., Frank, E., Hall, M., Karwath, A.: Large-scale attribute selection using wrappers. In: Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on. pp. 332–339. IEEE (2009)
19. Harris, C.M.: Absorption of sound in air versus humidity and temperature. The Journal of the Acoustical Society of America **40**(1), 148–159 (1966)
20. Hesler, N., Mundry, R., Dabelsteen, T.: Does song repertoire size in common blackbirds play a role in an intra-sexual context? Journal of Ornithology **152**(3), 591–601 (2011)
21. Huang, G., Huang, G.B., Song, S., You, K.: Trends in extreme learning machines: A review. Neural Networks **61**, 32 – 48 (2015)
22. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: Theory and applications. Neurocomputing **70**(1-3), 489–501 (2006)
23. ICML int. Conf.: Proc. 1st workshop on Machine Learning for Bioacoustics - ICML4B (2013), http://sabiod.univ-tln.fr
24. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planque, R., Rauber, A., Fisher, R., Müller, H.: Lifeclef 2014: Multimedia life species identification challenges. In: Kanoulas, E., Lupu, M., Clough, P., Sanderson, M., Hall, M., Hanbury, A., Toms, E. (eds.) Information Access Evaluation. Multilinguality, Multimodality, and Interaction, Lecture Notes in Computer Science, vol. 8685, pp. 229–249. Springer International Publishing (2014)

25. Keen, S., Ross, J.C., Griffiths, E.T., Lanzone, M., Farnsworth, A.: A comparison of similarity-based approaches in the classification of flight calls of four species of north american wood-warblers (parulidae). Ecological Informatics **21**, 25–33 (2014)
26. Laje, R., Mindlin, G.B.: Highly structured duets in the song of the south american hornero. Physical review letters **91**(25), 258104 (2003)
27. Leon, E.J., Beltzer, A.H., Olguin, P.F., Reales, C.F., Urich, G.V., Alessio, V., Cacciabué, C.G., Quiroga, M.A.: Song structure of the golden-billed saltator (saltator aurantiirostris) in the middle parana river floodplain. Bioacoustics **24**(2), 145–152 (2015)
28. Louette, M., Bijnens, L., Upoki Agenong'a, D., Fotso, R.: The utility of birds as bioindicators: case studies in equatorial africa. Belgian Journal of Zoology **125**(1), 157–165 (1995)
29. Marler, P.: Three models of song learning: Evidence from behavior. Journal of Neurobiology **33**(5), 501–516 (1997)
30. Narosky, T., Yzurieta, D.: Aves de Argentina y Uruguay–Birds of Argentina & Uruguay: Guía de Identificación Edición Total–A Field Guide Total Edition. Buenos Aires, 16 edn. (2010)
31. Päckert, M., Martens, J., Kosuch, J., Nazarenko, A.A., Veith, M.: Phylogenetic signal in the song of crests and kinglets (aves: Regulus). Evolution **57**(3), 616–629 (2003)
32. Payne, R.B.: Song traditions in indigo buntings: origin, improvisation, dispersal, and extinction in cultural evolution. Ecology and evolution of acoustic communication in birds pp. 198–220 (1996)
33. Planqué, B., Vellinga, W.P.: Xeno-cano.org, `http://www.xeno-canto.org`, accessed: 2015-07-10
34. Potamitis, I.: Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity. Ecological Informatics **26, Part 3**, 6–17 (2015)
35. Potamitis, I., Ntalampiras, S., Jahn, O., Riede, K.: Automatic bird sound detection in long real-field recordings: Applications and tools. Applied Acoustics **80**, 1–9 (2014)
36. Ptacek, L., Machlica, L., Linhart, P., Jaska, P., Muller, L.: Automatic recognition of bird individuals on an open set using as-is recordings. Bioacoustics **25**(1), 1–19 (2015)
37. Raposo, M.A., Höfling, E.: Overestimation of vocal characters in suboscine taxonomy (aves: Passeriformes: Tyranni): causes and implications. Lundiana **4**(1), 35–42 (2003)
38. Roch, M.A., Soldevilla, M.S., Burtenshaw, J.C., Henderson, E.E., Hildebrand, J.A.: Gaussian mixture model classification of odontocetes in the southern california bight and the gulf of california. The Journal of the Acoustical Society of America **121**(3), 1737–1748 (2007)
39. Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Mortillaro, M., Salamin, H., Polychroniou, A., Valente, F., Kim, S.: The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism. Proc. Interspeech, ISCA pp. 148–152 (Aug 2013)
40. Stowell, D., Plumbley, M.D.: Segregating event streams and noise with a markov renewal process model. Journal of Machine Learning Research **14**, 1891–1916 (2013)
41. Towsey, M., Wimmer, J., Williamson, I., Roe, P.: The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecological Informatics **21**, 110–119 (2014)

42. Ventura, T.M., de Oliveira, A.G., Ganchev, T.D., de Figueiredo, J.M., Jahn, O., Marques, M.I., Schuchmann, K.L.: Audio parameterization with robust frame selection for improved bird identification. Expert Systems with Applications **42**(22), 8463–8471 (2015)
43. Weninger, F., Eyben, F., Schuller, B.W., Mortillaro, M., Scherer, K.R.: On the acoustics of emotion in audio: what speech, music, and sound have in common. Frontiers in Emotion Science **4**(292), 1–12 (2013)
44. Woolley, S.M., Fremouw, T.E., Hsu, A., Theunissen, F.E.: Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nature neuroscience **8**(10), 1371–1379 (2005)
45. Yang, X., Wang, K., Shamma, S.A.: Auditory representations of acoustic signals. IEEE Transactions on Information Theory **38**(2), 824–839 (1992)
46. Zimmer, K.J., Whittaker, A.: The rufous cacholote (furnariidae: Pseudoseisura) is two species. The Condor **102**(2), 409–422 (2000)
47. Zollinger, S.A., Brumm, H.: Why birds sing loud songs and why they sometimes don't. Animal Behaviour **105**, 289–295 (2015)