Ecological Informatics 44 (2018) 76–81

Contents lists available at ScienceDirect

Ecological Informatics

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

Bird species identification via transfer learning from music genres

Stavros Ntalampiras

Department of Informatics, Università degli studi di Milano, via Comelico 39, 20135 Milan, Italy

ARTICLE INFO

Keywords:
Bird species identification
Transfer learning
Echo State Network
Biodiversity monitoring

ABSTRACT

Humans possess the ability to apply previously acquired knowledge to deal with novel problems quite efficiently. *Transfer Learning* is inspired by exactly that ability and has been proposed to handle cases where the available data come from diverse feature spaces and/or distributions. This paper proposes to transfer knowledge existing in music genre classification to identify bird species, motivated by the existing acoustic similarities. We propose a *Transfer Learning* framework exploiting the probability density distributions of ten different music genres for acquiring a degree of affinity between the bird species and each music genre. To this end, we exploit a feature space transformation based on Echo State Networks. The results reveal a consistent average improvement of 11.2% in the identification accuracy of ten European bird species.

1. Introduction

Monitoring animal communities is becoming increasingly important due to major environmental challenges including invasive species, infectious diseases, climate and land-use change, etc. The effectiveness of the adopted conservation measures is heavily based on the availability of accurate information regarding the range, population size and trends, such that one may assess the conservation status in a species-specific manner. To this end, classical observer-based survey techniques may be employed; nonetheless they are a) costly, b) subject to weather conditions, c) cover a limited amount of time, etc. Thus, in the recent years there has been an increasing interest for automated acoustic monitoring of sound-emitting animals, which may provide reliable information on the presence/absence of target species and on the general biodiversity status of an area.

As a consequence, autonomous recording units (ARUs) have been widely spread among biologists in the recent years to study and analyse different taxonomic groups of sound-producing animals, such as mammals, birds, amphibians, and insects (Fagerlund, 2007; Potamitis, 2015; Grill and Schlüter, 2017; Lopes et al., 2011; Harma, 2003; Fanioudakis and Potamitis, 2017; Potamitis, 2014; Buxton and Jones, 2012; Bardeli et al., 2010; Potamitis et al., 2014) despite their limitations (Wolfgang and Haines, 2016). As ARUs are capable of long-term continuous operation, one is able to collect enormous amount of data in relatively short periods of time. In this context, manual annotation of the acquired audio is not a feasible practise due to cost and time constraints. Thus, automatic analysis of the audio data is mandatory and may facilitate decision making in a series of topics, such as: a) monitoring of range shifts of animal species due to climate change, b) biodiversity assessment and inventorying of an area, c) estimation of species richness and abundance, and d) assessing the status of threatened species.

The basic problems which need to be faced by an automated mechanism dedicated to the classification of bird species are the following: a) it has to operate with reliability across a potentially large number of species (Stowell and Plumbley, 2014), b) it should be able to handle big data as ARUs can capture huge volumes (Aide et al., 2013), and c) be robust to non-stationary environmental noise and sound events (Ntalampiras et al., 2012). As current approaches for large scale bird species identification do not provide high performance, e.g. Wolfgang and Haines (2016) , we propose to complement the way the specific problem is handled so far, and exploit the *Transfer Learning* (TL) logic, i.e. instead of looking at the bird audio signals alone, we propose to statistically analyse them using their similarities with music genres. There are several motivations behind following the specific research path:

- Several musicologists share the belief that the development of music was affected by birdsong to a relatively large extent (Head, 1997; Clark and Rehding, 2005).
- Birds vocalize at traditional scales used in human music (e.g. pentatonic, diatonic) suggesting that birdsong may be thought as music the way humans perceive it (Rothenberg, 2006).
- Famous composers have employed birdsong as a compositional springboard in several genres, e.g. classical (Vivaldi, Beethoven, Wanger, etc.) and jazz (Paul Winter, Jeff Silverbush, etc.) (Franks,

E-mail address: stavros.ntalampiras@unimi.it.
URL: http://sites.google.com/site/stavrosntalampiras/home.

https://doi.org/10.1016/j.ecoinf.2018.01.006
Received 8 September 2017; Received in revised form 25 December 2017; Accepted 31 January 2018
Available online 09 February 2018

2016; Reich, 2010; Thompson, 2014) suggesting perceptual similarities in the respective acoustic structures.

These motivation points suggest that there exists a great variety of perceptual similarities between musical pieces and birdsongs. Thus, tools and methodologies for incorporating knowledge coming from the music information processing field into identifying birdsongs are worth investigating.

This work proposes to exploit the similarities between birdsongs and music genres as assessed by probabilistic modelling for bird species identification. This approach can be grouped under the umbrella of *transfer learning* (Pan and Yang, 2010), which is a relatively new learning framework, where the fundamental characteristic is the transferring of knowledge between two or more classification tasks. Transfer learning has not been extensively exploited by the audio signal processing community. An initial effort among speech and music signals is reported in Coutinho et al. (2014) , where a denoising autoencoder is employed for feature transformation. A more recent method is explained in Phan et al. (2016) where generalized audio events are classified based on their similarities to speech patterns. In the same direction, Lim et al. (2016) train a deep neural network using speech data for sound event classification.

This article aims at presenting the possibilities of transfer learning for bird sound identification. To this end, the dataset is composed of a selection of species covering common European bird species including regular breeding, wintering and migrant ones. On top of that, the Random Forest (RF) classification approach was exploited since it could be considered as a baseline system for many practical applications without making assumptions on the feature set side, such as sequentiality, normalization. More specifically, this work proposes a statistical model approximating the probability density function of each music genre, while taking into account the maximum classification rate criterion. Subsequently, these models are fed with transformed data coming from the birdsong dataset. The probabilities produced by the music genre models are used as features to characterize each species and subsequently classify them using an RF. This process is illustrated in Fig. 1.

2. Transfer learning for bird species identification

This section presents the TL direction followed in this work. We propose to exploit the statistical resemblance between bird calls and musical data as it is assessed via hidden Markov models (HMMs) trained on different music genres, while using the feature space transformation based on Reservoir Networks (*tRN*). The proposed approach encompasses a *tRN* learning a multiple input multiple output transfer function, responsible for transforming features extracted out of bird species to features representative of specific music genres. The following subsections are dedicated to explain the HMM modelling process as well as the training of the Reservoir Network (RN).

2.1. HMM based statistical similarity assessment

The idea behind this approach is that the similarity between music genres and bird species may reveal important information towards the identification of bird species. To this end, we propose to quantify the degree of resemblance between calls coming from a specific species and different music genres using HMMs approximating the distributions exhibited by the music genres. The block diagram of the proposed TL method is shown in Fig. 1.

We assume availability of a music genre dataset $\mathscr{M}$ including N different genres. The mechanism for recognizing music genres relies on well known techniques which have been proven efficient in the last years (Casey et al., 2008). In this work, we are not focussed on the features extraction or the pattern recognition phases; we rather wish to analyse potential transfer of knowledge derived from the music model to the bird call representation. Thus, the feature set comprised the Mel-frequency cepstral coefficients and the classifier is based on left-right HMMs composed of diagonal Gaussian mixture models.

In brief, we employed a triangular Mel scale filterbank for extracting 23 log-energies. Firstly the audio signal is windowed and the short-time Fourier transform (STFT) is computed. The outcome of the STFT passes through the filterbank and the logarithm is computed to adequately space the data. Lastly the discrete cosine transform is applied for decorrelating the data; a procedure which may enhance the performance of the pattern recognition algorithm. The velocity and acceleration coefficients were also appended to the final vector for capturing its dynamics. In the next, the function outputting the features F_v of the audio signal y_t is denoted as f, i.e. $F_v = f(y_t)$.

Pattern recognition is achieved by HMMs, which constitute probabilistic machines able to automatically identify a sequence of patterns within a stream of data (Rabiner, 1989). They are configured in a left-right topology as the nature of the music genres suggests, meaning that only left to right movements are allowed between the states (Ntalampiras, 2014; Panagakis et al., 2010). $\mathscr{M}$ is divided to train T_M and evaluation E_M sets. One HMM is created to represent a specific

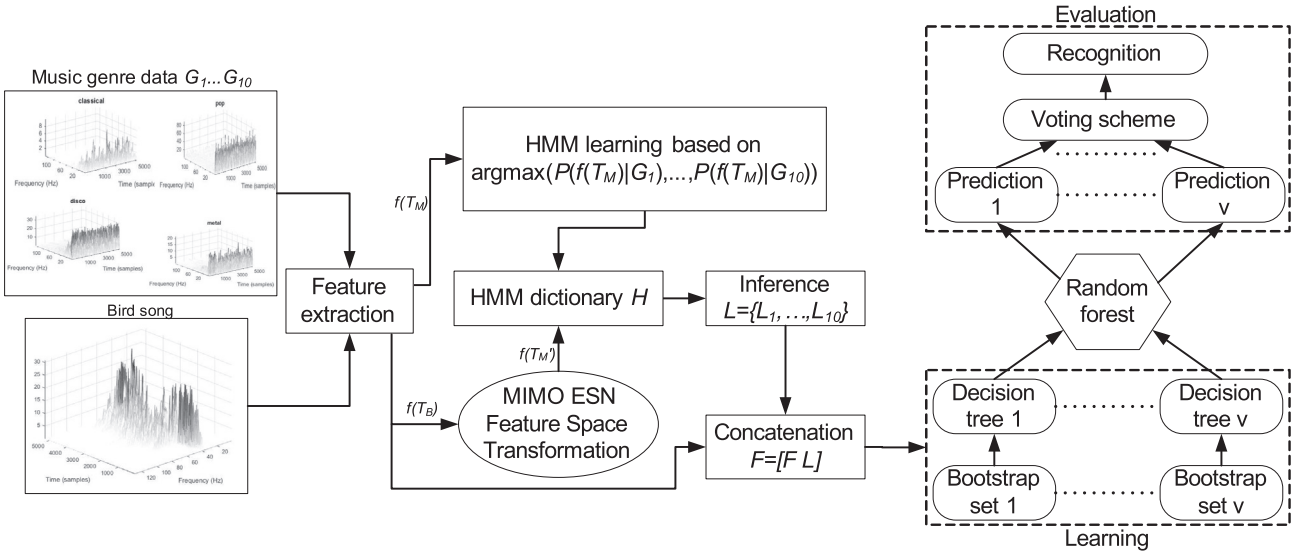

Fig. 1. The block diagram for transfer learning using the statistical affinity based on HMMs trained on music genres (*mHMM*).

genre, i.e. $\mathscr{H}_{\mathscr{M}} = \{H_1, ..., H_N\}$, using data taken from T_M, while the constructed models are used for computing a degree of resemblance (e.g. log-likelihood) between each model and the unknown input signal. This type of score is compared against the rest and the final decision is made based on the maximum log-likelihood criterion. The number of states and Gaussian modes are selected as the ones providing the highest classification accuracy on the testing data E_M.

In order to quantify the similarity between a bird call y_b and a music genre, the next steps are followed

- We apply f onto y_b, i.e. $F_b = f(y_b)$,
- Feed the feature vector F_b to the transformation module explained in Section 2.2,
- Provide the output of the transformation module to a set of models $\mathscr{H}_{\mathscr{M}}$,
- Obtain a vector of log-likelihoods $L = \{l_1,...,L_N\}$ which is now juxtaposed to F_b forming the TL feature vector F_{TL}.

Finally bird species identification is realized using Random Forests (RF), which comprise an integrated classifier composed of decision each one depending on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001; Wei and Li, 2013). After selecting the training instances from the set T_B corresponding to the bird call data, the RF algorithm generates ν training subsets. ν decision trees are "growing" based on data coming from each training subset. Each tree is trained in an independent from the rest process leafing to unique rules, while in the end they form a forest. Finally the prediction is based on a majority voting scheme based on the ν decision trees votes. The most popular class is assigned to the unknown testing instance. A schematic of the RF classification logic is included in Fig. 1.

2.2. Feature space transformation based on reservoir networks

In this work, feature space transformation is a necessary mechanism permitting a model trained on music signals to be used for identifying bird species, while addressing the diversities existing in the feature distributions. We overcome the particular obstacle by learning an RN-based transformation (Jalalvand et al., 2011; Verstraeten et al., 2006). It should be mentioned that this process could be vice-versa, i.e. exploiting bird species models for classifying music genres, which is part of our future study.

Initially each bird species is randomly associated with a music genre. Afterwards, a multiple input multiple output (MIMO) transformation is learnt using the training data of the music genre and the training data of the specific bird species.

RN modelling was employed at this stage as it is able to capture the non-linear relationships existing in the data. RNs represent a novel kind of echo-state networks providing good results in several demanding applications, such as speech recognition (Verstraeten et al., 2006), saving energy in wireless communication (Jaeger and Haas, 2004).

An RN, the topology of which is depicted in Fig. 2, includes neurons with non-linear activation functions which are connected to the inputs (input connections) and to each other (recurrent connections). These two types of connections have randomly generated weights, which are kept fixed during both the training and operational phase. Finally, a linear function is associated with each output node.

Recurrent neural networks aim at capturing the characteristics of high-level abstractions existing in the acquired data while designing multiple processing layers of complicated formations, i.e. non-linear functions. Reservoir computing argues that since back propagation is computationally complex but typically does not influence the internal layers severely, it may be totally excluded from the training process. On the contrary, the readout layer is a generalized linear classification/regression problem associated with low complexity. In addition any potential network instability is avoided by enforcing a simple constraint

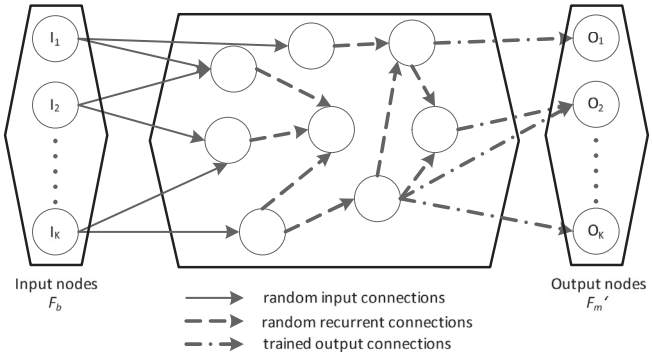

Fig. 2. The Reservoir Network used for feature space transformation.

on the random parameters of the internal layers.

In the following we explain a) how the *tRN* learns the transformation from the bird feature space $\mathscr{S}_{\mathscr{B}}$ to the music one $\mathscr{S}_{\mathscr{M}}$, and b) how the transformation is employed.

2.2.1. RN learning

The *tRN* is used to learn the relationships existing in the features spaces of bird and music signals. We assume that an unknown system model is followed, which may be described as a transfer function f_{RN}.

f_{RN} comprises an RN with N inputs and N outputs. Its parameters are the weights of the output connections and are trained to achieve a specific result, i.e. a bird feature vector. The output weights are learned by means of linear regression and are called read-outs since they "read" the reservoir state (Lukoševičius and Jaeger, 2009). As a general formulation of the RNs, depicted in Fig. 2, we assume that the network has K inputs, L neurons (usually called reservoir size), K outputs, while the matrices $W_{in}(K \times L)$, $W_{res}(L \times L)$ and $W_{out}(L \times K)$ include the connection weights. The RN system equations are the following:

$$x(k) = f_{res}(W_{in}u(k) + W_{res}x(k)) \tag{1}$$

$$y(k) = f_{out}(W_{out})x(k), \tag{2}$$

where $u(k)$, $x(k)$ and $y(k)$ denote the values of the inputs, reservoir outputs and the read-out nodes at time k respectively. f_{res} and f_{out} are the activation functions of the reservoir and the output nodes, respectively. In this work we consider $f_{res}(x) = \tanh(x)$ and $f_{out}(x) = x$.

Linear regression is used to determine the weights W_{out},

$$W_{out} = \arg\min_{W}\left(\frac{1}{N_{tr}}\left\|XW - D\right\|^2 + \epsilon\,\|W\|^2\right) \tag{3}$$

$$W_{out} = (X^TX + \epsilon I)^{-1}(X^TD), \tag{4}$$

where XW and D are the computed vectors, I a unity matrix, N_{tr} the number of the training samples while ϵ is a regularization term.

The recurrent weights are randomly generated by a zero-mean Gaussian distribution with variance ν, which essentially controls the spectral radius SR of the reservoir. The largest absolute eigenvalue of W_{res} is proportional to ν and is particularly important for the dynamical behaviour of the reservoir (Verstraeten et al., 2007). W_{in} is randomly drawn from a uniform distribution $[-InputScalingFactor, +InputScalingFactor]$, which emphasises/deemphasises the inputs in the activation of the reservoir neurons. It is interesting to note that the significance of the specific parameter is decreased as the reservoir size increases.

Here, f_{RN} adopts the form explained in Eqs. (1),(2) by substituting $y(k)$ with F_m and $u(k)$ with F_b, where F_b denotes an original bird feature vector.

2.2.2. Application of f_{RN}

After the learning process, f_{RN} may be thought as a multiple input

multiple output (MIMO) model of the form:

$$\begin{pmatrix} F_m^{1'}(t) \\ F_m^{2'}(t) \\ \vdots \\ F_m^{K'}(t) \end{pmatrix} = f_{RN} \begin{pmatrix} F_b^1(t) \\ F_b^2(t) \\ \vdots \\ F_b^K(t) \end{pmatrix}$$

where the bird features $F_b^1 \ldots, F_b^K$ at time t are transformed to observations belonging to music features $F_m^{1'} \ldots, F_m^{K'}$ using f_{RN}, where K denotes the dimensionality of the feature vector shared by both domains.

3. Experiments

This section offers a thorough analysis of the enhancement provided by TL with respect to identifying bird species. It is divided into three parts: a) the first part describes the datasets used in the experiments, b) the second part the parametrization of the proposed TL framework, c) while the third part presents and analyses the identification results.

3.1. Datasets

Following the proposed TL logic we employed two datasets: a) the GTZAN corpus, which has been acquired by Tzanetakis (Tzanetakis and Cook, 2002) and well-suits classification of music genres (Huang et al., 2014; Costa et al., 2012; Hamel et al., 2013). It is publicly available and includes the following ten genres: *blues*, *classical*, *country*, *disco*, *hiphop*, *jazz*, *metal*, *pop*, *reggae* and *rock*. b) a corpus of the following ten bird species: *Acrocephalus melanopogon*, *Calidris canutus*, *Carduelis chloris*, *Emberiza citrinella*, *Falco columbarius*, *Lanius collurio*, *Larus melanocephalus*, *Parus palustris*, *Sylvia sarda*, *Turdus torquatus*. It includes European bird species covering regular breeding, wintering and migrant ones, while is was obtained from http://www.xeno-canto.org/. All the signals were downsampled to 16 kHz with 16-bit analysis. In Tables 1 and 2, we tabulate the statistics of both datasets.

3.2. Parametrization of the proposed framework

3.2.1. Signal processing

To ensure robustness against possible misalignments, the feature extraction module operated on frames with size 30 ms, while the hop-size was 10 ms. The data frames were hamming-windowed to smooth any potential discontinuities and the FFT size was 512.

3.2.2. Audio pattern recognition

The construction of the probabilistic models was based on the Torch machine learning library (available at http://www.torch.ch). For each experimental phase the maximum number of k-means iteration was 50 and the limit with respect to the Baum-Welch algorithm was 25 while the threshold between subsequent iterations was 0.001.

The number of states was between 3 and 7 while the respective numbers of Gaussian components was taken from the set {2,4,8,16,32,64,128}. The models providing the highest recognition rates were finally selected. This phase comprises the most computationally intensive part of the proposed methodology, albeit the specific phase is meant to run only once and offline. During the classification phase the system only executes the Viterbi algorithm, which is computationally inexpensive as it is based on recursive dynamic programming, and simple log-likelihood comparisons (Viterbi, 1967).

The Random Forest training procedure originates from the general technique of bootstrap aggregating. The standard algorithm described in Frank et al. (2010) was employed, while the number of trees was 100.

Table 1
The characteristics of the dataset of musical genres used in the present work.

Music genre	Duration (s)
Blues	3000
Classical	3000
Country	3000
Disco	3000
Hiphop	3000
Jazz	3000
Metal	3000
Pop	3000
Reggae	3000
Rock	3000

Table 2
The characteristics of the dataset of bird calls used in the present work.

Bird species	Duration (s)
Acrocephalus melanopogon	1320
Calidris canutus	1249
Carduelis chloris	1300
Emberiza citrinella	1382
Falco columbarius	1426
Lanius collurio	1288
Larus melanocephalus	1585
Parus palustris	1651
Sylvia sarda	1485
Turdus torquatus	1362

3.2.3. MIMO ESN feature space transformation

The parameters of the RN were selected by means of exhaustive search. They were taken from the following sets:

- $SR \in \{0.8,0.9,0.95,0.99\}$,
- $L \in \{0,500,1000,5000,10000\}$, and
- $InputScalingFactor \in \{0.1,0.5,0.7,0.95,0.99\}$.

The combination of parameters providing the lowest reconstruction error on a validation set is including both feature spaces, i.e. $\mathscr{S}_{\mathscr{M}}$ and $\mathscr{S}_{\mathscr{B}}$. Its implementation was based on the Echo State Network Toolbox which is available at http://reservoir-computing.org/software.

3.3. Experimental results

We followed the stratified tenfold cross validation protocol during the learning and evaluation phases. Table 3 provide the experimental results with and without the proposed TL framework in the form of confusion matrix. The highest rate with respect to each species is emboldened. The superiority of the TL approach is evident as the respective average classification rates are 92.5% and 81.3%. The average performance improvement across all the bird species is 11.2%. The species which benefit the most are *Acrocephalus melanopogon* and *Calidris canutus*, while *Emberiza citrinella* and *Larus melanocephalus* demonstrate constant identification rate. This suggests that the similarities existing between the first two species and the music genres provide useful information for their identification. Overall, we observe that the TL approach improves the accuracy with respect to every partial identification task. Nonetheless the sources of misclassification are quite similar, i.e. *Acrocephalus melanopogon* is confused for *Turdus torquatus* by both approaches. Conclusively we infer that the proposed TL logic exploiting potential similarities between bird species sounds and music genres adds important distinctive information to the feature vector leading to improved classification rates.

4. Conclusions

This article detailed a novel solution for automatic bird species

Table 3
The confusion matrix which includes the classification rates reached by the proposed TL framework and the methodology without TL. The presentation format is the following: TL/no TL. Average values over 50 iterations are shown.

Presented	Responded									
	Acrocephalus melanopogon	*Calidris canutus*	*Carduelis chloris*	*Emberiza citrinella*	*Falco columbarius*	*Lanius collurio*	*Larus melanocephalus*	*Parus palustris*	*Sylvia sarda*	*Turdus torquatus*
Acrocephalus melanopogon	**79.4**/56.7	–/3.4	–/–	2.1/4.2	–/–	–/–	–/–	–/–	–/–	18.5/35.7
Calidris canutus	–/–	**87.4**/53.1	–/–	–/–	–/–	8.7/12.8	–/2	–/–	6.6/32.1	–/–
Carduelis chloris	–/–	2.5/3.2	**88.3**/79.1	–/–	–/4	–/–	–/–	6.1/9.6	–/–	3.1/4.1
Emberiza citrinella	–/–	–/–	–/–	**100**/**100**	–/–	–/–	–/–	–/–	–/–	–/–
Falco columbarius	5.5/8.2	–/–	8.1/12.4	–/–	**80.2**/67.4	–/–	6.2/12	–/–	–/–	–/–
Lanius collurio	–/–	–/4.9	–/–	–/–	–/1.9	**100**/93.2	–/–	–/–	–/–	–/–
Larus melanocephalus	–/–	–/–	–/–	–/–	–/–	–/–	**100**/**100**	–/–	–/–	–/–
Parus palustris	–/–	–/–	–/3.7	–/–	–/–	–/3.4	–/–	**100**/92.9	–/–	–/–
Sylvia sarda	–/–	–/–	–/–	1.3/5.2	3.9/6	–/–	–/–	–/–	**89.5**/76.3	5.3/12.5
Turdus torquatus	–/3.9	–/–	–/–	–/–	–/–	–/–	–/1.8	–/–	–/–	**100**/94.3

identification based on their acoustic emissions by exploiting statistical models trained on data associated with diverse music genres. A feature space transformation based on MIMO-RN was designed to characterize the transition among them. The results revealed that TL is able to offer distinctive information, which is particularly useful for classification. To the best of the author's knowledge, this is the first time that information from musical pieces has complimented a system towards bird species identification.

The results of this research may open the path for works of similar logic, i.e. exploiting data, knowledge, features, distributions from specific domain(s) to address problems existing in different but related one (s) and vice-versa. Such relationships could be proven particularly useful in the future by increasing the quantity of training data, thus enhancing the recognition performance in a cross-domain manner. Finally, an important point is that the TL approach reached quite good performance with a rather conventional feature set.

In our future work we aim to evaluate whether the TL-based solution is suitable for large and diverse feature spaces. In such scenarios, the TL module needs to operate in a potentially big data environment, thus the established data processing techniques might need to be updated for managing the involved complexity. However, the main logic for transferring knowledge between domains will remain unaltered. In addition, another important aspect concerns the efficacy of *TL* under real world conditions, where bird vocalizations might be corrupted by environmental noise. More precisely, when a bird species classification framework operates on signals contaminated by non-stationary noise, it is expected that the discrimination capabilities of the extracted feature vector, i.e. $f(T_B)$, will decrease. However, the feature vector part derived from the TL module, i.e. the likelihoods, might be less affected since they are a product of models trained on noise-free data. Thus, an interesting future direction would be the development of noise compensation methods specifically tailored for TL-based solutions.

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. PeerJ 1, e103. http://dx.doi.org/10.7717/peerj.103. Jul.

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 (12), 1524–1534. http://www.sciencedirect.com/science/article/pii/S0167865509002487 Pattern Recognition of Non-Speech Audio.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. http://dx.doi.org/10.1023/A:1010933404324. Oct.

Buxton, R.T., Jones, I.L., 2012. Measuring nocturnal seabird activity and status using acoustic recording devices: applications for island restoration. J. Field Ornithol. 83 (1), 47–60. http://dx.doi.org/10.1111/j.1557-9263.2011.00355.x. Feb.

Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M., 2008. Content-based music information retrieval: current directions and future challenges. Proc. IEEE 96 (4), 668–696 April.

Clark, S., Rehding, A., 2005. Music Theory and Natural Order From the Renaissance to the Early Twentieth Century. Cambridge University Press.

Costa, Y., Oliveira, L., Koerich, A., Gouyon, F., Martins, J., 2012. Music genre classification using {LBP} textural features. Signal Process. 92 (11), 2723–2737. http://www.sciencedirect.com/science/article/pii/S0165168412001478.

Coutinho, E., Deng, J., Schuller, B., 2014. Transfer learning emotion manifestation across music and speech. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 3592–3598 July.

Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP J. Adv. Signal Process. 2007 (1), 038637. http://dx.doi.org/10.1155/2007/38637. May.

Fanioudakis, L., Potamitis, I., 2017. Deep networks tag the location of bird vocalisations on audio spectrograms. eess abs/1711.04347. https://arxiv.org/abs/1711.04347.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., Trigg, L., 2010. Weka-A Machine Learning Workbench for Data Mining. Springer US, Boston, MA, pp. 1269–1277. http://dx.doi.org/10.1007/978-0-387-09823-4_66.

Franks, R., 2016. Six of the best: pieces inspired by birdsong. http://www.classical-music.com/article/six-best-birdsong-pieces February.

Grill, T., Schlter, J., 2017. Two convolutional neural networks for bird detection in audio signals. In: 2017 25th European Signal Processing Conference (EUSIPCO), pp. 1764–1768 Aug.

Hamel, P., Davies, M.E.P., Yoshii, K., Goto, M., 2013. Transfer learning in mir: sharing learned latent representations for music audio classification and similarity. In: 14th International Conference on Music Information Retrieval (ISMIR '13).

Harma, A., 2003. Automatic identification of bird species based on sinusoidal modeling of syllables. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on. vol. 5. pp. V-545-8 April.

Head, M., 1997. Birdsong and the origins of music. J. R. Music. Assoc. 122 (1), 1–23. http://www.jstor.org/stable/766551.

Huang, Y.-F., Lin, S.-M., Wu, H.-Y., Li, Y.-S., 2014. Music genre classification based on local feature selection using a self-adaptive harmony search algorithm. Data Knowl. Eng. 92, 60–76. http://www.sciencedirect.com/science/article/pii/S0169023X14000640.

Jaeger, H., Haas, H., 2004. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science 304 (5667), 78–80. http://science.sciencemag.org/content/304/5667/78.

Jalalvand, A., Triefenbach, F., Verstraeten, D., Martens, J., 2011. Connected digit recognition by means of reservoir computing. In: Proceedings of the 12th annual conference of the International Speech Communication Association, pp. 1725–1728.

Lim, H., Kim, M.J., Kim, H., 2016. Cross-acoustic transfer learning for sound event classification. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2504–2508 March.

Lopes, M.T., Gioppo, L.L., Higushi, T.T., Kaestner, C.A.A., Silla Jr, C.N., Koerich, A.L., 2011. Automatic bird species identification for large number of species. In: 2011 IEEE International Symposium on Multimedia, pp. 117–122 Dec.

Lukoševičius, M., Jaeger, H., 2009. Survey: reservoir computing approaches to recurrent neural network training. Comput. Sci. Rev. 3 (3), 127–149. http://dx.doi.org/10.1016/j.cosrev.2009.03.005. Aug.

Wei, J. m., Li, Y., 2013. Specific environmental sounds recognition using time-frequency texture features and random forest. In: Image and Signal Processing (CISP), 2013 6th International Congress on. vol. 03. pp. 1277–1281 Dec.

Ntalampiras, S., 2014. Directed acyclic graphs for content based sound, musical genre, and speech emotion classification. J. New Music Res. 43 (2), 173–182. http://dx.doi.org/10.1080/09298215.2013.859709.

Ntalampiras, S., Potamitis, I., Fakotakis, N., 2012. Acoustic detection of human activities in natural environments. J. Audio Eng. Soc. 60 (9), 686–695. http://www.aes.org/e-lib/browse.cfm?elib=16373.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359 Oct.

Panagakis, Y., Kotropoulos, C., Arce, G.R., 2010. Non-negative multilinear principal

component analysis of auditory temporal modulations for music genre classification. IEEE Trans. Audio Speech Lang. Process. 18 (3), 576–588 March.
Phan, H., Hertel, L., Maass, M., Mazur, R., Mertins, A., 2016. Learning representations for nonspeech audio events through their similarities to speech patterns. IEEE/ACM Trans. Audio Speech Lang. Process. 24 (4), 807–822 April.
Potamitis, I., 2014. Automatic classification of a taxon-rich community recorded in the wild. PLoS ONE 9 (5), 1–11. http://dx.doi.org/10.1371/journal.pone.0096936. 05.
Potamitis, I., 2015. Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity. Eco. Inform. 26 (Part 3), 6–17. http://www.sciencedirect.com/science/article/pii/S1574954115000102.
Potamitis, I., Ntalampiras, S., Jahn, O., Riede, K., 2014. Automatic bird sound detection in long real-field recordings: applications and tools. Appl. Acoust. 80, 1–9. http://www.sciencedirect.com/science/article/pii/S0003682X14000024.
Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 257–286.
Reich, R., 2010. Njit professor finds nothing cuckoo in serenading our feathered friends. http://www.nj.com/entertainment/music/index.ssf/2010/10/njit_professor_finds_nothing_c.html October.
Rothenberg, D., 2006. Why Birds Sing: A Journey Into the Mystery of Bird Song. Basic Books. https://books.google.gr/books?id=J1gemza3aOkC.
Stowell, D., Plumbley, M.D., 2014. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. CoRR abs/1405.6524. http://arxiv.org/abs/1405.6524.
Thompson, H., 2014. This birds songs share mathematical hallmarks with human music. http://www.smithsonianmag.com/science-nature/birds-songs-share-mathematical-hallmarks-human-music-180953227/?no-ist November.
Tzanetakis, G., Cook, P., 2002. Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. 10 (5), 293–302 Jul.
Verstraeten, D., Schrauwen, B., DHaene, M., Stroobandt, D., 2007. An experimental unification of reservoir computing methods. Neural Netw. 20 (3), 391–403. http://www.sciencedirect.com/science/article/pii/S089360800700038X Echo State Networks and Liquid State Machines.
Verstraeten, D., Schrauwen, B., Stroobandt, D., 2006. Reservoir-based techniques for speech recognition. In: Neural Networks, 2006. IJCNN '06. International Joint Conference on, pp. 1050–1053 July.
Viterbi, A.J., 1967. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE Trans. Inf. Process. 13, 260–269.
Wolfgang, A., Haines, A., 2016. Testing automated call-recognition software for winter bird vocalizations. Northeast. Nat. 23 (2), 249–258.