7.1 Multi-Instance Multi-Label Acoustic Classification of Plurality of Animals : birds, insects & amphibian

O. Dufour* H. Glotin† P. Giraudet‡ Y. Bas§
T. Artières¶

25/11/2013

1 Introduction

Nowadays, consulting firms on environment propose to evaluate impacts of transports and/or power production infrastructures on biodiversity using bioacoustic and adapted algorithms of signal processing. We present here our best algorithm (whose AUC score is 0.85%). This is our contribution to the "Neural Information Processing Scaled for Bioacoustics " (NIPS4B) workshop technical challenge [1] of NIPS 2013. Our objective was to obtain a bird-sound operational classification machine-learning model that environmental engineers (mostly ornithologists) could use to realise automatic inventories of acoustically active animals.

2 Description of the method

Our preprocessing is based on Mel-filter cepstral coefficients which have been proved useful for speech [12, 29] and bird song recognition [26]. A temporal signal is first transformed into a serie of frames (see figure 1 A and B) where each frame consists in 16 mfcc (Mel-filter cepstral coefficients), including energy (first coefficient). Each frame represents a duration of 11.6 ms (e.g 512 temporal bins of a signal sampled at 44 100 Hz). Two successive frames overlap of 33% i.e. 3.9 ms.

2.1 Detection and feature extraction

*LSIS, Université du Sud Toulon Var. olivierlouis.dufour@gmail.com
†Aix-Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13397 Marseille, France. glotin@univ-tln.fr
‡Université du Sud Toulon Var. giraudet@univ-tln.fr
§BIOTOPE. ybas@biotope.fr
¶LIP6, Université Paris 6. thierry.artieres@lip6.fr
[1] In proc. of int. symposium 'Neural Information Scaled for Bioacoustics' joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al.

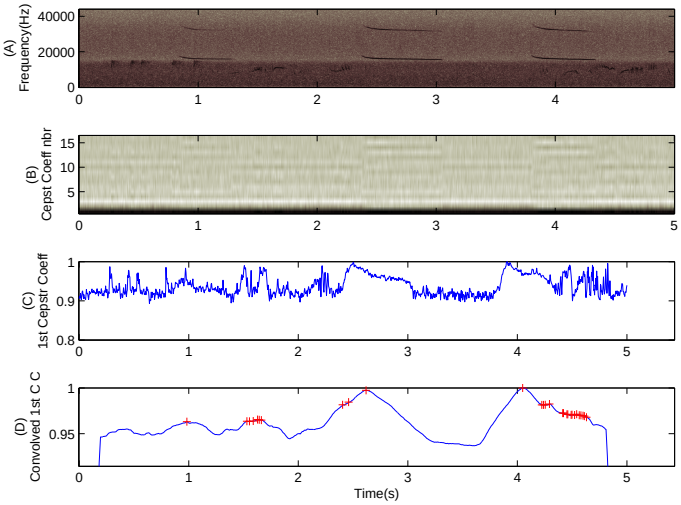

Figure 1: Main steps of the syllables detection.

Detection. To find frames of higher energy (liable to contain a bird syllable), we performed an energy-based detection step. The idea is close to the standard syllable extraction step that is used in most methods for bird identification [30, 10, 8].

1. We compute $E(t)$, $t = \{1, 2, 3, ..., N\}$ (figure 1 C). $E(t)$ is the set of values of the first MFCC from 1 to N. $E(t)$ is the value of the energy in the audio signal contained in the frame number t. N is the number of frames contained in an audio file. For a 5 seconds recording, $N \approx 860$.

2. We compute $Conv$. $Conv$ is the convolution of E by 1_{100}. 1_{100} is a 100-element vector of value "1". It corresponds to a 1.16 s duration.

$$Conv(t) = E(t) * 1_{100} \tag{1}$$

3. We note abscissas of all local maxima superior (figure 1 C) to Th such as:

$$Th = \frac{\sum_{i=50}^{N+50} Conv(i)}{N} \tag{2}$$

or we retain abscissas of the five higher local maxima.

4. For each of D dates, we consider the values of the 16 MFCC from 16 frames before to 15 frames after the frame of the detection. Considering segments of $n = 32$ frames (i.e. about 130 ms duration) means we use windowing.

2.2 Feature extraction

The final step of the preprocessing consists in computing a reduced set of features for any segment. Recall that each segment consists in a series of n 16-dimensional feature vectors (with $n = 32$).

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 165

- **96 coefficients featuring** To get new feature vectors that are representative of longer segments, our feature extraction first consisted in computing 6 values for representing the series of n values for each of the 16 mfcc features. Let consider a particular mfcc feature v, let note $(v_i)_{i=1..n}$ the n values taken by this feature in the n frames of a window and let note $\bar{v_i}$ the mean value of v_i. Moreover let note d and D the velocity and the acceleration of v, which are approximated all along the sequences with $d_i = v_{i+1} - v_i$, and $D_i = d_{i+1} - d_i$. The 6 values we compute are defined as:

$$f_1 = \frac{\sum_{i=1}^{n}(|v_i|)}{n} \tag{3}$$

$$f_2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(v_i - \bar{v_i})^2} \tag{4}$$

$$f_3 = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(d_i - \bar{d_i})^2} \tag{5}$$

$$f_4 = \sqrt{\frac{1}{n-3}\sum_{i=1}^{n}(D_i - \bar{D_i})^2} \tag{6}$$

$$f_5 = \frac{\sum_{i=1}^{n-1}|d_i|}{n-1} \tag{7}$$

$$f_6 = \frac{\sum_{i=1}^{n-2}|D_i|}{n-2} \tag{8}$$

  At the end, a segment in a window is represented as the concatenation of the 6 above features for the 16 cepstral coefficients. It is then a new feature vector s_t (with t the number of the window) of dimension 96.

- **112 coefficients featuring** To get new feature vectors, we also used an algorithm implemented by Vipin Vijayan [36]. Basically this algorithm first realises a PCA on the data and then compute a LDA on the dimensionally reduced data. Contrary to 96 coefficients featuring previously mentioned, in this case the number of features (i.e 112) isn't fixed by human operator but automatically chosen.

In all cases, each audio file is finally represented as a sequence of feature vectors s_t, each representing a duration of about 130 millisecond.

2.3 Training

What makes this challenge so difficult is the fact that:

- there isn't one Multiple-Instance Single-Label training recording per class. One single-label training recording has been provided for only N classes ($K > N$; $K = 87$; $N = 51$);

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 166

- in most of test and train signals, several classes are present. Each audio file is not only represented by multiple instances but also associated with multiple class labels.

Lets consider Tsoumakas definitions from [35]. We define problem transformation methods as those methods that transform the multi-label classification problem either into one or more single-label classification problems, for which there exists a huge bibliography of learning algorithms. We define algorithm adaptation methods as those methods that extend specific learning algorithms in order to handle multi-label data directly.

Problem transformation Method The most common problem transformation method learns $|L|$ binary classifiers ($|L| = 87$), one for each different label l in L. It transforms the original data set into $|L|$ data sets Dl that contain all examples of the original data set, labelled as l, if the labels of the original example contained l and as $\neg l$ otherwise. It is the same solution used in order to deal with a single-label multi-class problem using a binary classifier. We used this approach (dubbed PT) with a Support Vector Machine classifier.

Algorithm adaptation methods One strategy can consist in separating syllables of different classes in the same training recording during preprocessing like in [10, 9]. This is an signal-processing approach. According to [30, 10, 8, 9], we chose to use a machine learning approach. We trust in learning by bag-of-instances in order to realise the tricky task.

Multi-instance multi-label learning (MIML) is a recent learning framework where each example corresponds to a bag of instances as well as a set of labels [25, 41]. To handle this MIML task, we tested different matlab toolboxes from Nanjing University [38, 39, 37]:

- MIMLRBF (MIML Radial Basis Function) is an innovative neural network style algorithm. As its name implied, MimlRbf is derived from the popular radial basis function (RBF) method [4]. Connections between instances and labels are directly exploited in the process of first layer clustering and second layer optimization. Briefly, the first layer of MIMLRBF neural network consists of medoids (i.e. bags of instances) formed by performing k-Medoids clustering on Miml examples for each possible class, where a variant of Hausdorff metric [19] is utilized to measure the distance between bags [40]. Second layer weights of MimlRbf neural network are optimized by minimizing a sum-of- squares error function and worked out through singular value decomposition (SVD) [33].

- MIML-kNN (k-Nearest Neighbor Based Multi-Instance Multi-Label Learning Algorithm) is proposed for MIML by utilizing the populark-nearest neighbor techniques. Given a test example, MIML-kNNnot only considers its neighbors, but also considers its citers whichregard it as their own neighbors. The label set of the test example is determined by exploiting the labeling information conveyed byits neighbors and citers.

- M3MIML (Maximum Margin Method for Multi-instance Multi-label Learning) assumes a linear model for each class, where the output on one class is set to be the maximum prediction of all the MIML examples instances with

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 167

respect to the corresponding linear model. Subsequently, the outputs on all possible classes are combined to define the margin of the MIML example over the classification system. Obviously, each instance is involved in determining the output on each possible class and the correlations between different classes are also addressed in the combination phase. Therefore, the connections between the instances and the labels of an MIML example are explicitly exploited by M3MIML.

Based on the feature extraction step we described above (see section *Detection*) the simplest strategy was to train a MIML classifier from feature vectors s_t which are long enough to include a syllabe or a call. We retained the idea of aggregating all vectors s_t from the same test signal to constitute a bag of present syllables and then let the classifier decide which species are present (see section *Inference*).

2.4 Inference

At test time an incoming signal is first preprocessed as explained before in section 2.2 : interesting segments are selected and feature extraction is performed. Second, a MIML learned model is used as explained before in section 2.3 to compute prediction vectors from the same audio in one K-dimension vector ($K = 87$). This yields that an input signal is represented as one bag of variable number of 96-dimension vectors.

1 000 files compose the test set. The 1 000 bags of vectors obtained after preprocessing are processed by MIML-RBF classifier to get probabilistic scores of each one of the 87 labels sets provided in the train data set.

3 RESULTS

Our detection is based on peaks of energy in time-frequency representation of animals calls and songs. Our 0.85% best score to NIPS4B challenge reveals that it is relevant to focus on higher levels of energy inside an acoustic pattern in order to counteract the intraclass variability of patterns. It is a reasonable biological hypothesis to assert that even if a given species of bird composes complex and variable strophes, it insists more (in terms of signal intensity) on some precise syllables.

Model	NIPS4B Private AUC score	short description
M1	0.7226	5 higher maxima per file + 96 features per segment + PT
M2	0.8247	5 higher maxima per file + 96 features per segment + MIMLRBF
M3	0.6837	5 higher maxima per file + 96 features per segment + MIMLkNN
M6	0.5048	5 higher maxima per file + 96 features per segment + M3MIML
M4	0.8242	all local maxima in a file + 96 features per segment + MIMLRBF
M5	0.8521	all local maxima in a file + PCA/LDA + MIMLRBF
M6	0.8290	5 higher maxima per file + PCA/LDA + MIMLRBF
Mario	0.9175	best team of NIPS4B challenge

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 168

4 Discussion

Figure 2 gives the False Negative Rate (FNR) for each class computed from model M2 predictions on data test set. One can see that the global FNR (all classes included) turns around 25%.

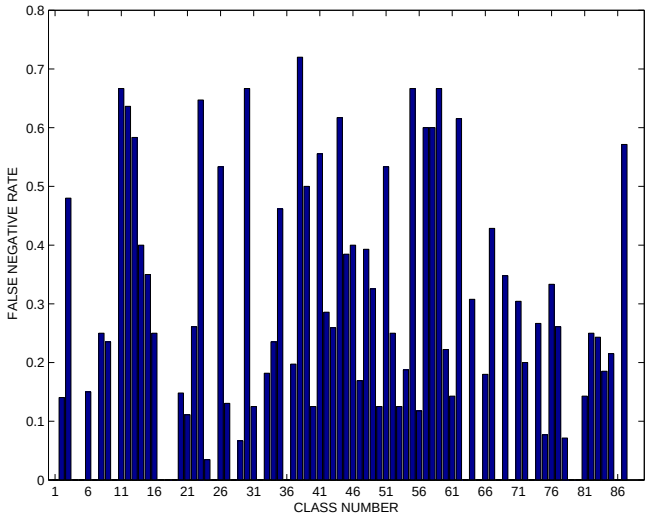

Figure 2: Class-dependant False Negative Rate for NIPS challenge

Expected comments

1. Scores are much better for classes corresponding to bird calls than for classes corresponding to bird songs. By instance, scores of classes number 36, 17, 1, 73, 18 are excellent because the concerned calls consist in strongly stereotyped signals.

2. Predictions remain generally very good for bird species whose songs stay simple and few variable (cl. 25, 65, 70).

3. A FNR of 33% for Subalpine Warbler (cl. 76) on a total of 36 test files is reasonable because it is one of the 4 most difficulty species of the challenge recognized by an ornithologist. most difficult bird species of the challenge.

4. It is well-known that European Robin produces complex and much variable songs (cl. 23). As a consequence, we reach a 65% FNR.

5. Song Thrush and European Serin (cl. 87 & 67) emit complex songs. Their respective scores are 57% et 43%. Although European Serin song is distinctive, it is also composed of a lot of syllables (50 per second). This comforts our hypothesis (see section *Improvements*) that in some cases our currently 130 ms fixed window function is well too large.

Unexpected comments

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 169

1. A 7% FNR regarding class 78 is very encouraging. Among birds species of the challenge, Sardinian Warbler is one of the 4 most difficulty recognized by an ornithologist.

2. Dartford Warbler is equally one of the 4 most complex bird species. Our flawless score must be tooked cautiously: only 3 test files contains this class.

3. Cetti's Warbler and Phylloscopus collybita (cl. 11 & 55) provide good examples of strongly stereotyped signals for which our FNRs keep too high (≈33%)).

4. The error is important concerning cl. 12 and 44 (European Greenfinch calls and Coal Tit songs) whereas their signals aren't particularly complicated because:

   - Train and test recordings providing European Greenfinch examples have a feeble signal-to-noise ratio (S.N.R);
   - In train and test recordings, Coal Tit always accompanies other species;

5. Overall, performances on insects remain disappointing

   - FNR turns around 40% for classes 82 and 14;
   - FNR regarding Common Cicada (cl. 38) is huge (72%) whereas the signal of this species is continuous and stable;

   except for

   - Pygmy Cicada (cl. 81) : 8% FNR. All train and test files concerning Pygmy Cicada come from the same location. Low FNR for this species is probably due to the fact that the model we built detect more the acoustic "signature" of the place rather than the signal of this insect;
   - and Fallow Bush-cricket (cl. 53) : 15% FNR. Its syllables keep similar to bird syllables: they are temporally-speaking punctual.

   This strengthen the idea according to which our current method isn't well compatible with uninterrupted signals. In all likelihood, a part of information concerning uninterrupted signals is lost during MFCC compression by spectral subtraction.

5 Improvements

1. According to figure 3, there is 36 classes for which we don't have any single-label recording. Plus, one can see that the volume of available training data (in seconds) varies much from one class to an other. It is very likely that this disequilibrium brakes performances of our classification algorithm. It will be interesting to watch carefully the differences of classification scores between classes and explain them: are they due to train data set disequilibrium, differences in signal complexities, variable S.N.R, acoustic properties of biotopes, etc. ?

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 170

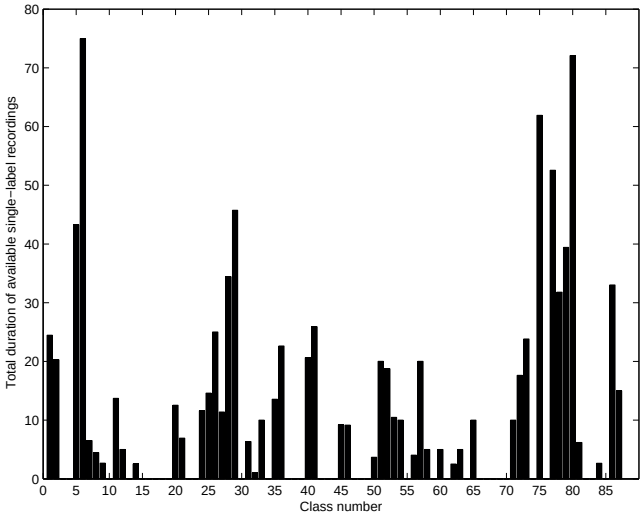

Figure 3: Available singlelabel training recordings total duration (in seconds) per class

2. Given encouraging performances of authors as Lecun, Bengio, Malikov or Abdel-Hamid models in [3, 28, 1, 23], we aim at testing in our future works MIML CNN (Convolutionnal Neural Network) algorithms.

3. One possible way of improvement consists in making variable the size of our currently fixed window function : 130 ms. Some species of birds emits more syllabes per second than others (between 1 up to 60 [6]). Moreover, we could improve learning vectors by adding the information: "Is there others detected syllables close to the considered syllable?".

4. Organizers of the challenge made the effort to label and to provide (Dr Yves Bas) 100 audio files containing only parasite sounds. Parasites constitute the most diversified class because then can be created by an infinity of different ways: car, bike or plane passage, wind, rain, walking sounds, etc. Parasites sounds designates the same type of frequency-and-temporal continuous signals than animals syllables. This is the reason why they complicate the classification task. An other way to improve our model consists in gathering all s_t vectors of all training files and separating them. On the one hand, we have the set S_a containing s_t vectors belonging to animals classes. On the other hand, we have the set S_p containing s_t vectors belonging to parasite class. It is easy to realise separately an optimized clustering of S_a (in $K1$ classes) and S_p vectors (in $K2$ classes). Thus, one can create a $K1 + K2$ multiclassification model by one-vs-all learning approach (binary relevance). This way, after the extraction of s_t from train and test files, we can identify and exclude s_t vectors similar to parasites s_t vectors. This should facilitates afterwards MIML classification task.

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 171

6 Acknowledgments

PhD funds of 1st author are provided by Agence De l'Environnement et de la Maîtrise de l'Energie (mila.galiano@ademe.fr) and by BIOTOPE company (Dr Lagrange, hlagrange@biotope.fr, R&D Manager). We thank Y. Bas and S. Vigant (from BIOTOPE company) who provided and labeled the challenge data.

References

[1] O. Abdel-Hamid, L. Deng, and D. Yu. Exploring convolutional neural network structures and optimization techniques for speech recognition. In *INTERSPEECH*, 2013.

[2] M. Acevedo, C. Corrada-Bravo, H. Corrada-Bravo, L. Villanueva-Rivera, and T. Aide. Automated classification of bird and amphibian calls using machine learning: A comparaison of methods. *Ecological Informatics 4 206214*, 2009.

[3] Y. Bengio and Y. Lecun. Convolutional networks for images, speech, and time-series, 1995.

[4] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.

[5] B. Bogert, M. Healy, and J. Tukey. The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe-cracking. In E. M. Rosenblatt, editor, *Symposium on Time Series Analysis, Chapter 15, p 209-243*, 1963.

[6] A. Bossus and F. Charron. Guide des chants d'oiseaux d'europe occidentale : Description et comparaison des chants et des cris, 2010.

[7] F. Briggs et al. The 9th Annual MLSP Competition: New Methods for Acoustic Classification of Multiple Simultaneous Bird Species in a Noisy Environment. In *IEEE Workshop on Machine Learning for Signal Processing, MLSP 2013*, 2013.

[8] F. Briggs, X. Fern, and R. Raich. Acoustic classification of bird species from syllables: an empirical study. Technical report, 2009.

[9] F. Briggs, X. Z. Fern, and J. Irvine. Multi-label classifier chains for bird sound. *CoRR*, abs/1304.5862, 2013.

[10] F. Briggs, B. Lakshminarayanan, L. Neal, X. Fern, R. Raich, M. Betts, S. Frey, and A. Hadley. Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach. *Journal of the Acoustical Society of America*, 2012.

[11] C.-C. Chang. Libsvm. http://www.csie.ntu.edu.tw/~cjlin/libsvm/, 2008.

[12] L. Chang-Hsing, L. Yeuan-Kuen, and H. Ren-Zhuang. Automatic recognition of bird songs using cepstral coefficients. *Journal of Information Technology and Applications Vol. 1, pp.17-23*, 2006.

[13] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A matlab-like environment for machine learning. In *Big Learning 2011 : NIPS 2011 Workshop on Algorithms, Systems, and Tools for Learning at Scale*, 2011.

[14] H. G. E. Deng, L. and B. Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *International Conference on Acoustic Speech and Signal Processing (ICASSP)*, 2013.

[15] O. Dufour, T. Artières, H. Glotin, and P. Giraudet. Clusterized mel filter cepstral coefficients and support vector machines for bird song identification. *International Machine Learning Conference*, 2013.

[16] O. Dufour, P. Giraudet, T. Artières, and H. Glotin. Automatic bird classification based on mfcc clusters, ranked 4th @ icml4b kaggle 2013 competition. In *Listening in the Wild*, page 11, 2013.

[17] O. Dufour, H. Glotin, T. Artières, and P. Giraudet. Classification de signaux acoustiques : Classification de matrices cepstre par support vector machine. Technical report, Laboratoire Sciences de l'Information et des Systèmes, Université du Sud Toulon Var, 2012.

[18] O. Dufour, H. Glotin, T. Artières, and P. Giraudet. Classification de signaux acoustiques : Recherche des valeurs optimales des 17 paramètres dentrée de la fonction melfcc. Technical report, Laboratoire Sciences de l'Information et des Systèmes, Université du Sud Toulon Var, 2012.

[19] G. A. Edgar. *Measure, topology, and fractal geometry*. Undergraduate texts in mathematics. Springer-Verlag, New York, Berlin, Paris, 1990. Réimpression en 1992, 1995.

[20] D. P. W. Ellis. PLP and RASTA (and MFCC, and inversion) in Matlab, 2005. online web resource.

[21] H. Glotin and O. Dufour. *Clusterized Mel Filter Cepstral Coefficients and Support Vector Machines for Bird Song Identification*. INTECH, 2013.

[22] H. Glotin and J. Sueur. Overview of the first international challenge on bird classification, 2013. online web resource.

[23] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. C. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R.-T. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, C. Zhang, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests. In *ICONIP (3)*, pages 117–124, 2013.

[24] A. Graves, A. rahman Mohamed, and G. E. Hinton. Speech recognition with deep recurrent neural networks. *CoRR*, abs/1303.5778, 2013.

[25] Z. hua Zhou and M. ling Zhang. Multi-instance multilabel learning with application to scene classification. In *In Advances in Neural Information Processing Systems 19*, 2007.

[26] joint to Int. Conf. on Machine Learning. *The 1st International Workshop onf Machine Learning for Bioacoustics (ICML 2013)*, Atlanta, USA, june 2013. Glotin H. et al. http://sabiod.univ-tln.fr/ICML4B2013$_p$*roceedings.pdf*.

[27] E. Kasten, M. Philip, and G. Stuart. Ensemble extraction for classification and detection of bird species. *Ecological Informatics 5 153166*, 2010.

[28] H. Lee, Y. Largman, P. Pham, and A. Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in Neural Information Processing Systems 22*, pages 1096–1104. 2009.

[29] A. Michael Noll. Short-time spectrum and cepstrum techniques for vocal-pitch detection. *Journal of the Acoustical Society of America, Vol. 36, No. 2, pp. 296-302*, 1964.

[30] L. Neal, F. Briggs, R. Raich, and F. X. Time-frequency segmentation of bird song in noisy acoustic environments. In *International Conference on Acoustics, Speech and Signal Processing*, 2011.

[31] A.-V. Oppenheim and R.-W. Schafer. From frequency to quefrency: a history of the cepstrum. *Signal Processing Magazine, Vol 21, Issue 5, pp 95 - 1015*, 2004.

[32] J. Placer and C. Slobodchikoff. A method for identifying sounds used in the classification of alarm calls. *Behavioural Processes 67: 8798*, 2004.

[33] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 1988.

[34] L. Ranjard, H. Ross, and H. Ross. Unsupervised bird song syllable classification using evolving neural networks. *Journal of the Acoustical Society of America, Volume 123, Issue 6, pp. 4358-4368*, 2008.

[35] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *Int J Data Warehousing and Mining*, 2007:1–13, 2007.

[36] H. Yu and J. Yang. A direct lda algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 34:2067–2070, 2001.

[37] M.-L. Zhang. A k-nearest neighbor based multi-instance multi-label learning algorithm. In *ICTAI (2)*, pages 207–212. IEEE Computer Society, 2010.

[38] M.-L. Zhang and Z.-J. Wang. Mimlrbf: Rbf neural networks for multi-instance multi-label learning. *Neurocomputing*, 72(16-18):3951–3956, 2009.

[39] M.-L. Zhang and Z.-H. Zhou. M3MIML: A Maximum Margin Method for Multi-instance Multi-label Learning. In *ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 688–697, Washington, DC, USA, Dec. 2008. IEEE Computer Society.

[40] M.-L. Zhang and Z.-H. Zhou. Multi-instance clustering with applications to multi-instance prediction. *Applied Intelligence*, 31(1):47–68, Aug. 2009.

[41] Z.-H. Zhou, M.-L. Zhang, S.-J. Huang, and Y.-F. Li. Miml: A framework for learning with ambiguous objects. *CoRR*, abs/0808.3231, 2008.

In: proc. of int. symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, joint to NIPS, Nevada, dec. 2013, Ed. Glotin H. et al. 175