An Expert System for Automatic Classification of Sound Signals

Krzysztof Tyburek and Piotr Kotlarz

Institute of Computer Science, Kazimierz Wielki University, Bydgoszcz, Poland

Abstract—In this paper, we present the results of research focusing on methods for recognition/classification of audio signals. We consider the results of the research project to serve as a basis for the main module of a hybrid expert system currently under development. In our earlier studies, we conducted research on the effectiveness of three classifiers: fuzzy classifier, neural classifier and WEKA system for reference data. In this project, a particular emphasis was placed on fine-tuning the fuzzy classifier model and on identifying neural classifier applications, taking into account new neural networks that we have not studied so far in connection with sounds classification methods.

Keywords—*audio descriptors, bird species, fuzzy classification of audio signals, MPEG-7, spectral features of sound.*

1. Introduction

Automation of sound recognition processes is an important part of modern-day technology, widely applied in different areas, such as, for instance, automatic speech-based emotion recognition [1] and speech analysis [2]. Nowadays, with well-developed human speech recognition techniques available, work is also performed on automatic recognition of musical instruments. Here, we focus on developing specific methods for automatic recognition of animal species. We commenced with bird species [3] that are recognized based on the sounds they generate. The methods may find practical application in agriculture, where birds often pose a significant threat to crops (fruit, vegetables, cereals). Proper identification of bird species present in the area where the crops are grown is a prerequisite for the application of appropriate mitigation measures.

Another aspect that inspired us to conduct this research is the potential use of this method for classifying bird species to support the work of ornithologists and foresters aiming to quantify bird populations inhabiting a given area. Hopefully, the methods we develop may also be relied upon to automatically classify sounds in other fields, for instance in speech recognition [4]. The authors have already achieved considerable success in recognizing musical instruments [5], [6] and other sounds generated by technical devices. This project constitutes a subsequent stage of the research activities planned. Both methods are relevant for agriculture, wind farms and airport monitoring. A bird identification method based on camera snapshots

and relying on deep learning techniques was presented in papers [7], [8]. Research [9] used deep learning techniques while processing images from an unmanned aerial vehicle. A classification based on machine learning is presented in [10]. On the other hand, in [11], decision trees were used. Here, approaches that differ from the ones described previously are presented. We propose to classify bird species by using such methods, as neural networks and fuzzy logic with sounds descriptors.

The first section contains a short introduction to problems faced while recognizing bird species. Section 2 is devoted to the research methodology applied. It contains information about the MPEG7 standard, the methodology used to detect bird sound features and a proposal to develop a proprietary research tool. It also presents the result of bird species classification work supported by some classification algorithms included in the WEKA package. It also shows research results obtained with the use of the neural classifier. Section 3 presents the concept of a linguistic audio signal classifier. The last two sections summarize the research conducted. A preliminary model of an expert system is also proposed.

2. Study Methodology

This work focuses on the concept of building an expert system, and on achieving some novel results in the area of using neural networks for the recognition of bird species based on their sound. In order to perform the study, the
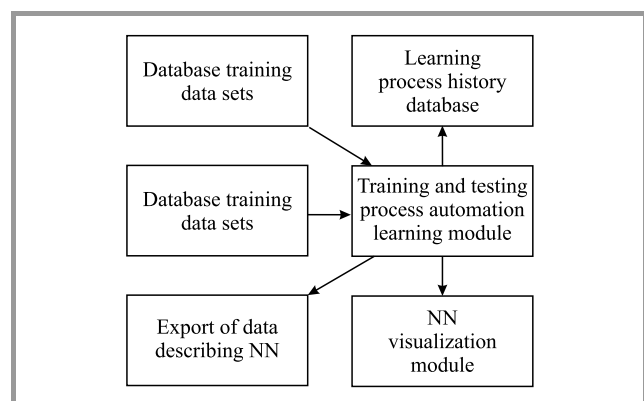
Fig. 1. Development of a proprietary research tool.

Table 1
Results retrieved for the WEKA classifier

	Corn crake	Hawk	Black-bird	Chiff-chaff	Eurasian pygmy owl	Meadow pipit	Lesser whitethroat	Firecrest	Cuckoo	House sparrow
Used k-NN, cross-validation method	98	91	69	84	96	91	98	96	100	78
Used random forest, cross-validation method	98	87	80	82	98	93	100	93	98	80
Used J48 decision tree, cross-vali-dation method	95	80	76	84	95	85	100	87	87	67

authors developed software supervising the learning process of neural networks. The concept of the solution under development is shown in Fig. 1. The development of a test tool facilitated full control over the learning process, offering the ability to relatively quickly assess how effective the specific networks were in the classification of the samples tested. The study itself was conducted using a three-layer neural network with three different neural activation functions. In the first phase, input vectors were determined by using low-level descriptors (LLD) from the MPEG-7 standard [12], [13]. The objective was to identify the vector of traits that will be best for automatic classification of bird species based on the sounds they record and produce. The methods developed in [3] have been used to determine window length in the frequency domain. Such a parameter is understood here as a fragment of the signal in the time domain, which was downloaded at the same time. The length of the window is determined by the spectrum resolution:

$$f_r = \frac{f_s}{n} \ , \tag{1}$$

where: f_r – adopted spectrum resolution (10 Hz), f_s – sampling frequency (44,100 Hz), n – number of samples for the experiment (4410).

If the tested sound is shorter than the window length ($n = 4410$), then the missing values are filled with zeros [5], [6]. In addition, the Blackman window was used for this study. Based on the MPEG-7 standard for audio signals, we have used the spectral centroid (SC) parameter describing the domain of signal frequency for the spectrum:

$$SC = \frac{\sum\limits_{i=0}^{n} A(i) \cdot i}{\sum\limits_{i=0}^{n} A(i)} \ , \tag{2}$$

where: $A(i)$ is the amplitude of the i-th partial (harmonic) and i is the frequency of the i-th partial.

The irregularity of spectrum Ir may be expressed as:

$$Ir = \log \left[20 \sum\limits_{i=2}^{N-1} \left| \log \frac{A(i)}{\sqrt[3]{A(i-1) \cdot A(i) \cdot A(i+1)}} \right| \right] \ , \tag{3}$$

where: $A(i)$ is the amplitude of the i-th partial (harmonic) and N is the number of the available harmonic.

Three characteristics were obtained from each spectrum: SC, Ir and Sum – the sum of all amplitudes in the spectrum. The combination of descriptors gives a 24-component feature vector that describes the bird sound.

2.1. Results obtained for the WEKA System

Reference values obtained from the popular WEKA system were used for testing and validation of neural classifier's capabilities. The methods used included the following: k-nearest neighbors, random forest, Jrip.

Table 1 shows the performance of the WEKA classifier while working on samples retrieved from 10 bird species. The numerical values in the table represent the recognition of the individual bird species.

2.2. Results Obtained for the Neural Classifier

Paper [3] shows that the characteristics vectors obtained by using the defined descriptors contain information allowing the classes that correspond to the bird species whose sound recordings were analyzed to be identified. Networks consisting of 2 and 3 layers and using selected activation functions were analyzed in the study. The learning set had 50 characteristics vectors (50 sets of input signals – expected output values) for 10 classes. The aim of the tests was to identify, in practice, the level of separation (distinction) of the particular classes corresponding to individual bird species. The tests were to determine the further direction of work, i.e. to select and modify the outcomes of the initial sample preparation phase (with sound descriptors included). The learning process was carried out using the Delta rule [14], [15].

$$r = \sigma = \left[d_i - f(w_i^t x) \right] f'(w_i^t x) \ , \tag{4}$$

$$\sigma w_i = c(d_i - y_i) f'(w_i^t x) x \ , \tag{5}$$

where: r – signal learner, σw – correction weights, c – constant learning process, d – error learning, y – matrix outputs, active function, x – matrix inputs.

The mean classification efficiency for the tested activation functions equaled 78%. It also needs to be noted here that such results were achieved for a relatively small learning set, containing 50 samples only. For the average best performing activation function:

$$f(x) = \frac{1}{1 + e^{-\beta x}} \ , \qquad (6)$$

the average of 1200 learning cycles was required. An important observation made at this stage is that some species were identified in a unanimous way, and some were "distorted" by other input data sets (belonging to another species), as shown in the following graphs generated for Eq. (6).

Therefore, we used a modest set of samples for the teaching set. The results obtained allow us to assume that this method is effective in classifying sound signals of various bird species. With a larger set of learners, the achieved level of recognition should be higher. The results obtained from the application of neural networks in various configurations, with different neural activation functions [3] prove to be useful for the development of an expert system. Furthermore, assuming that the expert system can also be deliberately developed as a sensory network (extensive area applications), the change/programming of the scope of actions may be limited to updating the value of the neurons' weights only. This will be a major advantage in the case of low-energy networks, for example. Hence, the use of neural networks to build one of the modules of the expert system seems to be justified.

2.3. Linguistic Audio Signal Classifier

In this research, we aim to develop a methodology for designing an expert system, with one of its modules based on fuzzy logic [16]. In previous studies, in order to develop a fuzzy classifier of sound signals, fuzzy sets were defined as an acceptable range of a given characteristic for the particular bird species. As a result, the values represented by triangular fuzzy sets [3], [16] indicating individual bird species are obtained within the range of a given linguistic variable. The fuzzy sets for the classifier inputs indicating the individual bird species are defined as:

$$Bird_i = \Lambda(x; x_{mean} - 2\Delta_L, \ x_{mean} + 2\Delta)$$
$$\Delta_L = X_{mean} - x_{min}, \ \Delta_L = x_{mean} - x_{min} \ , \qquad (7)$$

where: i – number of bird species x_{min}, x_{max}, x_{mean} – the minimum/maximum/mean value of the descriptor for a given species.

An assumption that is not usually used in fuzzy systems is a feature that is specific of the model adopted. In linguistic variables representing audio descriptors, we ignore the completeness and continuity of the model. We also do not assume that the values of the membership functions of adjacent fuzzy sets add up to unity [17], [18]. the introduction of a sufficiently wide media for the fuzzy sets reduces the risk of error in the operation of the proposed fuzzy system.
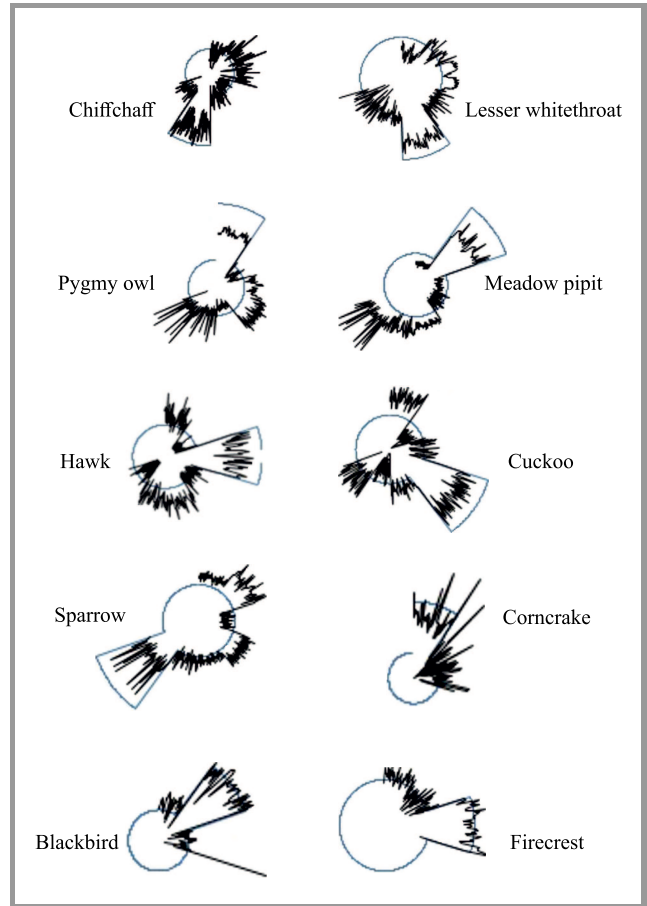

Fig. 2. Effectiveness of the neural classifier in identifying selected bird species.

Hence, the double delta defining the left- and right-hand side fuzzy bank in Eq. (5). Although such an omission of the recommended properties of fuzzy systems seems to be problematic, it is worth noting that a similar approach has already been successfully applied to the classification of lower subspecies of irises [19]. The adopted rule model is defined in the following way:

$$\begin{aligned} &\text{If } \ i_1 \ \text{ is } \ B_1 \wedge i_2 \ \text{ is } \ B_1 \wedge \ldots \ i_i \ \text{ is } \ B_1 \\ &\text{Then } B_1 \ \text{ is } \ rec \wedge B_2 \ \text{ is } \ rej \wedge \ldots \ B_i \ \text{ is } \ rej \end{aligned} \ , \qquad (8)$$

where: i_k – input number k, B – bird (species) number, rec – fuzzy set for the recognition and rej – for the rejection.

The resulting concept of building a modular fuzzy classifier has an advantage consisting in the form of a high scalability of the solution, as the existing structures does not have to be disturbed while adding more species of birds. the expansion may consist in simply adding another fuzzy set of blurred characteristic sets of the new species and in adding a rule that recognizes the new data class. The tests conducted so far have yielded results whose effectiveness ratio equals 65%. These results are not excellent, but we are currently conducting research on improving the data (descriptors) preparation stage and we assume that fuzzy and neural classifiers will complement each other as a result of higher/better classification results.

3. Expert System Concept

Figure 3 shows the modular structure of the designed expert system. It may be used in different hardware implementations, i.e. single-board computers and microcontroller-based embedded systems. The fuzzy logic module is to be an element that complements the neural classifier [20].
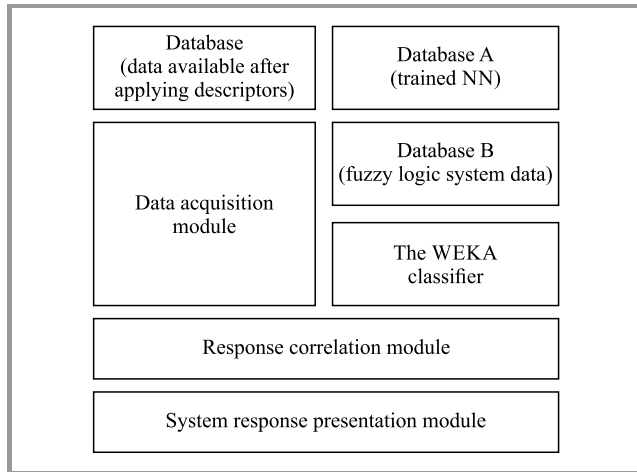

Fig. 3. Concept of an expert system.

4. Conclusions

This work summarizes the second stage of research focusing on sound recognition. In the first stage, the effectiveness of specific descriptors used in the MPEG-7 standard is identified. The present stage focuses on developing methodologies and models required for the expert system. Work planned in the future will allow to develop and test a prototype of the expert system and search for other effective descriptors, also for identifying objects other than bird species. In the future, we intend to develop methods and tools for the construction of expert systems, as indicated in [13], and applications that use the so-called deep learning techniques, for example those described in [20].

References

[1] J. Hook, F. Noroozi, O. Toygar, and G. Anbarjafari, "Automatic speech based emotion recognition using paralinguistics features". *Bull. of the Polish Acad. of Sci.: Techn. Sci.*, vol. 67, no. 3, pp. 479–488, 2019 (DOI: 10.24425/bpasts.2019.129647).

[2] Ł. Mik, A. Lorenc, D. Król, R. Wielgat, R. Święciński, and R. Jędryka "Fusing the electromagnetic articulograph, high-speed video cameras and a 16-channel microphone array for speech analysis", *Bull. of the Polish Acad. of Sci.: Techn. Sci.*, vol. 66, no. 3, pp. 256–266, 2018 (DOI: 10.24425/122106).

[3] K. Tyburek, P. Prokopowicz, P. Kotlarz, and M. Repka. "Comparison of the efficiency of time and frequency descriptors based on different classification conceptions", in *Proc. 14th Int. Conf. on Artif. Intell. and Soft Comput. ICAISC 2015* , Zakopane, Poland, 2015 (DOI: 10.1007/978-3-319-19324-3-44).

[4] M. Kubanek, J. Bobulski, and L. Adrjanowicz "Characteristics of the use of coupled hidden Markov models for audio-visual Polish speech recognition", *Bull. of the Polish Acad. of Sci.: Techn. Sci.*, vol. 60, no. 2, pp. 307–316, 2012 (DOI: 10.2478/V10175-012-0041-6).

[5] K. Tyburek "Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato (Classification of string instruments in multimedia database especially for pizzicato articulation)", Ph.D. Thesis, Institute of Fundamental Technological, Research Polish Academy of Sciences, Warsaw, 2008 (in Polish).

[6] K. Tyburek, W. Cudny, and W. Kosiński "Pizzicato sound analysis of selected instruments in the frequency domain", *Image Process. & Commun.*, vol. 11, no. 1, pp. 53–57, 2006.

[7] J. Niemi and J. T. Tanttu, "Deep learning case study for automatic bird identification", *Appl. Sci.*, vol. 8, no. 11, 2018 (DOI: 10.3390/app8112089).

[8] J. Niemi and J. T. Tanttu, "Automatic bird identification for offshore wind farms: A case study for deep learning" , in *Proc. of Int. Symp. ELMAR 2017*, Zadar, Croatia, 2017 (DOI: 10.23919/ELMAR.2017.8124482).

[9] S.-J. Hong, Y. Han, S.-Y. Kim, A.-Y. Lee, and G. Kim, "Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery", *Sensors*, vol. 19, no. 7, pp. 1–16, 2019 (DOI: 10.3390/s19071651).

[10] S. Balemarthy, A. Sajjanhar, and X. Zheng, "Our practice of using machine learning to recognize species by voice", 2018 [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1810/1810.09078.pdf

[11] M. Lasseck, "Improved Automatic Bird Identification through Decision Tree based Feature Selection and Bagging". in *Proc. Working Notes of CLEF 2015 – Conf. and Labs of the Eval. forum*, Toulouse, France, 2015 [Online]. Available: http://ceur-ws.org/ Vol-1391/160-CR.pdf

[12] B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG-7: Multimedia Content Description Interface*. Chichester: Wiley, 2002 (ISBN: 978-0-471-48678-7).

[13] H. G. Kim, N. Moreau, and T. Sikora, *MPEG7 Audio and Beyond: Audio Content Indexing and Retrieval*. Wiley, 2005 (ISBN: 9780470093344).

[14] K. K. Pathak, S. Panthi, and N. Ramakrishnan "Application of neural network in sheet metal bending process", *Defence Sci. J.*, vol. 55, no. 2, pp. 125–131, 2005 (DOI: 10.14429/dsj.55.1976).

[15] T. Hannagan "The delta rule does Bubbles", *J. of Vision*, vol. 13, no. 8, pp. 1–11, 2013 (DOI: 10.1167/13.8.17).

[16] L. C. de Barros, R. C. Bassanezi, and W. A. Lodwick, *A First Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics. Theory and Applications. Studies in Fuzziness and Soft Computing*, vol. 347. Berlin, Heidelberg: Springer, 2017 (ISBN: 9783662533222).

[17] P. Prokopowicz and D. Ślęzak, "Ordered fuzzy numbers: Definitions and operations", in *Theory and Applications of Ordered Fuzzy Numbers*, P. Prokopowicz, J. Czerniak, D. Mikołajewski, Ł. Apiecionek, and D. Ślęzak, Eds. *Studies in Fuzziness and Soft Computing*, vol. 356, pp. 57–79. Springer, 2017 (DOI: 10.1007/978-3-319-59614-3_4).

[18] P. Prokopowicz, "Processing direction with ordered fuzzy numbers", in *Theory and Applications of Ordered Fuzzy Numbers*, P. Prokopowicz, J. Czerniak, D. Mikołajewski, Ł. Apiecionek, and D. Ślęzak, Eds. *Studies in Fuzziness and Soft Computing*, vol. 356, pp. 81–98. Springer, 2017 (DOI: 10.1007/978-3-319-59614-3_5).

[19] W. Siler and J. J. Buckley, *Fuzzy Expert Systems and Fuzzy Reasoning. Wiley*, 2005 (ISBN: 9780471388593).

[20] P. Prokopowicz, "The use of ordered fuzzy numbers for modelling changes in dynamic processes", *Inform. Sci.*, vol. 470, pp. 1–14, 2019 (DOI: https://doi.org/10.1016/j.ins.2018.08.045).

[21] C. Marechal *et al.*, "Survey on AI-based multimodal methods for emotion detection", in *High-Performance Modelling and Simulation for Big Data Applications*, J. Kołodziej and H. Gonzalez-Velez, Eds. *LNCS*, vol. 11400, pp. 307–324. Springer, 2019 (DOI 10.1007/978-3-030-16272-6_11).

[22] M. Grochowski, A. Kwasigroch, and A. Mikołajczyk, "Selected technical issues of deep neural networks for image classification purposes", *Bull. of the Polish Acad. of Sci.: Techn. Sci.*, vol. 67, no. 2, pp. 363–376, 2019 (DOI: 10.24425/bpas.2019.128485).

Krzysztof Tyburek received his Ph.D. degree in Computer Science from the Institute of Fundamental Technological Research, Polish Academy of Science, Warsaw, Poland, in 2008. His research focuses on database systems, software engineering and digital signal processing, with a particular emphasis placed on sound recognition features. He has over 20 years of academic experience. Currently, Dr. Tyburek is an Assistant Professor at the Institute of Computer Science, Kazimierz Wielki University, Bydgoszcz, Poland.

https://orcid.org/0000-0002-1591-4740
Email: krzysztof.tyburek@ukw.edu.pl
Institute of Computer Science
Kazimierz Wielki University
Kopernika 1
85-074 Bydgoszcz, Poland

Piotr Kotlarz received his Ph.D. degree in Computer Science from the Institute of Fundamental Technological Research, Polish Academy of Sciences. He is currently working as an Academic Teacher at Institute of Computer Science, Kazimierz Wielki University. His areas of interest include robotics, cryptography, computer networks and artificial intelligence.

https://orcid.org/0000-0001-5004-2928
E-mail: piotrk@ukw.edu.pl
Institute of Computer Science
Kazimierz Wielki University
Kopernika 1
85-074 Bydgoszcz, Poland