Fuzzy System for the Classification of Sounds of Birds Based on the Audio Descriptors

Krzysztof Tyburek, Piotr Prokopowicz, and Piotr Kotlarz

Institute of Mechanics and Applied Computer Science
Kazimierz Wielki University, Bydgoszcz, Poland
{krzysiekkt,piotrekp,piotrk}@ukw.edu.pl
http://www.imis.ukw.edu.pl

Abstract. This paper presents an application of fuzzy systems for the classification of sounds coded by the selected MPEG-7 descriptors. The model of the fuzzy classification system is based on the audio descriptors for a few chosen species of birds: Great Spotted Woodpecker, Greylag, Goldfinch, Chaffinch. The paper proposes two fuzzy models that definitely differ by the description of the input linguistic variables. The results show, that both approaches are effective. However, second one is more flexible in a case of future expanding of the model with next descriptors or species of birds.

Keywords: fuzzy system, fuzzy classification, MPEG-7, audio descriptors, fuzzy classification of audio signals.

1 Introduction

Birds rely on auditory processing for survival. Listening to others enables an bird to classify them as worst enemy, neighbor or stranger, a mate or non-mate, etc. Juvenile songbirds can listen to adult for develop a memory of a normal song that they will use to guide their own life. Our paper presents we would like to present a method for recognize kind of bird by feature of their sounds efficiently. In our study we used definition of MPEG-7 descriptors and fuzzy logic for classification of result. The potential applications for detecting and identifying bird species, particularly automatically, are diverse but can be grouped into the following categories.
1. Species identification.
2. Identification of individuals within a species.
3. Detection of the presence of bird.
4. Approaches to bioacoustic identification.
5. Feature extraction from time domain and frequency domain of bird song.

2 Sound Description With the MPEG-7

The solutions of searching of multimedia data basing on label technique do not always give expecting results. It means that sending queries are not always in

L. Rutkowski et al. (Eds.): ICAISC 2014, Part II, LNAI 8468, pp. 700–709, 2014.

accordance with demanding of person or computer system. Correctly interpretation of sound source is the main issue which occurs during recognition process of sound signals. In this paper researching of sound come from birds: Great Spot–ted Woodpecker, Greylag, Goldfinch, Chaffinch. Researching of sound of bird can be useful for high level of recognizably each other. This problem can be solved by means of MPEG-7 standard which gives a lot of descriptors describing physical features of sound. These descriptors are defined on the base of analysis of digital signals and index of most important their factors. The MPEG-7 Audio standard comprises descriptors and description schemes that can be divided [4–6] into two classes: generic low-level tools and application-specific tools. The generic tools, referred to in the standard as the audio description framework apply to any audio signal and include the scalable series, low-level descriptors (LLDs) and the unform silence segment. The application-specific tools restrict their application domain as a means to afford more descriptive power and include general sound recognition and indexing tools and description tools. The low-level audio descriptors have very general applicability in describing audio. There are seventeen temporal and spectral descriptors [6] that can be divided into six groups. A typical LLD may be instantiated either as a single value for a segment or a sampled series. Then two names for those descriptors are used, as the application requires: AudioLLDScalarType and AudioLLDVectorType, the first type is inherited for scalar values and describing a segment with a single summary, such as power or fundamental frequency, the second one is inherited for vector types describing a series of sampled valued, such spectra. This paper deals with LLDs as well as application-specific tools to recognize audio signal coming from a group of birds. In order to find a feature vector of the group of birds the analysis has been performed in the temporal as well as in frequency domains.

2.1 Time Domain Parameterization

For the purpose of right describing of waveform of sound it is necessary to define descriptor. The descriptor is represented as a fraction of time of separating phases to time of all phases. Log - time of the ending transient (TET) l_{tk}, which is given by:

$$l_{tk} = log(t_{pk} - t_{max}), \tag{1}$$

where:
t_{max} is the time at which the maximal amplitude has been reached,
t_{pk} is the time at which the level of 10 % of maximal value has been reached in the decay stage.

2.2 Frequency Domain Parameterization

Since the frequency domain may contain important information concerning features of the sound it is worthwhile to introduce its parameterization. The base of parameterization of sound spectrum are Fourier transform, wavelet analysis, cepstrum or Wigner–Ville'a transform. The following parameters describing frequency domain of signal were applied:

1. Brightness

$$Br = \frac{\sum_{i=0}^{n} A(i) \cdot i}{\sum_{i=0}^{n} A(i)}, \tag{2}$$

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

2. Irregularity of spectrum

$$Ir = log(20 \sum_{i=2}^{N-1} \mid log \frac{A(i)}{\sqrt[3]{A(i-1) \cdot A(i) \cdot A(i+1)}} \mid), \tag{3}$$

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

3 Preparation of Audio Data

The objects of researching was sounds of the birds like Great Spotted Woodpecker, Greylag, Goldfinch, Chaffinch. One of the purposes of the experiments was searching for vector of features which allow to automatic classification of each bird. For parameterization of frequency domain state window length was proposed [8]. It was applied for all samples in experiment. State window length is the fragment of signal (in time domain) which was taken in the same point of time. State window length contains constant amount of samples. The beginning of this window was taken when the level of 10 % of maximal value has been reached. The length of window is determined by resolution of spectrum, according to the formula:

$$f_r = \frac{f_s}{n}, \tag{4}$$

where:
f_r is the spectrum resolution
f_s – sampling frequency (44100 Hz)
n - number of samples.
In the paper f_r equal to $4Hz$ was assumed. It means that number of samples which are assigned to experiment is equal 11025. If testing sound is shorter then

length of window ($n = 11025$) then absent values should be supplemented with zeros to $n = 11025$ [7, 8]. Selecting fragment of signals in time domain were treated DFT and this spectrum was analyzed.

4 Modeling of Fuzzy System

For the realization of classification a fuzzy system was used. The model of this system is based on audio descriptors for the chosen species of birds: Great Spotted Woodpecker, Greylag, Goldfinch, Chaffinch. Each descriptor is represented by a separate linguistic variable. Since the data shall be classified into four categories, we also accept four output variables. Each of them corresponds to a different species.

4.1 Basic Assumptions

All data used in defining the model were normalized to the interval $[0, 1]$. In research two different fuzzy models were proposed. The main difference between them lies in the way of definition of the input linguistic variables. However, common for both are the output variables, defined on the interval $[0, 1]$, where 0 means the lack of recognition of the given species, whereas 1 means the full identification. The output variable out_1 responsible for the recognition of woodpecker, is shown on the Fig.1. Other output variables for classifying the rest of the birds are described in the same way. In addition, both proposals assume that the system will use only four rules. Each rule will select one of the bird species. In this way we obtain the four answer from the fuzzy system, which belong to the numerical interval $[0, 1]$. Thanks to that, the final result of classification

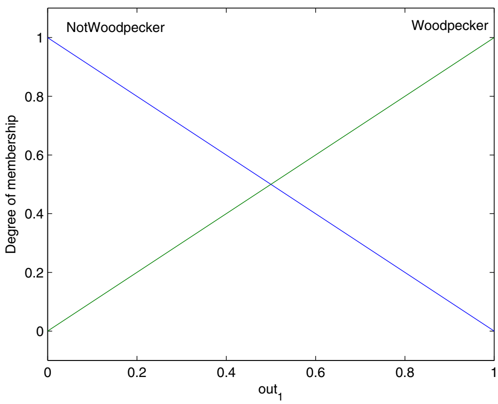

Fig. 1. Linguistic variable out_1

Table 1. Summary of the data used for modeling the fuzzy systems

	Brightness			TET			IR		
	min	max	avg.	min	max	avg.	min	max	avg.
Woodpecker	0	0,21	0,13	0	0,1	0,01	0,48	0,74	0,64
Greylag	0,26	0,43	0,34	0,68	1	0,82	0	0,62	0,37
Goldfinch	0,46	1	0,62	0,03	0,79	0,31	0,3	1	0,66
Chaffinch	0,31	0,55	0,44	0,17	0,7	0,37	0,2	0,73	0,47

can be easily and clear-cut determined with use the *winner takes all* principle. This means that the largest output value indicates an identified species of bird assigned to the given audio signal. Other parameters common for the both propositions:
– method of fuzzyfication - singleton,
– method of aggregation for premise parts of rules - min,
– operator of implication - min,
– defuzzyfication - middle of maximum.

4.2 Fuzzy System - Proposition 1

The input linguistic variables are divided into two values: Small, Big (see Fig.2). To determine the rules mean values for the input data assigned to the species of birds were calculated. On their base fuzzy sets for the premise part of rules were determined. For the average $\leq 0,5$ the *Small* set was taken, for $> 0,5$ fuzzy set *Big*. Tab.1 presents a summary of the minimum, average, and maximum for each of descriptors taking into account the species of birds.

As a result we receive the following rules:

- IF $Brgt$ is *Small* AND TET is *Small* AND IR is *Big* THEN out_1 is Wdp AND out_2 is $NotGrlg$ AND out_3 is $NotGldfch$ AND out_4 is $NotChfnch$
- IF $Brgt$ is *Small* AND TET is *Big* AND IR is *Small* THEN out_1 is $NotWdp$ AND out_2 is $Grlg$ AND out_3 is $NotGldfch$ AND out_4 is $NotChfnch$
- IF $Brgt$ is *Big* AND TET is *Small* AND IR is *Big* THEN out_1 is $NotWdp$ AND out_2 is $NotGrlg$ AND out_3 is $Gldfch$ AND out_4 is $NotChfnch$
- IF $Brgt$ is *Small* AND TET is *Small* AND IR is *Small* THEN out_1 is $NotWdp$ AND out_2 is $NotGrlg$ AND out_3 is $NotGldfch$ AND out_4 is $Chfnch$

where TET means linguistic variable *Time of the Ending Transient*, $Brgt$ - *Brightness* , IR - *Irregularity of Spectrum* , Wdp - fuzzy set $Woodpecker$, $Grlg$ - *Greylag*, $Gldfch$ - $Goldfinch$ and $Chfnch$ - $Chaffinch$.

The results of classification are presented in the Tab.2. The 40 audio signals was used, ten for every species of birds. First tenth lines represents woodpecker's signals, next ten - greylag, and next for goldfinch, and last - chaffinch. As can be noted, the classification is quite effective. However, the results clearly diverging from the average values are classified incorrectly - see lines numbers

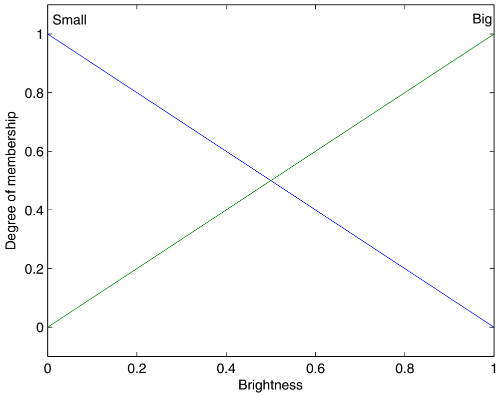

Fig. 2. Input linguistic variable *Brightness*

2,21,22,24,28,32,35,39 . It's a little alarming because the system is based on the fuzzy model, which takes account all available data, including those diverging from the average values.

4.3 Fuzzy System - Proposition 2

To improve the quality of the classification presented in the first proposition, the greater granulation of input variables can be a way. In addition to the terms $Small$, Big, could be introduced another fuzzy sets such as: $VerySmall$, $Average$, etc. Note, however, that when increasing the granularity, you need to check the uniqueness of the premise parts of the rules. If in the future we will expand the capabilities of our classification system for another species of birds, there is a rather complex process of updating and matching of linguistic variables and already defined rules.

An alternative solution is split the input variables for the fuzzy sets characteristic for each species. Similar solution to another problem - the classification of flowers (irises) - was presented in [3]. In this way, instead of fuzzy sets defining the size $Small$, Big, etc., we introduce sets $Woodpecker$, $Greylag$, $Goldfinch$, $Chaffinch$. Each of them is a triangular fuzzy set (see LR fuzzy sets notation in [1]) and is determined on the the available data. For example lets look at set $Woodpecker$ (Fig.3):

$$Woodpecker = \Lambda(x; x_{mean} - 2 \cdot \Delta_L, x_{mean}, x_{mean} + 2\Delta_R), \quad (5)$$

where $\Delta_L = x_{mean} - x_{min}$, $\Delta_R = x_{max} - x_{mean}$,
$x_{min}/x_{max}/x_{mean}$ – the minimum/maximum/mean value of the descriptor for the given species.

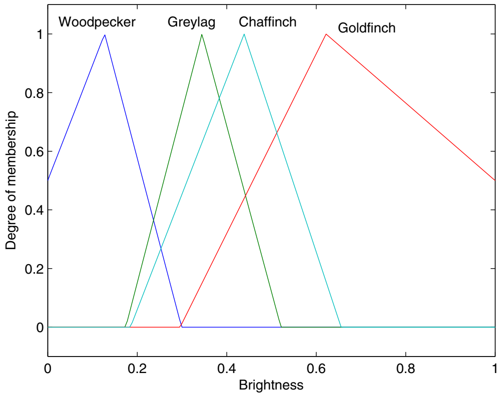

Fig. 3. New fuzzy sets for input linguistic variable *Brightness*

With such definitions of the fuzzy sets we are sure that minimum and maximum have at least $0,5$ membership level in the set assigned to a given species. Same way we deal with the all other input variables (see Fig.3). It is characteristic that we do not take into account properties expected from the fuzzy model such as the completeness or continuity (see [2]). We also do not expect that the values of the fuzzy membership functions sum to unity within a linguistic variable. The rule base looks as follows:

- IF $Brgt$ is Wdp AND TET is Wdp AND IR is Wdp THEN out_1 is Wdp AND out_2 is $NotGrlg$ AND out_3 is $NotGldfch$ AND out_4 is $NotChfnch$
- IF $Brgt$ is $Grlg$ AND TET is $Grlg$ AND IR is $Grlg$ THEN out_1 is $NotWdp$ AND out_2 is $Grlg$ AND out_3 is $NotGldfch$ AND out_4 is $NotChfnch$
- IF $Brgt$ is $Gldfch$ AND TET is $Gldfch$ AND IR is $Gldfch$ THEN out_1 is $NotWdp$ AND out_2 is $NotGrlg$ AND out_3 is $Gldfch$ AND out_4 is $NotChfnch$
- IF $Brgt$ is $Chfnch$ AND TET is $Chfnch$ AND IR is $Chfnch$ THEN out_1 is $NotWdp$ AND out_2 is $NotGrlg$ AND out_3 is $NotGldfch$ AND out_4 is $Chfnch$

We can see that the classification results (Tab.3) do not contain faulty detections. It is primarily the result of increased granulation of input variables comparing to the previous proposal. If we applied a similar granulation there, we would also get a similar effectiveness. However, the second proposition has one basic advantage over the previous one. It is related above all, to the simplicity of expansion defined classifier. Adding the next species of birds does not violate existing structure. You simply add the next fuzzy sets characteristic for the new species. Well, of course you should add a rule that recognizes a new class of data. Similarly, if we want to introduce another descriptor to the model, the changes will be much simpler, more intuitive and easier than with the previous proposition.

Table 2. Results of classification (proposition 1)

	Woodpecker	Greylag	Goldfinch	Chaffinch
1	0,84	0,16	0,16	0,16
2	0,24	0,24	0,24	0,76
3	0,86	0,14	0,14	0,14
4	0,88	0,13	0,13	0,13
5	0,86	0,14	0,14	0,14
6	0,82	0,19	0,19	0,19
7	0,82	0,19	0,19	0,19
8	0,82	0,19	0,19	0,19
9	0,79	0,21	0,21	0,21
10	0,81	0,19	0,19	0,19
11	0,26	0,75	0,26	0,26
12	0,25	0,75	0,25	0,25
13	0,31	0,7	0,31	0,31
14	0,18	0,82	0,18	0,18
15	0,22	0,79	0,22	0,22
16	0,2	0,81	0,2	0,2
17	0,2	0,8	0,2	0,2
18	0,25	0,76	0,25	0,25
19	0,21	0,79	0,21	0,21
20	0,25	0,75	0,25	0,25
21	0,28	0,28	0,28	0,73
22	0,5	0,5	0,5	0,5
23	0,15	0,15	0,85	0,15
24	0,27	0,27	0,27	0,73
25	0,16	0,16	0,84	0,16
26	0,26	0,26	0,74	0,26
27	0,19	0,19	0,81	0,19
28	0,5	0,4	0,5	0,4
29	0,3	0,3	0,71	0,3
30	0,2	0,2	0,81	0,2
31	0,24	0,24	0,24	0,76
32	0,32	0,68	0,32	0,32
33	0,23	0,23	0,23	0,78
34	0,23	0,23	0,23	0,78
35	0,79	0,22	0,22	0,22
36	0,23	0,23	0,23	0,78
37	0,24	0,24	0,24	0,77
38	0,24	0,24	0,24	0,76
39	0,77	0,24	0,24	0,24
40	0,27	0,27	0,27	0,73

Table 3. Results of classification with modified fuzzy system (proposition 2)

	Woodpecker	Greylag	Goldfinch	Chaffinch
1	0,75	0,25	0,25	0,25
2	0,75	0,25	0,25	0,25
3	0,75	0,25	0,25	0,25
4	0,75	0,25	0,25	0,25
5	0,75	0,25	0,25	0,25
6	0,75	0,25	0,25	0,25
7	0,75	0,25	0,25	0,25
8	0,75	0,25	0,25	0,25
9	0,75	0,25	0,25	0,25
10	0,75	0,25	0,25	0,25
11	0,18	0,82	0,18	0,18
12	0,14	0,87	0,14	0,14
13	0,25	0,75	0,25	0,25
14	0,08	0,92	0,08	0,08
15	0,25	0,75	0,25	0,25
16	0,02	0,98	0,02	0,02
17	0,25	0,76	0,25	0,25
18	0,13	0,88	0,13	0,13
19	0,22	0,79	0,22	0,22
20	0,13	0,87	0,13	0,13
21	0,25	0,25	0,75	0,25
22	0,14	0,14	0,86	0,14
23	0,25	0,25	0,75	0,25
24	0,15	0,15	0,86	0,15
25	0,19	0,19	0,81	0,19
26	0,25	0,25	0,76	0,25
27	0,04	0,04	0,97	0,04
28	0,25	0,25	0,76	0,25
29	0,14	0,14	0,86	0,14
30	0,25	0,25	0,76	0,25
31	0,2	0,2	0,2	0,81
32	0,25	0,25	0,25	0,76
33	0,22	0,22	0,22	0,78
34	0,25	0,25	0,25	0,76
35	0,25	0,25	0,25	0,75
36	0,04	0,04	0,04	0,97
37	0,08	0,08	0,08	0,92
38	0,1	0,1	0,1	0,91
39	0,1	0,1	0,1	0,91
40	0,25	0,25	0,25	0,75

5 Summary

Worth to emphasize again that the research presented here are a just preliminary step to further work on the idea of using fuzzy sets in the analysis of audio signals. It should also be noted that the set of data which is the source of the proposed models is too small to build an explicit and definitive conclusions. Nevertheless, the results presented here, clearly shows that the application of fuzzy systems as the classifiers for audio data described by MPEG-7 descriptors is the direction worth of further attention.

In the future, the authors also plan to work together on the search for new effective audio signal descriptors, where new model of fuzzy numbers the Ordered Fuzzy Numbers [9, 10] (developed by the second author) will be used. Good computational properties of this model [11], which support applications, deserve the particular attention.

References

1. Dubois, D., Prade, H.M.: Fuzzy sets and systems: Theory and applications. Academic Press, New York (1980)
2. Driankov, D., Hellendoorn, H., Reinfrank, M.: An Introduction to fuzzy control. Springer, Heidelberg (1996)
3. Siler, W., Buckley, J.J.: Fuzzy Expert Systems and Fuzzy Reasoning. Wiley (2005)
4. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7, Multimedia Content Description Interface. John Wiley & Sons, Chichester (2002)
5. Martnez, J.M.: MPEG-7 Overview, Klangenfurt (July 2002)
6. Lindsay, A.T., Burnett, I., Quackenbush, S., Jackson, M.: Fundamentals of audio descriptions. In: Manjunath, B.S., Salembier, P., Sikora, T. (eds.) Introduction to MPEG-7: Multimedia Content Description Interface, pp. 283–298. John Wiley and Sons, Ltd. (April 2002)
7. Tyburek, K.: Classification of string instruments in multimedia database especially for pizzicato articulation, Ph. D. thesis. Institute of Fundamental Technological Research Polish Academy of Sciences, Warsaw (November 2006) (in Polish)
8. Tyburek, K., Cudny, W., Kosiski, W.: Pizzicato sound analysis of selected instruments In the freguency domain. Image Processing & Communications 11(1), 53–57 (2006)
9. Kosiński, W., Prokopowicz, P., Ślezak, D.: Ordered fuzzy number, Bulletin of the Polish Academy of Sciences. Ser. Sci. Math. 51(3), 327–338 (2003)
10. Kosiński, W., Prokopowicz, P., Kacprzak, D.: Fuzziness – representation of dynamic changes by ordered fuzzy numbers. In: Seising, R. (ed.) Views of Fuzzy Sets and Systems from Different Perspectives. STUDFUZZ, vol. 243, pp. 485–508. Springer, Heidelberg (2009)
11. Prokopowicz, P.: Flexible and Simple Methods of Calculations on Fuzzy Numbers with the Ordered Fuzzy Numbers Model. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2013, Part I. LNCS, vol. 7894, pp. 365–375. Springer, Heidelberg (2013)