DISCRIMINATE ANIMAL SPECIES USING CEPSTRAL COEFFICIENTS AND TESPAR ANALYSIS

Petre G. POP
Comm. Dept., Technical University of Cluj-Napoca, petre.pop@com.utcluj.ro

Abstract: Identification of animal species based on the sounds emitted has already proven to be useful in biodiversity assessment but can also be useful in other biological research. Most researchers have used the cepstral coefficients for animal species discrimination. In this paper we explore the use of a combined cepstral - TESPAR analysis, which does not use directly the acoustic signal from animals, but cepstral coefficients derived from it (MFCC and Teager-MFCC) subject to TESPAR analysis, to discriminate between different animal species. Our experiments using this approach together with classification techniques shows that TESPAR S-matrices of cepstral coefficients can be successfully used to discriminate between different animal species, even in the conditions of small sets of training data with small length and different sampling frequencies.

Keywords: *animal species identification, cepstral coefficients, TESPAR analysis, Teager energy operator, Random Forest.*

I. INTRODUCTION

Identification of animals by their sounds is important for biological research and biodiversity assessment, especially in detecting and locating animals [1]. Many animals generate sounds either for communication or to accompany their living activities such as mating, eating, moving, flying, etc. Challenges related to processing animal sounds come from noisy data, imperfect data labeling, poor knowledge about how animals produce and perceive sound and state of the animal. Incorporation of noise models is important since the recordings are altered with many interfering noise sources and noise can significantly decrease classification accuracy, especially when the noise characteristics vary across the dataset. Detecting state of an animal is also a challenging task since humans can only guess as to what an animal is trying to communicate by sounds [2].

One of the important tasks when analyzing animal sounds is to measure the acoustically relevant features. Majority of bioacoustics signals processing systems use time-frequency techniques such as the Fourier analysis, wavelets and energy distributions [3]. Most authors use cepstral coefficients (especially Mel frequency cepstral coefficients - MFCC) as acoustic characteristics for training and then testing under different classification systems.

Previously we have approached the possibility of using the TESPAR S-matrix applied to individual cepstral coefficients to visual differentiate sounds from different animal species [4]. In this paper we investigate the use of MFCC and Teager-MFCC coefficients with TESPAR analysis to differentiate the sounds of different animal species. The choice is justified by the fact that some of the MFCC coefficients contain specific information about the emitter of sound, while other contain information about the message. The Teager (or Teager-Kaiser) operator contains information of both amplitude and frequency, so it can give us a good insight into the occurrence of events in the initial acoustic signal (what could be used to discriminate between species). On the other hand, TESPAR analysis allows obtaining a specific finger-print, of fixed length, for the sound emitter.

To test our idea we implemented a dedicated application for cepstral and TESPAR analysis (to generate final features) and a classification system using the Random Forest algorithm, which is one of the easiest machine learning tool used in the industry.

II. CEPSTRAL COEFICCIENTS

Spectral features are frequently used to process acoustic signals. One of the most used spectral feature is represented by the MFCC coefficients (Mel Frequency Cepstral Coefficients). The first MFCC coefficient (C_0) give information about the shape of the log spectrum. The next one (C_1) measures the balance between the upper and lower zones of the spectrum while the rest of coefficients are concerned with finer features in the spectrum [5]. Also, some of the MFCC coefficients contain specific information about the emitter of sound, while other MFCC coefficients contain information about the message contained in the acoustic signal.

Another type of MFCC coefficients are obtained if, first, apply the Teager (or Teager-Kaiser) energy operator (TEO) on initial acoustic signal and then repeat the processing necessary to obtain the MFCC coefficients. In this way, we obtain so-called Teager-MFCC coefficients (T-MFCC) [6]. The TEO, is capable under appropriate signal constraints of accurately tracking the instantaneous amplitude and instantaneous frequency of the signal [7] [8]. For discrete signals, the Teager operator is defined as:

$$\Psi[x_n] = x_n^2 - x_{n-1}x_{n+1} \qquad (1)$$

A variety of applications have been developed using the TEO, most of them concerning speech analysis.

Manuscript received September 8, 2017; revised September 20, 2017

III. TESPAR ANALYSIS AND TESPAR MATRICES

TESPAR (**T**ime **E**ncoded **S**ignal **P**rocessing **A**nd **R**ecognition) algorithm is intended to classify time domain signals using some specific signal shape parameters. This algorithm is based on the position estimations of the signal real and complex zeroes. The first category of zeros (real) correspond to the zero crossings of the initial signal while the complex zeroes are associated with local extremes (minima or maxima), points of inflexion etc. The real zeroes of a time domain signal and some complex zeroes can be simply estimated by visual inspection of time domain signal waveform (Fig. 1), however the detection of all complex zeroes is difficult task.

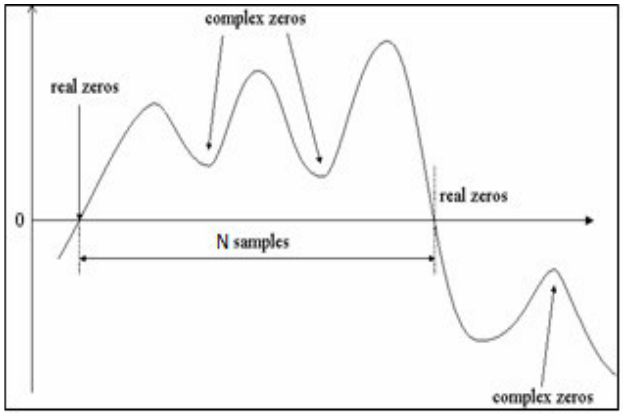

Figure 1. TESPAR analysis for a time domain signal.

This impediment can be overcome in this way: the time domain signal waveform is segmented between successive zero crossings (real zeroes), thereby generating a number of epochs with a certain durations (lengths); then, this duration information is combined with simple approximations of the shape (number of minima or maxima) between two successive real zeroes (zero crossings). In this way an important number of complex zeroes may be identified by analyzing the shape of the time domain signal waveform between its successive real zeroes.

In the first approach of the TESPAR method [9], two descriptors (attributes) are associated with each epoch of the time domain signal waveform:

- the duration (*D*) between two successive real zeroes expressed as the number of samples between two zero crossings;
- the shape (*S*) between two successive real zeroes expressed as number of minima (or maxima) of that epoch.

In this way, a signal in the time domain can be described as a succession of value pairs for these descriptors. Since many identical epochs with same duration and number of minima are likely to occur, the idea is to use an alphabet that associates a symbol for each value pair (D, S) so that the initial signal can be described as a sequence of symbols in that alphabet.

Most signals can be described by a series of discrete numerical descriptors based on TESPAR symbol alphabets. The TESPAR standard alphabet can be used to convert the sequence of epochs into an equivalent TESPAR symbol stream, mapping the duration/shape (D/S) attributes combination of each epoch to a single symbol by a coding process [9] (Fig. 2). For different classes of signals, we can have different alphabets, both in structure and number of symbols.

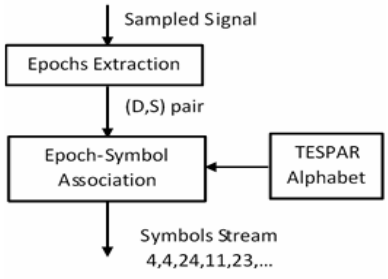

Figure 2. TESPAR coding process.

Another approach use an additionally signal descriptor *A*, for each epoch, such as epoch maximum amplitude, epoch energy, etc. In this case, the coding is based on a comparison of two successive epochs (for each descriptor D, S, A), as shown in Fig. 3. This is TESPAR DZ alphabet with fixed number of symbols [10] [11].

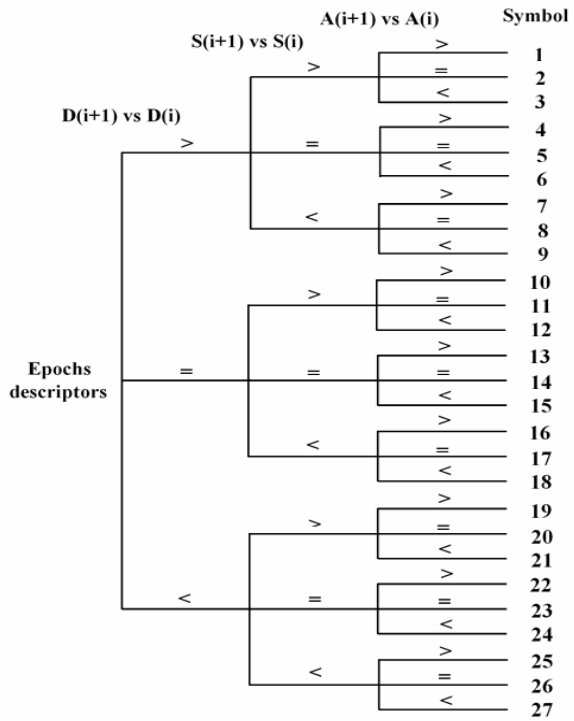

Figure 3. TESPAR DZ alphabet.

In this case, we get a fixed number of symbols (27) regardless of the duration or complexity of an epoch and the succession of epochs (duration, form, auxiliary descriptor value) is very important during the evolution of the signal over time.

Regardless of the alphabet used, the resulting TESPAR symbols string may be converted into fixed-dimension matrices [9]. The S-matrix is the histogram of TESPAR symbols. Given a symbol stream *s(i)* of length *M* (resulting

from the coding process using a specific alphabet with symbols *1,…,N*), the elements of *S matrix* can be expressed as:

$$S_n = \frac{1}{M}\sum_{i=0}^{M-1} T_i, \quad 1 \le n \le N \tag{2}$$

Where:

$$T_i = \begin{cases} 1 \; if \; s(i) = n \\ 0, \text{otherwise} \end{cases} \tag{3}$$

The contribution of each symbol to the S-matrix values can be weighted by the value of the associated *A* descriptor.

Another type of matrix is the A-matrix, a two dimensional *N x N* matrix that contains the number of times each pair of symbols appears, with a possible lag *L* (*L* ≥*1*) of symbols between the pair elements:

$$A_{mn} = \frac{1}{M}\sum_{i=L}^{M-1} T_i, \quad 1 \le m, n \le N \tag{4}$$

Where:

$$T_i = \begin{cases} 1 \; if \; s(i) = m \;\; and \;\; s(i-L) = n \\ 0, \text{otherwise} \end{cases} \tag{5}$$

Again, the contribution of each pair of symbols can be weighted with a value that implies values of the *A* descriptor from each epoch associated with that symbols.

IV. EXPERIMENTS, RESULTS AND DISCUSSIONS

To extract the attributes required for classification we realized a dedicated application that allows:

- selecting and open a file with an acoustic signal (*wav* format);
- determine and eliminate silence zones;
- pre-processing the acoustic signal (preemphazis, frame blocking, windowing);
- selecting a type of features (LPC, LPCC, MFCC, T-MFCC) and computing a specified number of coefficients for each frame;
- TESPAR analysis (zero crossings, epochs, minima number for each epoch) of a selected feature values, with the possibility to visualize the epochs;
- generating symbol stream (coding) specifying the alphabet used (standard or *DZ*) and the type of additional epoch descriptor (maximum amplitude, average energy, threshold value) for each feature type;
- generate and visualize TESPAR matrices, *S* and *A* (based on a user specified value for lag parameter *L*), with the possibility of applying a weight to each symbol or pair of symbols;
- saving the TESPAR matrices generated in the *arff* format (Attribute-Relation File Format) specific to the Weka application [12].

In our experiments we used sounds from different animal species (domestic and wild: bear, cat, cow, elephant, lion and sheep), taken from free sources ([15] [16] [17]) and with great diversity in the quality of recordings and the acoustic parameters used. It has to be said that it is a challenge to find enough of such free recordings on the internet (for example, it is very likely to find the same sounds under different names or with different audio parameters) so you can use this data in a training-prediction process using classification algorithms. On the other hand, it is to be expected that for some wild animals there will be few quality records and long enough for a better training.

In our experiments we used the following parameters:

- window length: 10ms, 15ms;
- features: MFCC, T-MFCC;
- number of cepstral coefficients (N): 8, 12, 16, 20;
- TESPAR-DZ alphabet, using epoch energy as additional *(A)* descriptor with a threshold value of 1%.

The developed application allows the selection of these parameters and the generation of an *arff* file (containing TESPAR S matrix values for each feature coefficient along with the associated class) for a set of audio files containing animal sounds. Finally, we used a total of 25 audio recordings for each animal, of which 20 were used to train the classification algorithm (a total of 120 recordings for training) and the remaining 5 records were used for predictions (a total of 30 recordings for predictions).

Using Weka we tested some classification algorithms and finally we stopped at the Random Forest (RF) algorithm [13]. A Random Forest consists of a collection of simple decision trees, each capable of producing a response when presented with a set of predictor values. For classification problems, this response takes the form of a class membership, which associates a set of independent predictor values with one of the categories present in the dependent variable (in our case, the animal species). The RF algorithm was developed by Breiman [14].

Using tree ensembles (collections) can lead to significant improvement in prediction accuracy.

We used the following values for the Random Forest model parameters:

- number of attributes to randomly investigate (K), with values between 12 and 30, depending on the total number of instances (using as the starting point the value given by the product number_TESPAR_symbols * number_feature_cofficients);
- number of iterations or the number of trees (I), with values between 200 and 500, with a step of 20.

Evaluation of predictive model in the training phase was done with the 10 folds cross-validation. For predictions, the confusion matrix and the *TP Rate* accuracy parameter were retained.

The Weka application allows you to save the results in text files, files that were then processed with another dedicated application to extract the previously specified parameter values and save them to *csv* files. These values were then used in the Excel application to generate pivot tables and associated charts. A pivot table allows a synthetic representation of a data set, with the possibility to easily change the perspective of the data (e.g. parameters with selectable values from a list), with passes from aggregation levels to detail levels (drill down) and vice versa (drill up).

In our analysis:

- we used T (length of analysis window in ms) and N (feature coefficients number) as independent (selectable) parameters;
- on the horizontal (lines) we represented the number of iterations (I parameter) in the RF algorithm;

- on vertically (columns) we represented the parameter K (number of attributes) in the classification algorithm;
- at the lowest level we have the number of correctly classified cases (maximum 5) for each animal species.

In the case of T-MFCC feature, the best global (for whole prediction set) results (Sum (TPRate) = 27 meaning a 90% percentage) were obtained for the following parameter values (Fig.4, Fig.5):
- acoustic parameters T = 10ms, N = 8;
- parameters of the Random Forest algorithm: K = 15, I = 320, 400, 420.

In the case of MFCC, the global results are weaker and the best ones (TPRate = 80%) were obtained for the following values (Fig.7, Fig.8):
- acoustic parameters T = 15ms, N = 12;
- parameters of the Random Forest algorithm: K = 15, I = 300, 320.

Some comments on the results:
- the use of TESPAR analysis allows us to obtain a fixed length matrix S (in our case 27) regardless of the length of the analyzed sound signal; of course, a longer acoustic signal allows to obtain a more representative S-matrix for the individual of that species;
- we used the TESPAR-DZ alphabet because it is invariable at the sampling frequency used to record the sounds (alphabet symbols are determined based on the ratio between successive epochs); in our case this was necessary due to the problems encountered in obtaining records for different animal species;
- the T-MFCC feature gives better results than MFCC using shorter analysis windows and a lower number of coefficients; this can be explained by TEO ability to obtain representative time-frequency information for each animal species even on the basis of a small number of training records; we have the belief that for enough training data, the results would be comparable;
- the pivot tables for the two types of features (Fig. 4, Fig. 7) show that in the case of T-MFCC, the overall prediction results range are more widely than in the case of MFCC; but between species there are much greater differences in MFCC versus T-MFCC.

Figures 6 and 9 show the classification results for each species (having the same entities represented on the axes as in Fig. 5 and Fig. 8) for the two types of characteristics used (T-MFCC / MFCC). It is noted that:
- for some species (cat, lion) the results are almost constant for the variation of classification parameters;
- in some cases (T-MFCC/MFCC cow, MFCC sheep), for several values of parameter *K*, the results are maintained for a wide range of parameter *I* values;
- in other cases (bear, elephant) the results vary strongly with respect to the values of the classification parameters.

In our training-prediction experiments we used an instance (a set of attribute values together with associated class) for each animal sound recording, which means a total number of attributes equal to the product of the number of symbols in the TESPAR alphabet (in our case 27) and the number of T-MFCC / MFCC coefficients.

T[ms]	10
nCoef	8

Sum of TPRate	Column Labels															
Row Labels	200	220	240	260	280	300	320	340	360	380	400	420	440	460	480	500
12	**18**	**18**	**18**	**18**	**19**	**20**	**20**	**22**	**22**	**22**	**23**	**23**	**23**	**23**	**23**	**23**
13	**22**	**23**	**25**	**25**	**25**	**25**	**24**	**25**	**24**	**25**	**24**	**25**	**25**	**25**	**25**	**25**
14	**23**	**23**	**22**	**24**	**25**	**24**	**23**	**22**	**22**	**23**	**22**	**23**	**21**	**23**	**24**	**23**
15	**24**	**25**	**25**	**24**	**25**	**26**	**27**	**26**	**25**	**26**	**27**	**27**	**25**	**24**	**23**	**23**
bear	4	4	4	3	4	4	4	4	3	3	4	4	3	3	3	3
cat	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
cow	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4
elephant	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
lion	3	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4
sheep	3	3	3	4	3	4	5	4	4	5	5	5	4	4	3	3
16	**23**	**23**	**22**	**23**	**26**	**26**	**24**	**24**	**24**	**24**	**24**	**24**	**24**	**24**	**24**	**24**
17	**23**	**23**	**23**	**24**	**24**	**24**	**25**	**26**	**26**	**24**	**23**	**23**	**23**	**24**	**24**	**26**
18	**20**	**21**	**22**	**24**	**23**	**23**	**24**	**24**	**24**	**25**	**26**	**25**	**25**	**25**	**25**	**25**

Figure 4. Pivot table with confusion matrix values for T-MFCC feature

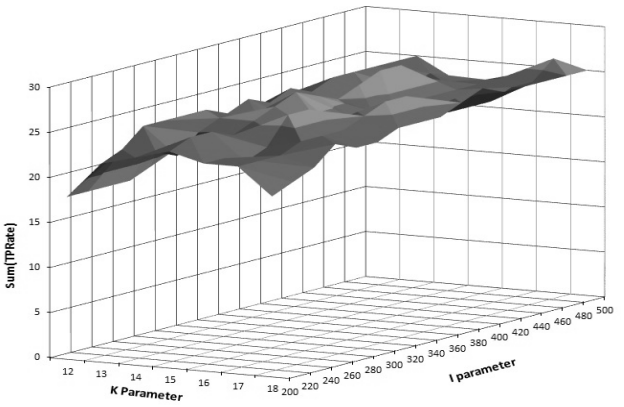

Figure 5. Pivot chart with confusion matrix values for T-MFCC feature, all species (T=10, N=8).

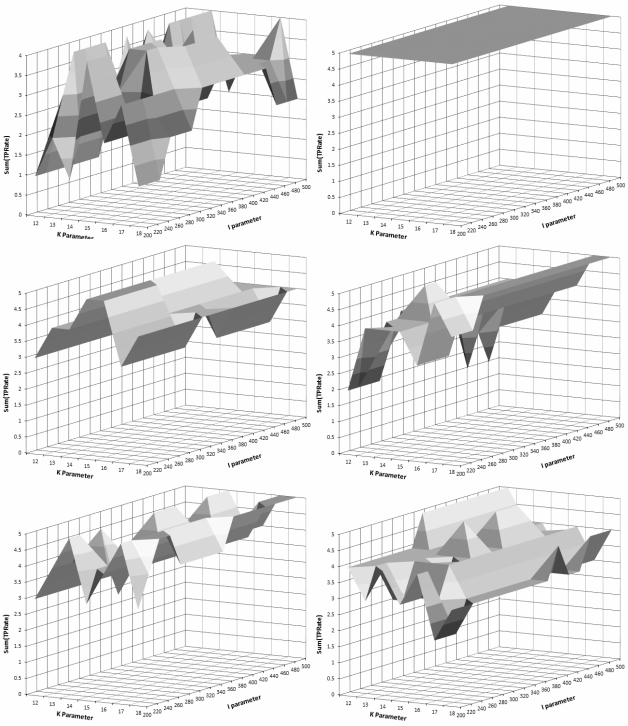

Figure 6. Confusion matrix values for T-MFCC feature, for each species; from top-left to bottom-right: bear, cat, cow, elephant, lion, and sheep (T=10, N=8).

T[ms]	15
nCoef	12

Sum of TPRate	Column Labels															
Row Labels	200	220	240	260	280	300	320	340	360	380	400	420	440	460	480	500
14	21	20	19	19	20	20	20	21	21	20	20	20	20	20	20	20
15	21	21	23	23	23	24	24	23	23	23	23	23	23	23	22	22
bear	3	2	3	3	3	4	4	3	3	3	3	3	3	3	3	3
cat	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
cow	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
elephant	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
lion	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
sheep	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4
16	22	21	22	21	21	21	22	23	23	21	22	23	23	23	23	23
17	20	21	21	22	22	22	21	21	21	21	21	21	20	21	21	21
18	20	22	21	23	21	21	21	22	22	22	22	20	21	21	21	21
19	22	23	23	23	23	23	24	24	23	23	23	23	23	23	23	22
20	21	21	20	21	20	19	21	21	21	21	21	22	21	21	20	21
21	20	19	20	20	20	20	20	20	20	20	21	21	21	20	20	21
22	20	22	20	20	21	19	18	19	19	19	20	21	20	20	20	21

Figure 7. Pivot table with confusion matrix values for MFCC feature.

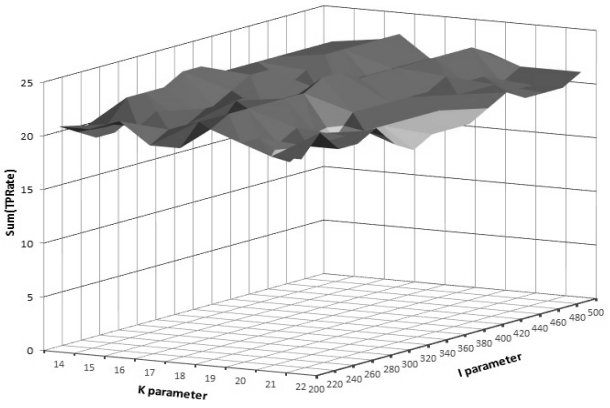

Figure 8. Pivot chart with confusion matrix values for MFCC feature (T=15, N=12).

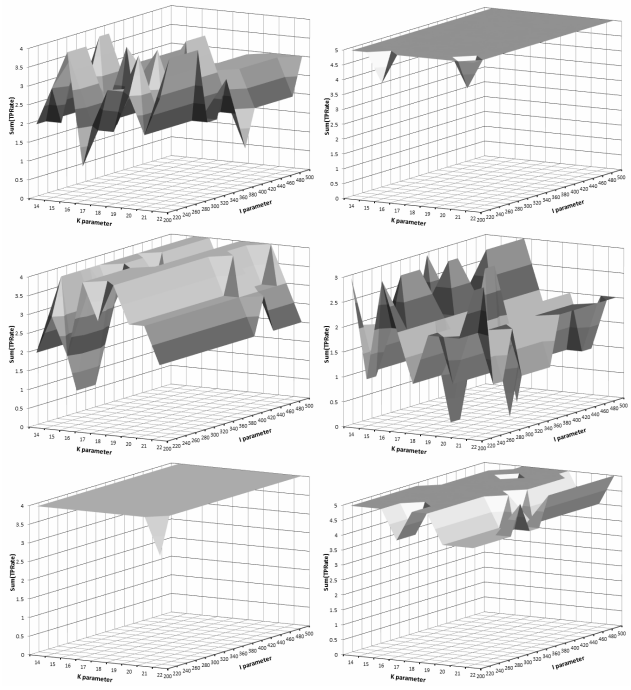

Figure 9. Confusion matrix values for MFCC feature, for each species; from top-left to bottom-right: bear, cat, cow, elephant, lion, and sheep (T=15, N=12).

A possible improvement would be to reduce the number of attributes (using machine-learning techniques), keeping as far as possible the classification (training-prediction) performances. This can lead to a significant reduction in the number of values for each feature coefficient to be used for the classification or even the exclusion of some coefficients from the classification process.

Another option would be the use of an instance for each feature coefficient. In this case, we may find it easier to determine what coefficients have significant weightings in the classification process and thus significantly reduce the number of attributes used in the classification if prediction performances are maintained.

V. CONCLUSIONS

In this work, we investigated the use of TESPAR S-matrices of cepstral coefficients (MFCC, T-MFCC) and Random Forest classification algorithm to discriminate between animal species based on their sounds. First results show that the TESPAR S-matrices could be successfully used for this purpose. T-MFCC coefficients allow for better results than MFCC, even in the conditions of small sets of training data with small length and different sampling frequencies. More experiments on a larger database with animal sounds are needed.

REFERENCES

[1] Chang-Hsing Lee, Yeuan-Kuen Lee, Ren-Zhuang Huang (2006), "Automatic Recognition of Bird Songs Using Cepstral Coefficients", *Journal of Information Technology and Applications,* Vol. 1 No. 1, pp.17-23, May, 2006.
[2] Deshmukh O., Rajput N., Singh Y., Lathwal S., "Vocalization patterns of dairy animals to detect animal state", *21st Int. Conf. Pattern Recognition*, Tsukuba, Japan, Nov. 11-15, 2012.
[3] E. D. Chesmore, "Automated bioacoustic identification of species", *An. Acad. Bras. Ciênc*. 76(2), pp. 435 – 440, 2004.
[4] Pop, P.G., "Discriminate Animal Sounds Using TESPAR Analysis", *Int. Conf. on Advancements of Medicine and Health Care through Technology*, Romania, 59, pp. 185-188, Oct. 2016.
[5] Huang, X., Acero, A., Hon, H., *Spoken Language Processing*, Prentice Hall, 2001.
[6] H.A. Patil, T. K. Basu, "Identifying perceptually similar languages using teager energy based cepstrum", *Engineering Letters*, 16(1), pp.151–159, 2008.
[7] Kaiser JF, "On a simple algorithm to calculate the 'energy' of a signal", *IEEE International Conference Acoustic Speech Signal Process.,* Albuquerque, pp 381–384, 1990.
[8] Kaiser JF, "Some useful properties of Teager's energy operators", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp 149–152, 1993.
[9] R.A. King, T.C. Phipps, "Shannon, TESPAR And Approximation Strategies", *ICSPAT 98*, Vol. 2, pp. 1204-1212, Toronto, Canada, Sept. 1998.
[10] Lupu, E., Emerich, S., Beaufort, F. „On-line signature recognition using a global features fusion approach", *Acta Technica Napocensis, Electronics and Telecommunications*, vol.50, pp. 13-20, 2009.
[11] Emerich, S.; Lupu, E., Arsinte, R. "A New Approach to Iris Recognition", *10th International Symposium on Signals, Circuits and Systems (ISSCS)*, Iasi, Romania, 2011.
[12] http://www.cs.waikato.ac.nz/ml/weka/
[13] Witten IH, Frank E, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.
[14] Breiman L, "Random Forests", *Machine Learning*, 45 (1): 5–32, 2001.
[15] http://www.grsites.com/
[16] http://soundbible.com/tags-animal.html
[17] http://www.findsounds.com/types.html