Volume 58, Number 4, 2017
ACTA TECHNICA NAPOCENSIS
Electronics and Telecommunications

LINEAR PREDICTIVE CEPSTRAL COEFFICIENTS IN WILDLIFE DETECTION SYSTEMS

Elena Roxana BUHUS, Lacrimioara GRAMA, Corneliu RUSU
Signal Processing Group, Basis of Electronics Department, Faculty of Electronics, Telecommunications and Information Technology, Technical University of Cluj-Napoca
26 George Baritiu Street, Cluj-Napoca, Romania,
Email: {Roxana.Buhus, Lacrimioara.Grama, Corneliu.Rusu}@bel.utcluj.ro

Abstract: The paper presents the impact of cepstrum computation in Linear Predictive Coding for audio signals that belong to wildlife areas. In this sense, we compare on diverse number of features both Linear Predictive Coding coefficients and Linear Predictive Cepstral coefficients for the audio signal characterization. The scope is to identify the appropriate number of coefficients for both methods, together with the best method that can provide a certain accuracy within a defined framework, such as an intruder detection system. The experimental results prove that the Linear Predictive Cepstral coefficients method provides a better accuracy for tested scenarios than Linear Predictive Coding. The highest accuracy is obtained using the order of 16 for the predictor filter together with the KStar classifier.

Keywords: *LPC, LPCC, KStar.*

I. INTRODUCTION

The number of events that imply illegal logging, hunting, or trespassing of natural reservations, parks, or forests increased so much in the past decade, that on a high demand became the design of wildlife surveillance systems. These systems are meant to detect in time unwanted activities within the protected areas and help the authorities to take an action. To create a reliable design, one needs to know what kind of wildlife is present in the surveyed area and what sound events are meant to be reported. This implies a proper selection of the algorithm to meet the expectations of the application.

Previously we have proposed different solutions for the problem of wildlife intruder detection [1]: a solution that uses TESPAR (Time Encoded Signal Processing and Recognition) as a method for sound encoding and classification and two standard sound classification methods, which proved to be more robust. Both of these methods were using as features the Mel Frequency Cepstral Coe-cients (MFCCs), while for training and classification were used Gaussian Mixture Models (GMM) and Support Vector Machines (SVMs). In [2] the averaged binary sparsogram implementation was presented and we have shown that this approach may be suitable for applications such as wildlife intruder detection.

In [3] we have compared some classification algorithms applied on different number of features (11, 51, 101, 151, 201, 251, 301) given by LPC (Linear Predictive Coding) in order to detect audio signals from wildlife areas. The best classifier was Logistic Model Trees regardless the number of features, having a constant classification accuracy greater than 95%. For a reduced number of features, i.e. 11, both Random Forest and Lazy IBk provide good results; the classification accuracy was greater than 98%. In [4] were presented some possible intruder detection systems and the influence of impulse-like signals upon the overall classification accuracy, based on LPC coefficients. We have used two different scenarios: in the first one five sound classes were considered (last class belongs to impulsive sounds: 'gunshot'), while in the second one we dropped out the impulsive sound class. The highest accuracy for the first scenario was for J48 classifier using 51 features (94.8%), while for the second scenario the highest accuracy was attained for Simple Logistic classifier with 101 features (98.7%). In [5] the main objective was to determine an audio signal classification system based on LPC and Random Forests. The proposed system achieved an overall correct classification rate of 99.25%. The experimental results proved that LPC coefficients can be used in the context of sound source detection. However, for wildlife intruder detection, the choice of the classifier can be essential.

One approach in selecting a proper classifier would be experimental. The purpose of our work is to define experiments using different classification algorithms and different number of features for two scenarios and evaluate the results. In our first scenario, we use as features the Linear Predictive Coding coefficients (LPCs), while in the second scenario we use the Linear Predictive Cepstral Coefficients (LPCCs). We assume that the surveillance system shall recognize the nature of the audio signal by classifying it in one of five possible classes: 'bird', 'chainsaw', 'gunshot', 'human voice' or 'tractor'. The main goal is to find the recommended number of coefficients for both LPCs and LPCCs in the context of wildlife assessment, and to choose the best approach based on the obtained results.

There are works that investigate similar aspects in the

Manuscript received September 11, 2017; revised October 14, 2017

context of speech recognition [6-10] however, for our context, meaning wildlife detection, this approach is somehow new. In [6], for a speech recognition system for Indonesian language, the highest accuracy obtained was 91.4%, using 25 samples per word, and an Artificial Neural Network (ANN) with one hidden layer (five neurons). In [8] a comparative analysis of features like LPCs and LPCCs is presented, for an automatic speech recognition system. They established that both LPCs and LPCCs reduced their performance when it comes in contact to noisy environment. In [9] a classification of lung sounds is conducted using SVMs, based on LPCCs from coefficients of 3-level of Discrete Wavelet Transform. They reported 95.42% accuracy for 42 subjects (21 used for training and 21 used for testing). The accuracy reported in [10] for LPCCs and ANN is 95.88% for isolated words, 97.02% for paired words, and 96.62% for hybrid words.

The paper is organized as follow: in Section II we recall some theoretical background regarding Linear Predictive Coding and Linear Predictive Cepstral coefficients, and we provide the used classifiers. Section III presents the experimental setup and the database used to measure the acoustic classification performance. The experimental results are the subject of Section IV, while the overall conclusion is presented in Section V.

II. THEORETICAL BACKGROUND

A. Linear Predictive Coefficients to Cepstral Coefficients

Linear Predictive Coding is mostly known and used in speech recognition. The basic steps consist of five stages [6]. First (pre-emphasis), the digital signal is passed through a low-order digital high-pass filter, in order to increase the high frequency contribution in the LPC spectrum [7]. In the second and third stage (framing and windowing), the signal is carried out for the next stage autocorrelation analyses [7]. If the order of the LPC filter is p, the autocorrelation method is used to model the filter coefficients, by converting each frame of p+1 autocorrelations into an LPCs parameter set. In the last stage, the coefficients are obtained by solving $\mathbf{Xa} = \mathbf{b}$, where

$$\mathbf{X} = \begin{bmatrix} x(1) & 0 & ... & 0 \\ x(2) & x(1) & ... & ... \\ ... & x(2) & ... & 0 \\ x(m) & ... & ... & x(1) \\ 0 & x(m) & ... & x(2) \\ ... & ... & ... & ... \\ 0 & ... & 0 & x(m) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 1 \\ a(2) \\ ... \\ a(p+1) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ ... \\ 0 \end{bmatrix}$$

However, normal equations are needed to solve the least squares problem, giving $\mathbf{X}^H\mathbf{Xa} = \mathbf{X}^H\mathbf{b}$. Usually Yule Walker equations follow, which can be solved using the Levinson's algorithm. More details about LPCs are presented in [3-5].

B. Linear Predictive Coefficients to Cepstral Coefficients

Once the LPCs have been determined we can evaluate the cepstral coefficients c_n, using a simple recursion, as in (1) [11], where σ^2 is the prediction error variance, a_k are the LPC coefficients, p is the order of the predictor, and n is the number of LPCCs.

$$c_n = \begin{cases} \ln \sigma^2, & n = 0 \\ a_n + \dfrac{1}{n}\sum_{k=1}^{n-1}(n-k)a_k c_{n-k}, & n = \overline{1,p} \\ \dfrac{1}{n}\sum_{k=1}^{n-1}(n-k)a_k c_{n-k}, & n > p \end{cases} \tag{1}$$

C. Classification Algorithms

We have selected several types of classifiers for this work [12]:

- Logistics – classification is based on a model for multinomial logistic regression having a ridge estimator;
- KStar – is based on instance classification; the class of the test instance is using the class of those training instances like it, as described by some similarity function;
- PART – classification is based on a PART decision list, using the separate-and conquer approach;
- Random Trees (RT) – classification is based on a tree that considers k randomly chosen attributes at each node, without pruning;
- REP Trees – implements a fast decision tree learner, that builds a decision/regression tree using the gain/variance and prunes it using reduced-error pruning;
- Sequential Minimal Optimization (SMO) – the classification is based on the sequential minimal optimization algorithm for training a support vector classifier, named John Platt algorithm;
- Logistic Model Trees (LMT) – classification is done by using the classification trees that use at leaves logistic regression functions;
- J48 – classification is done using a C4.5 pruned or unpruned decision tree.

III. EXPERIMENTAL SETUP

All the audio signals used to define the experiments were sampled at 16 kHz and quantized on 16 bits. The signals for 'bird' (654), 'chainsaw' (356), 'gunshot' (120) and 'tractor' (260) were recorded outside, while the 'human voice' audio signals (207) were recorded inside. None of them are studio recordings, thus they are subject of additional noise from the surroundings. The audio signals database employed for current work is the one used in [3-5, 13].

The generic design of an audio detection system is defined by a feature extracting phase and a classification phase. For audio signal characterization, we have used the two feature extraction methods: for Scenario 1 the LPCs approach, while for Scenario 2 the LPCCs approach. The feature extraction phase was conducted in MATLAB, with the help of the built-in functions present in the Signal Processing Toolbox. Similar work using LPC was done by us in [3-5].

The novelty of this research is the explore of LPCCs behavior within the known database and the use of some new classifiers. To obtain the LPCCs, first we should obtain the LPCs and the prediction error variance. In this sense p+1 features are extracted. For LPCs the feature vector of each audio signal is $[\sigma^2 \quad a_1 \quad \ldots \quad a_p]$, where σ^2 is the prediction error variance and a_k, k=1, 2, …, p are the last p prediction filter coefficients. The first one is discarded since it is always 1. This feature vector is used in advance to compute the cepstral coefficients, as in (1). The feature vector for each audio signal, for LPCCs, is $[\ln\sigma^2 \quad c_1 \quad \ldots \quad c_p]$, where the cepstral coefficients c_n, n=1, 2, …, p, were computed based on the prediction error σ^2.

In our experiments, six different features sets were computed for each class, based on the predictor's order. The chosen orders for the predictor are: 10, 12, 14, 16, 18, and 20, respectively. The reason of this choice is due to the results we have obtained in [3, 4], where we have noticed that for some type of classifiers selecting an order for the prediction filter greater than 20 would not have any noticeable impact on the obtained classification accuracy. A larger order will only increase the complexity of the detection system, which is not desired.

With the help of a specialized free tool WEKA [12], dedicated for data mining, we have extracted several classifiers as it was presented in Subsection II.C, and we have defined several experiments. To this end, each classifier was evaluated for each set of LPCs and LPCCs obtained for the corresponding order for the prediction filter. This lead to a total number of 96 experiments, that provided information regarding the performances of the correspondent algorithm.

For the validation of the experimental setup, we have used 10 times stratified 10-fold cross validation, meaning that only 90% of the data is used for the training stage, while 10% of the data is used for the testing stage [14]. We report the overall average classification accuracy, for all the experiments, while for the best result we also provide the description of the confusion matrix. A comparison between the obtained results for Scenario 1 and Scenario 2 is presented in Section IV.

IV. EXPERIMENTAL RESULTS

In this work, we propose an experimental approach in selecting a proper classifier, that might serve in designing the wildlife detection systems. In this sense, from the large variety of sounds encountered in the natural environment, as in [5, 12], we considered that the system checks if a recorded event belongs to an intruder, such as a 'human' or 'gunshot' or even 'chainsaw', signaling illegal logging.

Within Tables 1 and 2 we present the results for Scenario 1 and Scenario 2, regarding the overall average classification accuracy. We have set a light-grey highlight for the results that provided a margin greater than 95%, except the best results marked with black.

A. Scenario 1

The results from Table 1 show that the best experiment for this scenario is KStar using 14 as prediction order for the LPC filter; in this case an overall classification accuracy equal to 98.7% is obtained.

One can notice, from Table 1, that the highest accuracies, greater than 98.2% regardless the prediction filter order, are obtained for KStar. A lower performance is obtained for Random Trees and Logistic Model Trees, that provide a low margin of 95.7% for Random Trees, while the high margin does not exceed 97.9% for Logistic Model Trees.

Table 1: Scenario 1 – overall average classification accuracy (CCR [%]).

Classifier	**LPC-*p***					
	10	**12**	**14**	**16**	**18**	**20**
Logistics	92.7	93.9	93.6	94.0	94.5	94.8
KStar	98.5	98.5	**98.7**	98.7	98.6	98.2
PART	95.0	94.6	95.5	95.5	95.4	95.5
RT	96.2	95.9	95.7	96.0	95.8	95.7
REP Trees	92.3	92.3	92.7	92.9	92.7	92.4
SMO	86.6	85.9	86.5	86.2	87.3	87.8
LMT	96.7	96.6	97.2	97.9	97.8	97.9
J48	94.7	94.8	94.7	95.1	94.9	95.0

B. Scenario 2

As Table 2 presents, the second method used for feature extraction (LPCCs) provides improved performance for all the classifiers during all experiments. An extended list for the classifiers that provided results greater than 95% is defined by: Logistics, KStar, PART, Random Trees, Logistic Model Trees and J48.

Table 2: Scenario 2 – overall average classification accuracy (CCR [%]).

Classifier	**LPC-*p***					
	10	**12**	**14**	**16**	**18**	**20**
Logistics	92.7	93.4	93.0	93.0	94.9	95.5
KStar	99.3	99.2	99.2	**99.3**	99.3	99.2
PART	95.8	96.1	95.9	96.1	96.1	96.5
RT	97.0	96.5	96.6	96.7	96.5	97.0
REP Trees	92.9	93.1	93.2	93.4	93.7	93.6
SMO	90.7	91.0	90.2	91.2	91.9	92.0
LMT	97.3	98.0	98.3	98.3	98.1	98.3
J48	94.9	95.2	96.0	96.0	96.0	95.7

Outstanding results were obtained for KStar with a constant low margin greater than 99% for all the experiments, while for Logistic Model Trees 97.3%. Their best hit turned out to be for order 16 of the prediction filter, in KStar case, while for Logistic Model Trees order 14.

The best experiment from all, is pair KStar with order 16 for the prediction filter. Let us take apart this case, to study it further, based on the confusion matrix illustrated in Fig. 1.

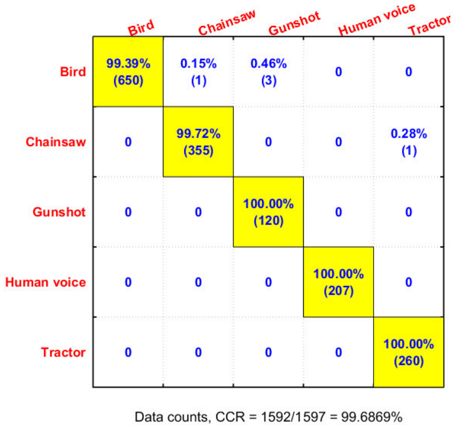

Figure 1. Confusion matrix for Scenario 2 – experiment (KStar, 16).

The rows represent true classes and the columns represent the classifier's prediction. On the main diagonal, the yellow cells represent the class probability of detection in percent and the number of correctly classified audio signals.

According with the confusion matrix from Fig. 1 we can notice that the classes with misclassified audio signals are 'bird' and 'chainsaw'. For the other three classes, all the audio signals are correctly classified.

In the case of 'bird' class, 650 instances are correctly classified, one instance is classified as 'chainsaw' sound, while three instances are classified as 'gunshot' sounds.

In the case of 'chainsaw' class, 355 instances are classified as 'chainsaw' sounds, while one instance is classified as 'tractor' sound.

The probability of detection for 'bird' class is 99.39%, for 'chainsaw' class is 99.72%, while for 'gunshot', 'human voice' and 'tractor' classes is 100%. Thus, the miss rate is 0.61% for 'bird' class and 0.28% for 'chainsaw' class. For the other three classes, the miss rate is 0%.

The precision for 'bird' and 'human voice' classes is 100%, thus there is no probability of false alarms. For 'chainsaw' class the precision is 99.72%, thus a probability of false alarm of 0.28%, for 'gunshot' class the precision is 97.56%, meaning a probability of false alarm of 2.44%, while for 'tractor' class the precision is 99.62%, thus a probability of false alarm of 0.38%.

Apparently the LPCCs method solves the problem with the gunshot sounds observed by us in works [8-10], since the broad spectrum of the impulse signal is absorbed when computing the smoothed autoregressive power spectrum, providing a more accurate characterization of the 'gunshot' signal. The results show that all 'gunshot' sounds are classified correctly, with no false alarm triggering, just one mismatch for the 'chainsaw' sounds that is being confused with a sound of 'tractor'.

V. CONCLUSION

In this work, we have performed experiments using different classification algorithms and different number of features for two scenarios and finally we have evaluated the results. In our first scenario, we have used as features the Linear Predictive Coding coefficients, while in the second scenario we have implemented the Linear Predictive Cepstral Coefficients.

The experimental results prove that LPCCs method improves greatly the classifiers performance in comparison with LPCs, in the context of wildlife detection systems. In this sense, we can consider LPCCs a better option for a more accurate characterization for these type of audio signals, within audio wildlife intruder surveillance applications.

The highest classification accuracy, 99.67%, was obtained for LPCC-16 using the KStar classifier, where 5 audio signals out of 1597 were correctly classified. We have also proved that LPCCs method solves the problem with the gunshot sounds reported elsewhere. These results have shown that all gunshot sounds are classified correctly, with no false alarm triggering.

Acknowledgment

The work of the last two authors was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI–UEFISCDI, project number PNIII-P2-2.1-BG-2016-0378, 54BG/2016, within PNCDI III.

REFERENCES

[1] M. V. Ghiurcau, C. Rusu, R. C. Bilcu, J. Astola. "Audio based solutions for detecting intruders in wild areas," *Signal Processing* vol. 92, no. 3, pp. 829–840, 2012.

[2] L. Grama, C. Rusu, "Averaged Binary Sparsogram for Wildlife Intruder Detection," *Acta Technica Napocensis - Electronics and Telecommunications*, vol. 56, no. 3, pp. 42-47, 2015.

[3] L. Grama, E. R. Buhus, C. Rusu, "Acoustic Classification using Linear Predictive Coding for Wildlife Detection Systems," in *Proc. 13th Int. Symposium on Signals, Circuits and Systems*, Iasi, Romania, July 2017.

[4] E. R. Buhus, L. Grama, C. Rusu C, "Several Classifiers for Intruder Detection Applications," in *Proc. 9th Conf. on Speech Technology and Human-Computer Dialogue*, Bucharest, Romania, July 2017, pp. 1-6.

[5] L. Grama, C. Rusu, "Audio Signal Classification Using Linear Predictive Coding and Random Forests," in *Proc. 9th Conf. on Speech Technology and Human-Computer Dialogue*, Bucharest, Romania, July 2017, pp. 1-9.

[6] T. Wijoyo, S. Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot," in *Proc. Int. Conf. on Information and Electronics Engineering*, vol. 6, 2011, pp. 179-183.

[7] S. Furui. *Digital Speech Processing, Synthesis, and Recognition*. 2nd edn. CRC Press, 2001.

[8] H. Gupta, D. Gupta, "LPC and LPCC method of feature extraction in speech recognition system," in *Proc. 6^{th} Int. Conf. on Cloud System and Big Data Engineering (Confluence)*, Noida, India, Jan. 2016, pp. 498-502.

[9] M. M. Amzy, "Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine," in *Proc. IEEE Jordan Conf. on Applied Electrical Engineering and Computing Technologies*, Ammam, Jordan, Nov. 2015, pp. 1-5.

[10] T. Gulzar, A. Singh, A. Sharma "Comparative Analysis of LPCC, MFCC and BFCC for the Recognition of Hindi Words using Artificial Neural Networks," *Int. J. of Computer Applications*, vol. 101, no. 12, pp. 22-27, Sept. 2014.

[11] S. Pazhanirajan, P. Dhanalakshmi, "EEG Signal Classification using Linear Predictive Cepstral Coefficient Features," *Int. J. of Computer Applications*, vol. 73, no. 1, pp. 28-31, 2013.
[12] WEKA – The University of Waikato. *Weka 3: Data Mining Software in Java*. [Online]. available: http://www.cs.waikato.ac.nz/ml/weka, version 3.8.0
[13] L. Grama, L. Tuns, C Rusu, "On the Optimization of SVM Kernel Parameters for Improving Audio Classification Accuracy," in *Proc. 14th Int. Conf. on Engineering of Modern Electric Systems*, Oradea, Romania, June 2017, pp. 224-227.
[14] I. Witten, E. Frank, M. A. Hall, C. J. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*. 4th edn. Elsevier: The Morgan Kaufmann Series in Data Management Systems, 2016.