Analysis of Linear Predictive Coefficients for Gunshot Detection Based on Neural Networks

Martin Hrabina
Brno University of Technology
Department of Radio Electronics
Technicka 12, 61600 Brno, Czech Republic
hrabina@phd.feec.vutbr.cz

***Abstract*—This work deals with analysis of the linear predictive coefficients with respect to their use for acoustic gunshot detection. First, coefficient stability was observed when changing length and position of an analysed signal frame. Then, the optimal prediction order was investigated. Finally, false alarms and correct gunshot detections were tested for various numbers of coefficients. The best experimental results achieved were 8.6% false alarm rate and 88% gunshot detection rate.**

Keywords*—*signal processing, feature extraction, gunshot detection, LPC coefficients

I. Introduction

Gunshots are well-known impulsive sounds and most people can reliably distinguish a gunshot from other acoustic events without any special training. Our long-term goal is to develop a simple system for automatic real-time detection of gunshots. This system should help in protecting wild elephants from poachers doing illegal business with the ivory. The motivation for this research is to assist the international organisation "Save Elephants" which is engaged in protecting elephants in Central Africa.

In general, commonly used types of features in sound recognition are autocorrelation coefficients (ACC), linear predictive coding (LPC) coefficients, linear predictive cepstral coefficients (LPCC), mel-frequency cepstral coefficients (MFCC) as well as MPEG-7 based descriptors. A detailed introduction of ACC, LPC, LPCC and MFCC including basic properties and formulae for calculation may be found in [1]. Note that MFCC is a widespread acoustic feature set inspired by the model of the human auditory system. Some useful MPEG-7 descriptors are divided into primary descriptors (obtained from audio spectrogram) and secondary descriptors (based on variations of the primary descriptors) and are introduced in [2]. There are many strategies for extracting and selecting features from the audio signal for recognizing of specific events. Simple systems for gunshot detection based on distance measures and low-level features are described, for example, in [3] and [4]. Feature selection and feature reduction by a principal component analysis was applied for acoustic event detection in [5] and [6]. An efficient algorithm using matching pursuit for classifying of eight acoustic events including gunshot is presented in [7]. A comprehensive review of current gunshot detection technologies is provided in [8].

II. Investigated Features

A. Sound Data

Acoustic signals used in our experiments can be classified into two sound categories. The first category contains gunshots, and the second one covers a variety of sounds that can occur at the same places as gunshots but must be ignored by the gunshot detector. In total, five weapons (different instances of Tikka and Ruger Six) were selected from the database Firearm Sounds [9] using shots from different angles, both front and back. Other sounds were obtained partly from the Urban Sounds Dataset [10] and partly from our own recordings. These sounds cover a siren, motorcycle engine, children playing, loud speech, barking dogs, plane flying by, sounds of wood processing (e.g., breaking of branches, chopping) as well as elephant sounds. All sounds are recordings of real-life acoustic events, no simulated sound was used. All acoustic signals are in WAV format (mono, 44.1 kHz, 16 bit). Some recordings originally had a higher sampling frequency, but for purposes of our experiments were resampled at 44.1 kHz. Audio input was segmented into frames of 3 to 10 ms using a rectangle window.

B. Feature Sets

In our analysis, we have used four feature sets for gunshot detection, namely ACC, LPC, MFCC, and LPCC. All features were estimated using Matlab software. Two feature sets were extracted using native Matlab functions: ACC by *autocorr* and LPC by *lpc*. The *lpc* function calculates LPC coefficients using the Levinson-Durbin recursive algorithm [11]. For calculating MFCC, the standard approach based on energy in mel-frequency subbands [12] was implemented. LPCC were derived from LPC coefficients [13] by applying the following recursive transformation

$$c_1 = a_1 \tag{1}$$

$$c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k} \quad \text{for} \quad m = 2,\ldots,M \tag{2}$$

where a_m are the LPC coefficients, c_m denotes the cepstral coefficients and M is the total number of coefficients considered in each feature set. Figure 1 shows the overall feature extraction procedure.

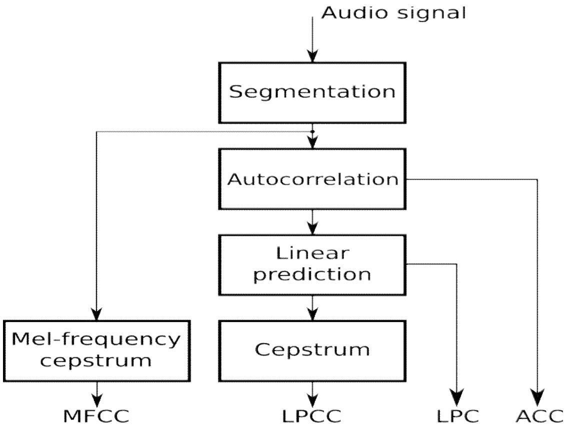

Fig. 1. Block diagram for feature extraction.

III. Experimental Results

A. Comparison of Feature Sets

Primarily, the performance of all feature sets was estimated and compared using *Neural Net Pattern Recognition* in Matlab with a default configuration of 10 neurons in one hidden layer. The default configuration was retained because this paper does not explore the effects of network configuration and the network is used only to reflect the feature's discriminatory capacity so our metric can be compared. More on optimization of neural network hyper-parameters can be found in [14]. In total, 22523 signal frames were processed in the experiments. For each number of coefficients, the total amount of sound data was randomly divided into three parts as follows: 70% for training, 15% for validation, and 15% for testing. Table 1 summarizes achieved results from the viewpoint of false alarm rates (in percent). False alarm reduction seems to be more important than maximization of correct gunshot detections due to the fact that multiple gunshots occur during hunting.

As can be seen in Tab. 1, the best results were mostly achieved using LPC coefficients, closely followed by LPCC and then by MFCC. On the other hand, autocorrelation coefficients performed very poorly in this task. Therefore, our further investigation was focused on evaluating individual coefficients from the best feature set, i.e. LPC coefficients. In the past, linear prediction analysis was used predominantly for speech processing, but it can also be successfully applied for recognizing specific acoustic signals [15], [16].

TABLE I. False Alarms Detected with Different Feature Sets

Number of coefficients	**Feature set**			
	AC	***LPC***	***LPCC***	***MFCC***
8	90.00 %	16.70 %	15.40 %	15.60 %
12	60.00 %	12.20 %	13.70 %	13.10 %
16	40.00 %	11.50 %	12.60 %	16.60 %
20	46.00 %	10.67 %	11.62 %	16.80 %
Average performance	59.00 %	12.77 %	13.33 %	15.53 %

B. Variability of Coefficients Values

Variability of individual LPC coefficients was investigated in terms of their sensitivity to different frame lengths and frame positions in relation to the whole gunshot signal. A desirable coefficient characterizing a gunshot should keep a constant value when both frame length and position change. The individual gunshots are usually shorter than 10 ms having maximal instantaneous power at the gunshot´s beginning and decreasing at the end, as depicted in Fig. 2.

Note that the LPC coefficients incorporate spectral properties of the analysed signal frame and thus, change of coefficient values represents a change of spectrum. Depending on the LPC order, i.e. number of LPC coefficients used in series, the spectrum is more or less smoothed. This is useful especially in signals which have some local maxima (formants) in the spectrum. Such spectra are typical, for instance, in vowels [17] but a similar structure also appears in gunshots.

C. Effect of Frame Length

Four frame lengths (3 ms, 5 ms, 8 ms, and 10 ms) have been used for coefficient extraction to explore the effect of frame lengths on coefficient variability. Time relation of these lengths to the gunshot wave is illustrated in Fig. 2. In the experiment, various orders of LPC coefficients from 6 to 30 were taken into account. An example of the values of all 20 coefficients for 20th LPC order for different frame lengths can be seen in Fig. 3. In a brief graphical evaluation–the straighter the line, the less variable the coefficient is (i.e. more suitable feature). Achieved results for the best 3 individual coefficients within each LPC order from 6th to 30th are shown in Tab. 2.

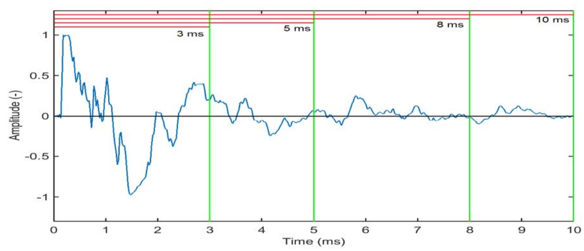

Fig. 2. Relation of various frame lengths (red lines) to gunshot duration.

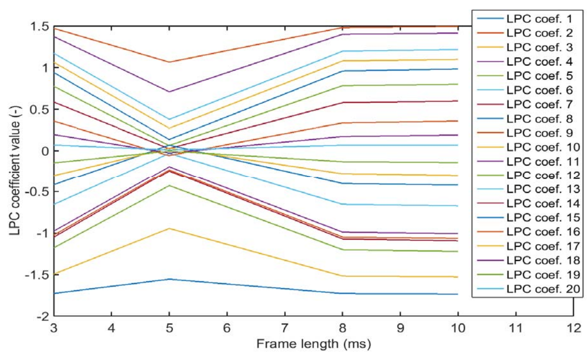

Fig. 3. Values of 20 LPC coefficients for various frame lengths.

TABLE II. VARIABILITY DUE TO DIFFERENT FRAME SIZE

LPC order	Best three coefficients			Average of triad
	1st	*2nd*	*3rd*	
6	0.059	0.143	0.148	0.117
8	0.066	0.165	0.165	0.132
10	0.055	0.105	0.159	0.106
12	0.091	0.186	0.204	0.161
14	0.073	0.179	0.196	0.149
16	0.080	0.223	0.223	0.176
18	0.089	0.232	0.232	0.184
20	0.097	0.234	0.244	0.192
22	0.090	0.239	0.261	0.197
24	0.056	0.104	0.220	0.126
26	0.087	0.210	0.269	0.189
28	0.064	0.193	0.277	0.178
30	0.068	0.181	0.256	0.169

The first, second and third best coefficients across all LPC orders were in 10th order (highlighted in Tab. 2). Note that this table does not indicate series indices of the best coefficients. Highlighted values in the Tab. 2 represent the last coefficients in series for 10th LPC order, namely a_{10}=0.055, a_9=0.105, a_8=0.159 giving the minimal average of 0.106. The next low averages are 0.117 for 6th order and 0.126 for 24th order.

D. Effect of Frame Shift

In this test, coefficient variability with respect to the frame position is explored. A frame with a constant length of 3 ms was chosen, which is small enough in comparison with the duration of the gunshot wave. The frame was shifted step-by-step along the whole gunshot wave with 50% overlapping as illustrated in Fig. 4. All frame operations carried out in this and previous tests, i.e. changes of length and changes of position, were realized by means of signal windowing using a flexible rectangular window. The obtained results were processed similarly as in the previous section. Table 3 shows the best 3 individual coefficients within each LPC order. The best order seems to be the 24th order (highlighted values in Tab. 3) with an average of 0.231 followed by the 30th order with an average of 0.232.

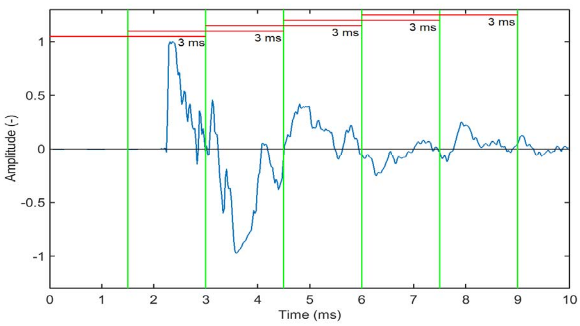

Fig. 4. Illustration of various frame positions marked by red line segments.

TABLE III. VARIABILITY DUE TO DIFFERENT FRAME SHIFT

LPC order	Best three coefficients			Average of triad
	1st	*2nd*	*3rd*	
6	0.2617	0.7805	0.8785	0.640
8	0.3234	0.8274	0.9682	0.706
10	0.1600	0.5244	0.8992	0.528
12	0.1888	0.3505	0.4734	0.338
14	0.1787	0.3591	0.4634	0.334
16	0.1484	0.3510	0.4026	0.301
18	0.1917	0.3517	0.5377	0.360
20	0.1889	0.2643	0.4906	0.315
22	0.2680	0.2850	0.3393	0.297
24	0.0986	0.2302	0.3655	0.231
26	0.1489	0.2662	0.3315	0.249
28	0.2121	0.2673	0.3016	0.260
30	0.1464	0.2679	0.2815	0.232

Comparing the results presented in Tab. 2 and Tab. 3, the 24th order seems to give the most efficient feature set among all tested LPC orders. Hence, our further analysis was focused more in detail on this coefficient set. Table 4 shows the variability of individual coefficients a_m estimated as average difference

$$\overline{\Delta}_m = \frac{\sum_{k=1}^{K}\sum_{p=1}^{P-1}(a_{m,k,p+1} - a_{m,k,p})}{K(P-1)} \tag{3}$$

where m is series index of coefficients, $1 \leq m \leq 24$, k is gunshot index, p is index of frame position, and $a_{m,k,p}$ are corresponding LPC coefficients.

TABLE IV. AVERAGE DIFFERENCES IN 24TH ORDER LPC COEFFICIENTS ACROSS FRAME SHIFT

Coefficient series index	Average difference	Coefficient series index	Average difference
1	1.056	13	0.788
2	2.371	14	0.846
3	2.694	15	0.873
4	2.586	16	0.834
5	2.172	17	0.716
6	1.805	18	0.587
7	1.513	19	0.517
8	1.231	20	0.442
9	0.838	21	0.457
10	0.559	22	0.366
11	0.535	23	0.230
12	0.716	24	0.099

E. Performance of the LPC Coefficients

Due to the results obtained in previous tests, linear prediction of the 24th order was chosen for more complex testing of individual LPC coefficients. False alarms were investigated primarily. The number of coefficients used in individual tests started at 2 coefficients and then increased to the full set of 24 coefficients by adding the individual coefficients in two ways, according to the principles "best *N*" and "first *N*". In the case of "best *N*", coefficients were ordered from the best to worst (in terms of numerical stability expressed by average differences in Tab. 4) and added in the following order: a_{24}, a_{23}, a_{22}, a_{20}, a_{21}, …, a_3. In the case of "first *N*", coefficients were added by increasing index number, i.e.: a_1, a_2, a_3, a_4, a_5, …, a_{24}. Fig. 5 compares false alarms achieved with various numbers of coefficients for signal frame length of 3 ms, 5 ms, 8 ms, and 10 ms. Minimal false alarm rates are 8.6% using 19 of 24 coefficients in the "best *N*" case and 9.2% using 15 of 24 coefficients in the "first *N*" case, both for frame length of 10 ms. Generally, the frame length 10 ms gives better results than other lengths.

In the final experiment, the gunshot detection rate was investigated for various numbers of coefficients in the same manner as in the previous tests aimed at false alarm rate. Figure 6 shows the results for a frame length of 10 ms. The best detection rate achieved is 88%.

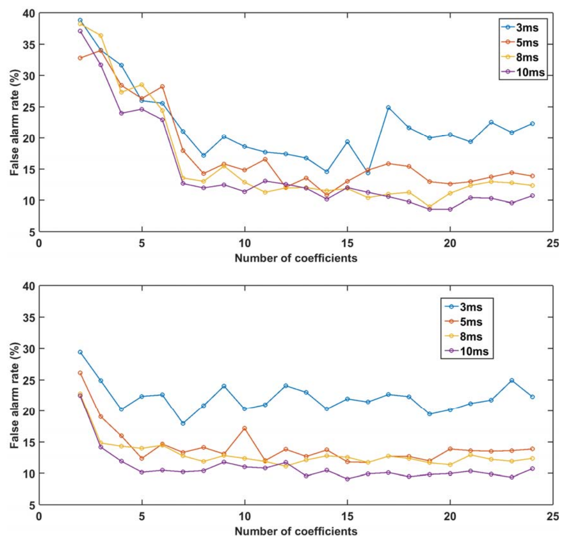

Fig. 5. Effect of the number of used coefficients on false alarm rate for various frame lengths when adding coefficients as "best *N*" (upper graph) and "first *N*" (lower graph).

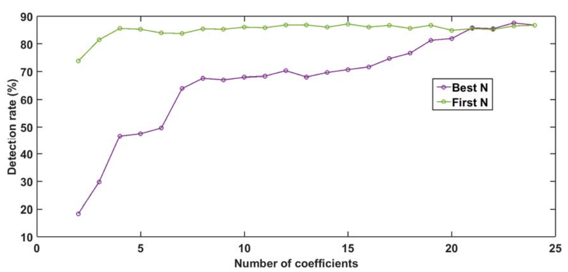

Fig. 6. Effect of the number of used coefficients on detection rate.

IV. Conclusion

Comparing both principles "best *N*" and "first *N*", the "first *N*" performs better for a small number of coefficients in general. Here, good results in false alarms as well as in gunshot detections were achieved using the first 5 coefficients from the total of 24 coefficients associated with the 24th LPC order. However, when absolute low false alarm rate is required, the principle "best *N*" should be applied. The order 24 seems to be very good for this purpose. Our future work will focus on the influence of nonstationary noises [18], overlapping acoustic events and a combination of different features.

Acknowledgment

Research described in this paper was financed by the Czech Ministry of Education within the frame of the National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used. The research was also financially supported by the Brno University of Technology Internal Grant Agency under project no. FEKT-S-17-4707.

References

[1] L.R. Rabiner and R.W. Schafer, Theory and Applications of Digital Speech Processing. London: Prentice Hall, 2011.

[2] J. Luque, D. Larios, E. Personal, J. Barbancho, and C. Leon, "Evaluation of MPEG-7-based audio descriptors for animal voice recognition over wireless acoustic sensor networks," Sensors, 2016, vol. 16, no. 5, pp. 7462-7498.

[3] A. Chacon-Rodriguez, P. Julian, L. Castro, P. Alvarado, and N. Hernandez, "Evaluation of gunshot detection algorithms," IEEE Transactions on Circuits and Systems, 2011, vol. 58, no. 2, pp. 363-373.

[4] M. Hrabina and M. Sigmund, "Acoustical detection of gunshots," in Proc. 25th Intern. Conference Radioelektronika, 2015, pp. 150-153.

[5] X. Zhuang, X. Zhou, T. S. Huang, and M. Hasegawa, "Feature analysis and selection for acoustic event detection," in Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing, 2008, pp. 17-20.

[6] E. Vozáriková, P. Viszlay, J. Juhár, and A. Čižmár, "Acoustic events detection via PCA-based feature extraction," Journal of Computer Science and Control Systems, 2010, vol. 3, no. 2, pp. 99-102.

[7] Q. Nguyen and J. Choi, "Matching pursuit based robust acoustic event classification for surveillance systems," Computers & Electrical Engineering, 2017, vol. 57, pp. 43-54.

[8] J.R. Aguilar, "Gunshot detection systems in civilian law enforcement," Journal of Audio Engineering Society, 2015, vol. 63, no. 4, pp. 280-291.

[9] http://www.airbornesound.com/

[10] J. Salamon, C. Jacoby, and J.P. Bello, "A dataset and taxonomy for urban sound research," in Proc. of 22nd ACM International Conference on Multimedia, Orlando, 2014, pp. 1041-1044.

[11] B. Fazlali and M. Eshghi, "A pipeline design for implementation of LPC feature extraction system based on levinson-durbin algorithm," in Proc. of 19th Iranian Conference on Electrical Engineering, 2011, pp. 1-5.

[12] A. C. Kelly and Ch. Gobl, "A comparison of mel-frequency cepstral coefficient (MFCC) calculation techniques," Journal of Computing, 2011, vol. 3, no. 10, pp. 62-66.

[13] E. Salomon and P. Havinga, "A survey on the feasibility of sound classification on wireless sensor nodes, Sensors, 2015, vol. 15, no. 4, pp. 7462-7498.

[14] S. C. Smithson, G. Yang, W. J. Gross, and B. H. Meyer, "Neural networks designing neural networks: multi-objective hyper-parameter optimization", in Proc. of 35th International Conference on Computer-Aided Design, 2016, pp. 1-8.

[15] M. Sigmund, “Search for keywords and vocal elements in audio recordings,” Elektronika ir Elektrotechnika, 2013, vol. 19, no. 9, pp. 71-74.

[16] K. Morio, H. Wada, Y. Tadahira, “The method of identification for the type of aircrafts,” in Proc. of 44th Intern. Congress and Exposition on Noise Control Engineering, San Francisco, 2015, pp. 181-190.

[17] M. Sigmund, “Influence of psychological stress on formant structure of vowels,” Elektronika ir Elektrotechnika, 2012, vol. 18, no. 10, pp. 45-48.

[18] P. Zelinka and M. Sigmund, “Hierarchical classification tree modeling of nonstationary noise for robust speech recognition,” Information Technology and Control, 2010, vol. 39, no. 3, pp. 202-210.