Comparison of Feature Performance in Gunshot Detection Depending on Noise Degradation

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

***Abstract*—This paper compares three different features and various feature orders for the purpose of determining the best feature for gunshot detection under adverse noise condition. Compared features cover LPC, LPCC and MFCC with orders from 8 to 30. All features were extracted from sounds with the sound-to-noise ratios 30, 20, 10, and 0 dB. The background noise was simulated by white noise. Experimental results indicate that LPC coefficients are the most efficient features, especially for low noise. On the other hand, MFCC performed well in noisy environments at 10 dB and 20 dB.**

***Keywords*—gunshot detection, feature analysis, linear predictive coding coefficients, cepstrum, noise**

I. Introduction

Our research work is motivated by the goal of developing an effective system for automated detection of gunshots in open nature that alerts police of poachers. The system should be primarily used in the protection of wild elephants against poaching for ivory. Currently, many elephant herds are collared with a Global Positioning System (GPS) radio module allowing to monitor the elephant's mobility and migration. The gunshots detection module will be integrated in the GPS collar and after an incident, information about the poaching event together with its location can be sent to police. Then, an anti-poaching team could react promptly to apprehend the poachers at the place of the crime.

We have recently developed quite a good working gunshot detection system [1] having a relatively simple structure of signal processing. In order to gain early experience with practical integration of the gunshot detection module to the existing monitoring network, this system was implemented in C for a real-time application by means of a signal processor [2] and the source code was given to an industrial company which is currently finalizing production of some prototypes. In parallel with the hardware realization of the first version of the developed system, further research is continuing which is focused on searching for specific features that are very effective for gunshot detection.

The aim of this paper is to determine which commonly used features for acoustic event detection are most suitable for gunshot detection under various noise conditions. In general, features are often chosen by comparing their performance in previous similar situations. Alternatively, big feature sets are composed and then feature reduction techniques are applied, e.g. principal component analysis [3]. More systematic feature selection methods are usually based on measures expressing class separability [4] or statistical measures [5]. A useful approach of feature selection for acoustic event detection based on maximum relevance and minimum redundancy can be found in [6].

II. Used Signals and Methods

This section describes used data, observed features and methods employed in the evaluation.

A. Used Signal Data

Sounds used include gunshots from free the Firearm Sounds dataset [7] and non-gunshot sounds obtained partly from the Urban Sounds Dataset [8] and partly from own previous recordings. All sounds were considered in two-class pattern recognition: gunshots and non-gunshots. The non-gunshots include jackhammers, drills, traffic, music, etc. Noise was introduced using the Matlab function *awgn* with parameter 'measured' and signal-to-noise ratios 30, 20, 10 and 0 dB. White noise was chosen for its flat spectral characteristics, thus the possibility of worst degradation of features. In the following tests, it was planned to add other noise types and also natural noises (such as wilderness or urban sounds). All sounds are stored in WAV format (mono, 44.1 kHz sampling rate, 16 bit resolution). Some recordings having originally higher sampling frequency were resampled at 44.1 kHz for purposes of our investigation. The processed input signals were segmented into frames of 11 ms using a rectangle window. This frame size resulted from findings in previous tests with gunshot signals [9], where we obtained good results in terms of detection accuracy for 11 ms and 23 ms (1024 samples) frames, slightly worse results for shorter frames (8 ms, 5 ms) and significantly worse results for 3 ms.

In order to present used signals, Fig. 1 illustrates a typical gunshot from an AK-47 in time domain. Figure 2 shows the corresponding spectrum represented by the maximum and minimum boundaries for smoothened spectra of 7 separate gunshots from an AK-47 (caliber 7.62x39 mm). All signals presented in Fig. 1 and Fig. 2, were recorded in an open field with a microphone placed 100 m in front of the shooter.

978-1-5090-4591-4/17/$31.00 ©2017 IEEE

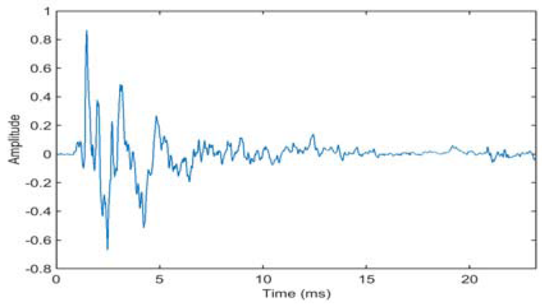

Fig. 1. An example of single gunshot (AK-47) in time domain.

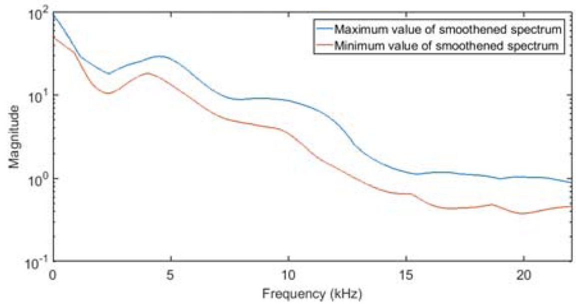

Fig. 2. Maximum and minimum values for smoothened spectrum of several gunshots.

B. *Investigated Features*

In this work, we compare widely used linear predictive coding coefficients (LPC), linear predictive cepstral coefficients (LPCC) as well as mel-frequency cepstral coefficients (MFCC) having various orders incremented step-by-step from 8 up to 30. These coefficients were chosen due to their successful use in similar applications, such as [10], [11] and our previous experience with them when searching for vocal elements in audio recordings [12]. A detailed description of these features may be found, for example in [13] and [14]. LPC coefficients were computed using native Matlab function. LPCC were derived from the LPC. MFCC were extracted using an implemented algorithm. Figure 3 illustrates numerical values of LPC coefficients of order 30 for 7 individual gunshots recorded under the same conditions.

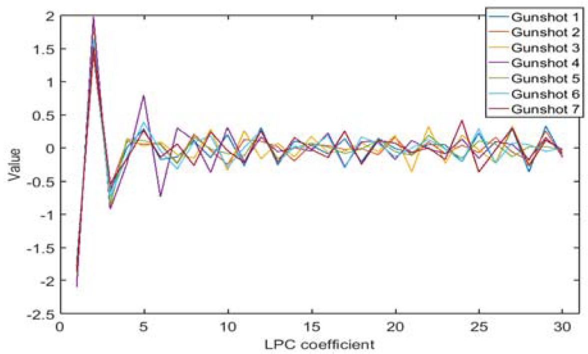

Fig. 3. Values of individual LPC coefficients for order 30.

C. *Artificial Neural Networks*

In our experiments, the neural networks available in Matlab as a pattern recognition tool were used for training and as a detector. These networks are feedforward networks with one hidden layer with a variable number of neurons that use the sigmoid function. The properties of neural networks, i.e. hyper-parameters, can be complicated to optimize. Paper [15] discusses automatic optimization of hyper-parameters for both complexity and performance. The number of neurons in the output layer corresponds to the number of output classes and these neurons use the softmax function. A training algorithm uses the scaled conjugate gradient backpropagation function. The default division of input data is 70% for training, 15% for validation and 15% for testing with the possibility of changing the division. However, similar data division are used in other (published) detectors such as in [16] where approximately 68% of data is used for training or in [6] where 90% of data is used. The data is divided randomly by default, however, the user can select different methods such as using contiguous blocks, the interleaved method or specify which data goes where. Our data was divided into non-overlapping sets using a random division in Matlab. In each experiment, the data was assigned newly to the mentioned groups. For evaluating network performance, the cross-entropy function was used. The algorithm heavily penalizes strong inaccuracies.

Due to different initial conditions and distribution of data in the above mentioned groups, each trained network can perform differently. As a consequence, all results obtained from neural networks are an average of five training results. Deep neural networks with MFCC as features were successfully used for example, on the first and second heart sound recognition [16].

III. EXPERIMENTAL RESULTS

All experiments were performed on a desktop using Matlab's neural networks with audio files recorded in real life environments. First, we have worsened acoustic signals with different levels of noise. Then each feature set LPC, LPCC and MFCC was investigated separately in order to estimate their efficiency for gunshot detection in noisy environments. The number of coefficients used in all tests begins at 8 and increases in steps of 2 up to the full set of 30 coefficients. The obtained experimental results are summarized in Figs. 4 to 7.

Figure 4 presents false alarms rate for feature orders from 8 to 30 for a signal without noise. LPC and LPCC coefficients achieved similar results, and MFCC performed comparatively worse. A slight decrease in false alarm rate can be seen for an initial increase in order, but a further increase of coefficients does not cause significant improvement.

The acceptance of false alarm plays a significant role in acoustic event detection. For instance, an error rate less than 5% is considered acceptable in automatic speech recognition. In gunshot detection, more insight into the application is required (human factor is very important). The achieved results present individual performance of coefficients and serve only as a comparison. The aim here was not to compile the best set of features for minimal false alarms or maximal true detections.

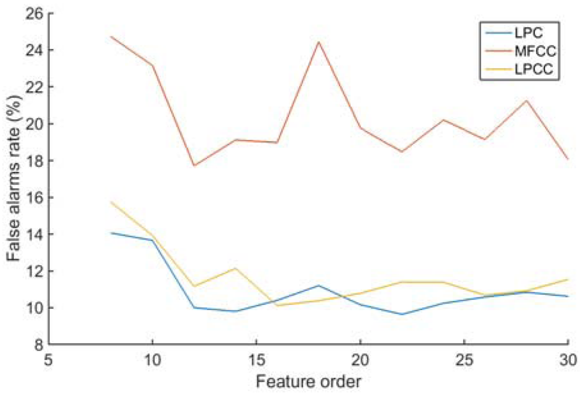

Fig. 4. False alarms rate for different features for clean signals.

Figures 5 to 7 compare true detection rates for all observed feature sets and for orders from 8 to 30 with different signal-to-noise ratios (SNR). LPC and LPCC again performed relatively well, above all for clean signals and signals with SNR=30 dB.

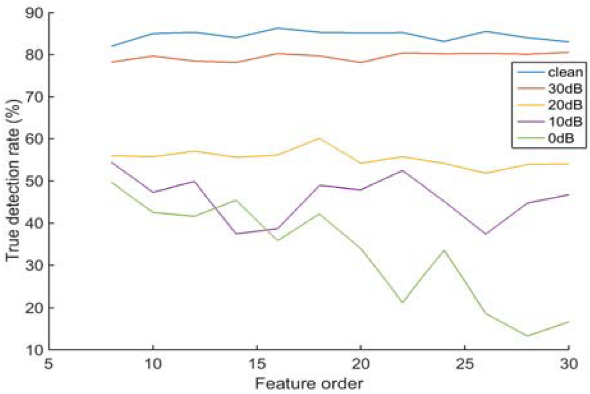

Fig. 5. LPCC true detection rate for different noise conditions and different feature order.

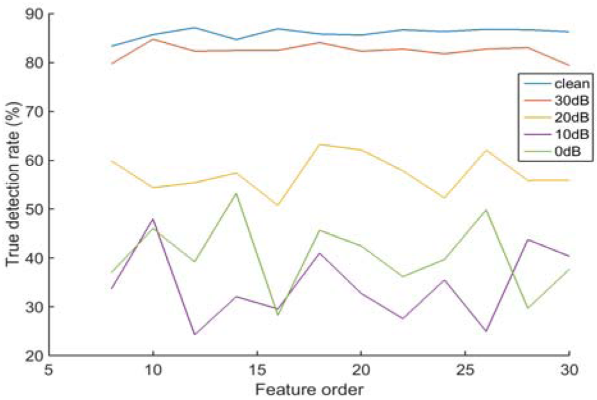

Fig. 6. LPC true detection rate for different noise conditions and different feature order.

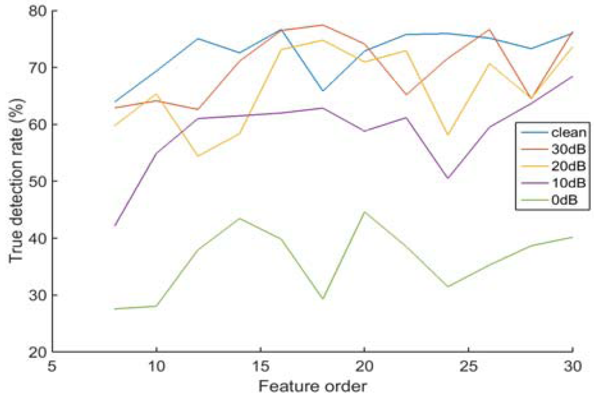

Fig. 7. MFCC true detection rate for different noise conditions and different feature order.

To evaluate the detection performance, besides the graphical representations, a quantitative criterion as weighted difference of true detection (*TD*) and false alarm (*FA*) was used as follows

$$\Delta = \alpha \cdot TD - \beta \cdot FA, \tag{1}$$

where α and β are weights depending on requirements of the application or user. Tables 1 and 2 illustrate the results using unitary weights, i.e. α=1 and β=1, as well as different weights α=0.33 and β=0.66, respectively.

TABLE I. RESULTS FOR DIFFERENT NOISE LEVELS USING UNITARY WEIGHTS α AND β

SNR	LPC			LPCC			MFCC		
	***TD* [%]**	***FA* [%]**	***Δ* [%]**	***TD* [%]**	***FA* [%]**	***Δ* [%]**	***TD* [%]**	***FA* [%]**	***Δ* [%]**
clear	86,7	9,6	77,1	86,3	10,1	76,2	75,1	17,7	57,4
30dB	82,7	13,9	68,8	80,5	18,7	61,8	77,5	16,8	60,7
20dB	62,1	27,4	34,7	60,0	29,6	30,4	74,8	19,4	55,4
10dB	43,8	32,9	10,9	54,4	30,2	24,2	68,5	22,5	46
0dB	49,8	34,6	15,2	49,7	35,3	14,4	44,6	30,9	13,7

TABLE II. RESULTS FOR DIFFERENT NOISE LEVELS USING WEIGHTS α=0.33 AND β=0.66

SNR	LPC			LPCC			MFCC		
	***TD* [%]**	***FA* [%]**	***Δ* [%]**	***TD* [%]**	***FA* [%]**	***Δ* [%]**	***TD* [%]**	***FA* [%]**	***Δ* [%]**
clear	86,7	9,6	22,3	86,3	10,1	21,8	75,1	17,7	13,1
30dB	82,7	13,9	18,1	80,5	18,7	14,2	77,5	16,8	14,5
20dB	62,1	27,4	2,4	60,0	29,6	0,3	74,8	19,4	11,9
10dB	43,8	32,9	-7,3	54,4	30,2	-2,0	68,5	22,5	7,8
0dB	49,8	34,6	-6,4	49,7	35,3	-6,9	44,6	30,9	-5,7

In both cases, LPC achieved the best results for clean signals and SNR of 30 dB. For 20 dB and 10 dB, MFCC provided significantly better results. At 0 dB MFCC and LPC were similarly effective, with MFCC achieving less false alarms but LPC achieving more true detections.

The above described should not serve primarily as a gunshot detector for real-life usage, but should provide insight into performance of these coefficients under different noise conditions. However, for informative purposes, Tab. 3 provides execution times of computations in Matlab expressed in milliseconds. Note that the data in Tab. 3 were measured without signal preprocessing. Each script was "warmed up" before timing in order to eliminate compilation times and other overhead. Timing itself was performed using the Matlab function pair *tic-toc* and the execution times in Tab. 3 are averages of three executions.

TABLE III. EXECUTION TIMES OF DIFFERENT FUNCTIONS IN MILLISECONDS

Function	**Order**		
	10	**20**	**30**
LPC	0.621	0.443	0.581
LPCC	0.754	0.553	0.567
MFCC	3.937	4.149	5.394
Neural Net	0.497	0.560	0.519

As can be seen, MFCC take the longest to compute, which might be caused by how it was written (since it is the only non-native function used). Furthermore, each function has its time inflated by input checking which would not be needed in application-specific code. In addition, the execution times would be different depending on the device on which the process runs. We were using a computer equipped with an Intel Core Quad CPU (3 GHz) and 8 GB RAM, running Matlab 2016b under Windows 7.

IV. CONCLUSION

The experimental results indicate that feature order is not a significant factor in determining recognition performance beyond a certain threshold. For higher order, true detection rates as well as false alarm rates tend to follow almost a constant mean value within each SNR condition. The only exception seems to be LPCC at 0 dB, where a decrease in true detections was observed with increasing feature order. LPC coefficients provided best overall performance followed by LPCC with detection rates between 80% and 87% for a signal without added noise and a signal with SNR=30 dB. MFCC coefficients exhibited more random behavior when detection rates vary more dramatically with changing order without an observable trend, but performed best at 20 dB and 10 dB SNR. In general, the detection rate achieved in our simulations is not optimal by any means. However, note that the primary goal of this work is not to develop the best detector but rather to compare the performance of noisy features for a given classifier. The main contribution of this paper is comparing features frequently used in recognition of acoustic events for the task of gunshot detection in various noise conditions.

In future work, we plan to extend the tests on noisy signals by adding fluctuated noises and impulsive noises [17]. We will also search for the most effective combinations of individual features and feature subsets using a lower sampling frequency.

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.

REFERENCES

[1] M. Hrabina and M. Sigmund, "Acoustical detection of gunshots," in Proc. of 25th Int. Conf. Radioelektronika, Pardubice, 2015, pp. 150-153.

[2] M. Hrabina and M. Sigmund, "Implementation of developed gunshot detection algorithm on TMS320C6713 processor," in Proc. of SAI Computing Conference, London, 2016, pp. 902-905.

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

[4] L. Gerosa, G. Valenzise, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection in noisy environments", in Proc. of 15th European Signal Processing Conference, 2007, pp. 1216-1220.

[5] E. Kiktova, M. Lojka, M. Pleva, J. Juhar, and A. Cizmar, "Gun type recognition from gunshot audio recordings", International Workshop on Biometrics and Forensics, 2015, pp. 1-6.

[6] E. Vozarikova, J. Juhar, and A. Cizmar, "Acoustic Event Detection Based on MRMR Selected Feature Vectors", Journal of Electrical and Electronics Engineering, vol. 5, no. 1, pp. 277-282, 2012. Firearm

[7] B. Jaszczak and B. Nelson, (2014). Free Firearm Sound Effects Library. [online] Available at: http://www.airbornesound.com/downloads/free-firearm-sound-effects-library-expanded-edition [Accessed March 2017].

[8] 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, USA, 2014, pp. 1041-1044. UrbanSound

[9] M. Hrabina and M. Sigmund, "Analysis of linear predictive coefficients for gunshots detection based on neural network", in Proc. of 26th Int. Symposium on Industrial Electronics (ISIE), Edinburgh, 2017, in press.

[10] S. Yang, J. Cao, and J. Wang, "Acoustics recognition of construction equipments based on LPCC features and SVM", in Proc. of 34th Chinese Control Conference (CCC), 2015, pp. 3987-3991.

[11] I. L. Freire and J. A. Apolinário Jr., "Gunshot detection in noisy environments", in Proc. of 7th International Telecommunications Symposium, Manaus, Brazil, 2010, pp. 1-4.

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

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

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

[15] 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.

[16] T.-E. Chen et al., "S1 and S2 Heart Sound Recognition Using Deep Neural Networks", IEEE Transactions on Biomedical Engineering, vol. 64, no. 2, pp. 372-380, Feb. 2017.

[17] 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.