Efficient Feature Set Developed for Acoustic Gunshot Detection in Open Space

Milan Sigmund*, Martin Hrabina
Faculty of Electrical Engineering and Communication, Brno University of Technology,
Technicka 12, CZ-61600 Brno, Czech Republic
sigmund@feec.vutbr.cz

***Abstract*—This paper presents an efficient approach to automatic gunshot detection based on a combination of two feature sets: adapted standard sound features and hand-crafted novel features. The standard features are mel-frequency cepstral coefficients adapted for gunshot recognition in terms of uniform gamma-tone filters linearly spaced over the whole frequency range from 0 kHz to 16 kHz. The first 18 coefficients calculated from the 41 filters represent the best set of the optimized cepstral coefficients. The novel features were derived in the time domain from individual significant points of the raw waveform after amplitude normalization. Experiments were performed using single and ensemble neural networks to verify the effectiveness of the novel features for supplementing the standard features. The novelty of the work is the proposed feature combination, which allows to achieve very effective detection of gunshots from hunting weapons using 23 features and a simple neural network. In binary classification, the developed approach achieved an accuracy of 95.02 % in gunshot detection and 98.16 % in disregarding other sounds (i.e., non-gunshot).**

***Index Terms*—Acoustic signal processing; Gunshot detection; Neural networks; Parameter estimation.**

I. Introduction

Automatic gunshot detection in audio streams can help protect property or increase security. Although a gunfire sound can today be detected with a smartphone by both civilians [1] and police [2], it still makes sense to develop reliable detection methods for specific situations. Our research into gunshot detection has been initiated by the Save Elephants society to protect elephants against poachers in Central Africa. Some wild elephants today wear collars that are equipped with a GPS module to track the movement of elephant herds. The collar equipment, called "smart collar", will be complemented by a gunshot detection module that sends an alarm signal along with location information. In such a case, game rangers can act very promptly.

The widespread features in acoustic pattern recognition, such as mel-frequency cepstral coefficients (MFCCs) or linear prediction coefficients (LPCs), have their origins in speech recognition, but their application was extended to various acoustic scenes, including gunshot recognition [3], [4]. An introduction of these features may be found in [5]. In particular, the MFCCs are successfully used in many acoustic event recognitions (e.g., see [6], [7]). The authors of [4] used three MFCC variants as a neural network input for detecting two sounds - gunshot and glass breaking. The 3-set combination of MFCCs, i.e., static coefficients, delta coefficients, and delta-delta coefficients became the standard technique for speech recognition. Some researchers tested gunshot recognition using methods that were primarily developed for image recognition. For instance, the study in [8] describes the successful use of two-dimensional sound visualizations based on a spectrogram, MFCC, and a self-similarity matrix showing signal correlation. An overview of successful approaches developed by academics can be found in the proceedings resulting from the competition "Detection and Classification of Acoustic Scenes and Events (DCASE)" [9], a challenge that invited the authors to compare their sound detection algorithms where gunshots were the target sounds. A comprehensive review of gunshot detection technologies in urban environments, including a history of gunshot detection, can be found in [10].

Most of the developed methods and autonomous systems are intended for military, urban, or in-building applications. A surveillance system for automatic detection of gunshots in an indoor environment is proposed in [11]. In recent years, many commercial products for gunshot detection have been developed; for example, "Shooter Detection" [12] built into smartphones for various personal applications, "Boomerang" [13] installed on military patrol vehicles, and "ShotSpotter" [14] designed for urban use. Some practical aspects considering the implementation of ShotSpotter in an urban environment are discussed in [15]. Study in [16] describes the experience of using automatic gunshot detection in US cities.

There are very few studies dealing with gunshot detection developed to protect endangered animals, such as elephants or rhinos in the wild. Paper in [17] presents a prototype of an anti-poaching system built on elephants' collars - it is the same application as the goal of our work. The system is based on 10 features calculated from the shockwave. The authors did not disclose the overall accuracy of the detection. All the features used in [17] differ from our features defined in this paper.

The rest of the paper is organized as follows. The next

Manuscript received 8 February, 2021; accepted 6 May, 2021.

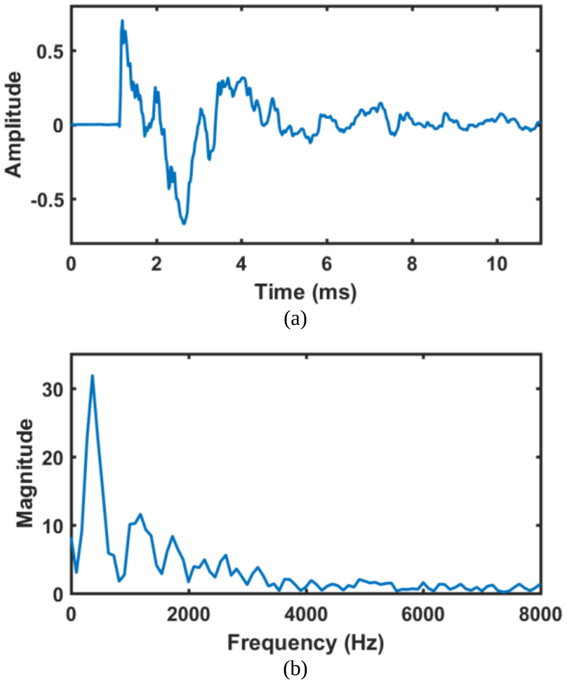
This research was supported by the Brno University of Technology Internal Grant Agency under Grant No. FEKT-S-20-6361.

section provides basic information about gunshot sounds. Section III introduces a combination of new efficient features developed to detect gunshot events in open space. The experiments carried out, including the evaluation of the results obtained, are presented in Section IV. Finally, Section V concludes the paper and outlines the continuation of work for the intended application.

II. GUNSHOT SOUND CHARACTERISTICS

The sound of a gunshot depends on the generating mechanisms and differs in detail according to the type of firearm, especially according to its caliber and barrel length. The sounds are naturally impulsive signals characterized by very high intensity and short duration (a few milliseconds). A typical gunshot wave consists of two parts: an initial high-intensity signal, which has a (usually) N-shaped waveform, and a subsequent ending phase with falling intensity. It therefore makes sense to analyse both parts separately in addition to the whole signal. Due to the psychoacoustic properties of the human hearing organs [18], [19], our subjective perception of gunshots is much longer than the millisecond duration of a purely physical sound signal. It should also be noted that the intensity of the gunshot drops non-linearly with increasing distance from the source (as well as the intensity of other sounds) by absorption in the air and spherical propagation. For rifles, the typical sound-pressure (SPL) level is around 160 dB (at a distance of 1 m from the barrel), the main frequency components cover the range of 250 Hz–450 Hz and the velocity of the projectile is between 800 m/s and 900 m/s.

A typical waveform of the gunshot and its corresponding spectrum is depicted in Fig. 1.



(a)

(b)

Fig. 1. Example of a typical gunshot sound signal measured for a Tikka hunting rifle: (a) waveform and (b) its spectrum.

In this case, the initial N-wave has a symmetrical oscillation. There are three clear local peaks in the spectrum in the range of 0 kHz–2 kHz with a dominant frequency below 500 Hz. Generally, the height and frequency of individual peaks differ for different types of firearms, and in some details they can be influenced by the current technical condition of the firearm. A similar spectral phenomenon is well known, e.g., in speech signal processing, where local peaks (so-called “formants”) in vowel spectra serve primarily to distinguish individual vowels [20], but their small changes may reflect the speaker’s state [21].

III. EFFECTIVE COMBINATION OF TWO FEATURE SETS

A key step in the classification of audio signals is the extraction of appropriate signal features that represent and distinguish the audio event of interest. To reliably detect gunshots from hunting weapons, we have created and combined two feature sets from different domains. The first one is based on standard sound features resulting in an optimized cepstral coefficient set. The second feature set includes new individual shooting-specific features derived directly from the shot waveform.

A. *Optimized Cepstral Coefficients*

First, the performance of standard sound features, such as MFCCs, LPCs, autocorrelation coefficients, and energy in frequency bands, was tested, and then the feature sets were subsequently tweaked for gunshot recognition. The best results were achieved with optimized MFCCs, which differ from standard MFCCs in two filtering parameters: the original triangular filters spaced in mel-frequency scale were replaced by uniform gamma-tone filters spaced linearly on the frequency axis [22]. The arrangement of the entire filter bank is shown in Fig. 2 (alternating red and blue colors are used for adjacent filters to increase readability). These features will be called “linear frequency cepstral coefficients” (LFCCs) to emphasize the linear distribution of filter boundaries.

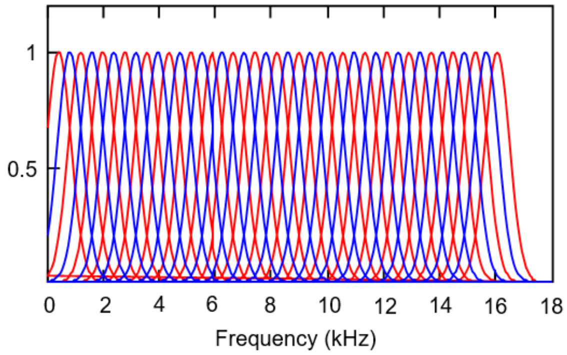


Fig. 2. Linearly distributed filter bank of 41 uniform gamma-tone filters applied to the calculation of 18 LFCCs.

The gamma-tone filter response synthesizes an impulse response from nerve cells in the auditory fiber. A gamma-tone filter is described in the time domain by the impulse response given by the product of the gamma envelope and the sinusoidal tone

$$g(t) = a t^{n-1} \exp(-2\pi b t) \cos(2\pi f_0 t + \varphi), \qquad (1)$$

where a, n, b, f_0, and φ are the peak value, filter order, bandwidth, characteristic (i.e., center) frequency, and initial phase, respectively [23]. In the frequency domain, the filter is approximately symmetric around f_0.

The optimum LFCC structure found for gunshot detection consists of 41 filters covering the frequency range of 0 kHz–16 kHz with an equidistant center frequency spacing of 390 Hz. Then, the set of the first 18 coefficients proved to be the most efficient coefficients. These coefficients were used in our experiments.

After filtering, the calculation of the coefficients continues further according to the rules of standard MFCC as follows

$$LFCC_m = \sum_{n=1}^{N} \log(S_n) \cos\left[\frac{\pi}{N} m(n-0.5)\right], \tag{2}$$

where $N = 41$ is the number of filters used and S_n stands for the output power of the n-th filter of the filter bank. The signal spectrum was calculated by the Fourier transform. In our calculation, the index m ranges from 1 to 18. The steps of the complete LFCC algorithm are graphically summarized in Fig. 3. First, the audio signal is segmented into short frames. Then, fast Fourier transform (FFT) is applied to the signal in each frame to get a short-term spectrum. The magnitude spectrum is filtered by the linear gamma-tone filter bank. The outputs of the bandpass filters are processed into a logarithmic value of energy. Finally, the LFCC coefficients are calculated using discrete cosine transform (DCT).

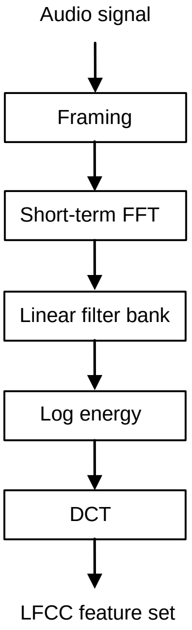


Fig. 3. Block diagram of LFCC set calculation.

The gamma-tone LFCCs seem to be more powerful than MFCCs in recognizing impulsive signals, such as gunshots, as the filter bank simulates a slightly different property of the auditory organ. The mel-frequency scale reflects overall auditory perception, and standard MFCCs based on this scale have been proposed to automatically recognize speech/speaker. A speech signal contains many small details in pronunciation. However, the gunshots are very short (< 10 ms). In addition, their acoustic signals are mechanically generated sounds.

B. *Shooting-Specific Features*

We were looking for new specific features derived directly from the gunshot acoustic waveform, which is usually clearly N-shaped at the beginning of the curve. The time-domain based feature set will be called “TDF”.

The significant TDFs applied herein are the time interval *T* between the dominant peaks (positive and negative) measured as the interval between the located points of maximum and minimum in the gunshot waveform, as well as the area *P* defined by the peak-to-peak curve and the horizontal time axis depicted as the filled area in Fig. 4. Both features *T* and *P* characterize the N-shape at the time of the onset of the wave, which distinguishes gunshots well from other sounds that may occur.

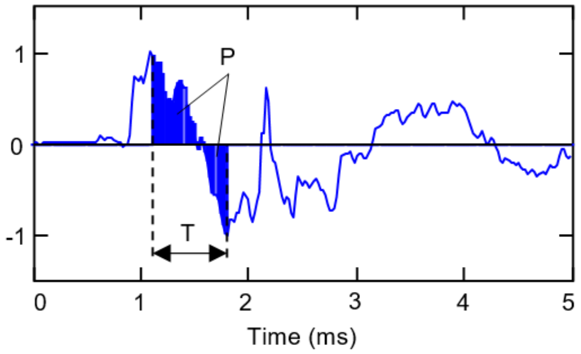
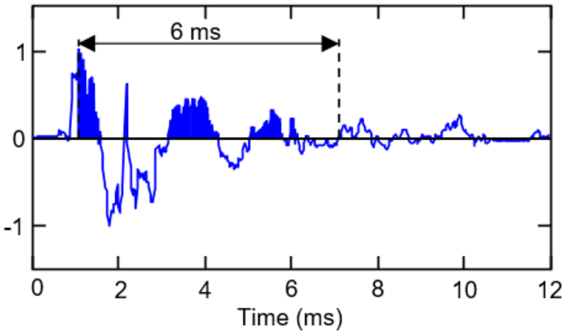


Fig. 4. Time interval *T* between two dominant peaks in the N-wave and area *P* (filled area) used as features.

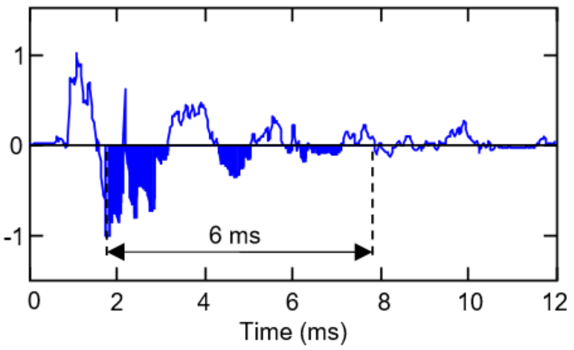
Other features are related to the overall decreasing part of the curve. With respect to damped irregular oscillations, the two areas were calculated separately for positive and negative amplitudes. Positive area *A* (above the horizontal axis) begins with the dominant positive peak and ends after a time of 6 ms. Negative area *B* (below the horizontal axis) begins with the dominant negative peak and ends after 6 ms too. This means that both areas correspond to a time period of 6 ms, but are offset from each other as can be seen in Fig. 5. In addition, the ratio of areas *A* and *B*

$$R = A/B \tag{3}$$

was considered. The features defined above are included in the TDF set. An overview structure of the TDF extraction is shown in Fig. 6.



(a)



(b)

Fig. 5. Determining of (a) positive area *A* (top) and (b) negative area *B* (bottom) in the decreasing part of the gunshot signal.

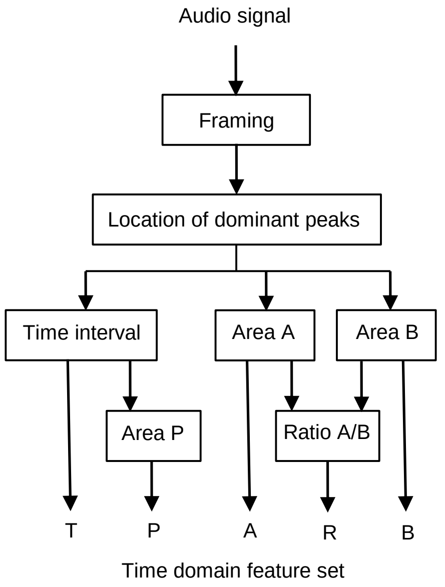


Fig. 6. Block diagram of individual TDF extraction.

C. Overall Feature Extraction

The overall process of feature extraction is shown in Fig. 7.

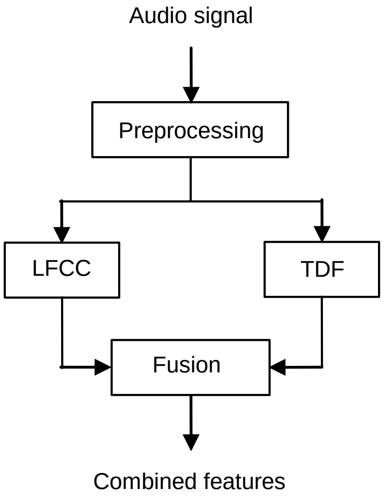


Fig. 7. Overall structure of feature extraction.

The first step in sound processing is the preprocessing of digital audio data. Preprocessing is used to prepare the input signal for reliable extraction of acoustic features. In our case, preprocessing involves two procedures: segmenting the signal into frames and normalizing the amplitude, so that all signal points are scaled in the range from −1 to +1. Regarding TDF, it is also important to remove the DC component (if an offset occurs). A rectangular window was used for framing all signals (gunshots and non-gunshots). The created frames have a length of 10 ms with an overlap of 50 %. This means that new features are calculated every 5 ms.

The feature sets LFCC and TDF are independent of each other, so the calculation can be implemented simultaneously in parallel processing. The obtained values are then combined into one vector of features. In the following gunshot detection, each signal frame is represented by a feature vector of 23 elements (18 LFCCs and 5 TDFs). Neither LFCC nor TDF alone outperform other feature sets used in audio event recognition. However, our experiments prove that their combination creates a very powerful feature vector for gunshot detection.

IV. EXPERIMENTS AND RESULTS

All experiments are based on real acoustic signals representing both gunshots and non-gunshots. The described calculations and approaches were implemented using the MATLAB programming environment on a desktop computer with a straightforward configuration. Due to the short duration of the gunshot (< 10 ms), the signal analysis is carried out on a one-frame basis, i.e., only one feature vector is extracted from each gunshot.

A. Used Data

Most of the sound recordings used in this study were from the GUDEON corpus [24], created specifically for research into gunshot detection in open nature. In addition, some signals were taken from two sources: Still North Media [25] for gunshots and Urban Sound Datasets [26] described in [27] for non-gunshots. The gunshot category includes 1500 gunshots from various hunting weapons such as AK-47, AR-15, Carl Gustaf m/45, Tikka T3, etc. Note that the AK-47 (7.62 mm × 39 mm) is not a typical hunting weapon, but it is often used by poachers in the territory of the intended application of gunshot detection in Central Africa. The non-gunshot category is divided into 6 different classes as follows: barking dog, sounds from elephants, sound of rain and storm, car horn, engine sounds, and human sounds (including short shouts). Each class contains 4000 individual sound frames, i.e., the non-gunshot category covers a total of 24000 sound frames. The sound classes were chosen due to the high frequency in which they appear around elephants. Some recordings in external databases originally had a different data bitrate, i.e., sampling frequencies, quantization levels, and one or two channels. We have converted the recordings with different parameters so that all signals used in our experiments were single-channel sounds, sampled at 44.1 kHz and quantized by 16 bits. All recordings were in WAV format.

B. Evaluation Criteria

To evaluate the performance of the features, we have used a fully connected neural network (NN) algorithm with two hidden layers of 20 neurons each. This architecture was chosen after simply searching the grid of from 1 to 3 hidden layers with 10, 20, and 30 neurons. More complex networks do not need to be considered due to the relatively low number of input features. The total dataset was divided into training, validation, and testing subsets. The training subset contains 600 gunshots and 6 × 600 non-gunshots (i.e., 600 sounds from each non-gunshot group) randomly selected, the validation subset contains 200 gunshots and 6 × 200 non-gunshots, and the testing subset contains 700 gunshots and 6 × 3200 non-gunshots. To limit the effect of a specific training subset, each subset was generated 5 times with a predefined random seed and the results presented were calculated as an average of 5 different training/testing phases.

Two appropriate event-based metrics were applied to the evaluation of gunshot detection performance using the developed features and neural networks, namely, the true positive rate (*TPR*) and the true negative rate (*TNR*), respectively, defined as follows:

$$TPR = \frac{TP}{TP + FN}, \tag{4}$$

$$TNR = \frac{TN}{TN + FP}, \tag{5}$$

where *TP* is the number of true positives (i.e., gunshots identified as gunshots) and *FN* is the number of false negatives (i.e., missed gunshot detection). Analogously, *TN* stands for the number of true negatives (i.e., non-gunshots identified as such) and *FP* stands for the number of false positives, also known as false alarm (i.e., number of non-gunshots identified as gunshots). In the overall evaluation, *TNR* was preferred for reliability measurement due to the prevailing non-gunshot sounds in continuous acoustic scene monitoring. Furthermore, frequent false alarms would dull the administrator’s attention.

C. Achieved Results

First of all, the power of LFCC and MFCC to detect gunshots from hunting weapons was tested. Each feature set was optimized separately for this purpose. LFCC parameters are described in Section III. The search for MFCC parameters resulted in the values as follows: frequency range of 0 kHz–22 kHz, linear scale from 0 Hz to 1000 Hz, and mel-scale for higher frequencies, triangular filter shapes, 28 filters, 12 coefficients. Table I shows a comparison of the results obtained using the optimized settings of the LFCC and MFCC algorithms. As can be seen, LFCC gives better results according to both *TPR* and *TNR* criteria.

Under the conditions described in the preceding section, various setups of 18 LFCCs and 5 TDFs were investigated and evaluated. The basis was the separate performance of the LFCC set and TDF set. Subsequently, the performance of the merged feature set LFCC + TDF (i.e., 23 features in total) was evaluated. Since TDFs are extracted from the time domain and their character is quite different from the LFCCs, low mutual information between the two feature sets is expected. Table II shows the results obtained in these tests with a single neural network. As can be seen, compared to LFCCs only, the combined set of LFCC + TDF performs significantly better in terms of *TPR*, but with a slight decrease in *TNR*.

In another test, the combination of LFCCs with TDFs was tested in an ensemble approach with a separate network trained for LFCC features and another one for TDF features. In this case, the recognition algorithms run in parallel in two branches, each resulting in a probability of the binary classification gunshot/non-gunshot. The final decision is then given by the sum of the probability scores from both separate networks. The results of this approach are shown in the last row of Table II. Overall, the ensemble network provides the best results in terms of both *TPR* and *TNR*.

The individual features of the TDF set have different potential for gunshot detection. In the investigation, each feature was added separately to the complete LFCC set and the change of performance was observed. Based on the performance improvement, the TDFs were sorted in descending order as follows: *T*, *A*, *R*, *B*, and *P*. Thereafter, the features from the TDF set were gradually added to the LFCCs in this order.

Table III summarizes the effect of increasing the number of added TDFs on the detection performance in terms of *TPR* for both single and ensemble neural networks. As can be seen, the inclusion of *T* and *A* in the ensemble network increases *TPR* by approximately 10 %. On the other hand, feature *B* has no significant benefit. Thus, in practical applications where the amount of data to be computed plays a role, the benefit of increased performance and the associated increase in computational costs should be taken into account.

TABLE I. COMPARISON OF MFCC AND LFCC IN TERMS OF TRUE POSITIVE RATE AND TRUE NEGATIVE RATE.

Feature set	Frequency range	TPR	TNR
MFCC 12 coefficients	0 kHz–22 kHz	79.4 %	85.1 %
LFCC 18 coefficients	0 kHz–16 kHz	83.86 %	98.13 %

TABLE II. COMPARISON OF PERFORMANCE OF THE INVESTIGATED FEATURE SETS IN TERMS OF TRUE POSITIVE RATE AND TRUE NEGATIVE RATE.

Feature set	Neural network	TPR (%)	TNR (%)
LFCC	single	83.86	98.13
TDF	single	72.43	87.24
LFCC + TDF	single	88.17	97.27
LFCC/TDF	ensemble	95.02	98.16

TABLE III. POSITIVE EFFECT OF ADDING TDF TO LFCC SET ON TRUE POSITIVE RATE (IN PERCENT).

Features	Single network	Ensemble network
LFCC	83.86	83.86
LFCC, T	84.02	92.54
LFCC, T, A	87.11	93.23
LFCC, T, A, R	86.43	94.18
LFCC, T, A, R, B	86.46	94.37
LFCC, T, A, R, B, P	88.17	95.02

Table IV shows a brief overview of the accuracy achieved by methods of mono-channel gunshot detection that have been published in recent years. In comparison to other methods, we achieved a high accuracy of 95.02 %

with a low number of 23 features. For example, the study in [28] reported the best accuracy of 96.10 % when applying 338 features. Moreover, we use a relatively simple neural network. In Table IV, SVM stands for Support Vector Machine [29], CNN stands for Convolutional Neural Network [30], and kNN for k-Nearest Neighbors [28].

TABLE IV. COMPARISON OF METHODS PUBLISHED IN PAPERS.

Study (Ref. No.)	**Year**	**Features and classifier**	**Environment**	**Accuracy (%)**
[29]	2020	spectral parameters; SVM	railway station	91.78
[30]	2020	MFCC-spectrogram, spectrum; CNN	open space	95.21
[31]	2020	spectrograms; SVM	open space	90.00
[11]	2021	energy bands, MFCC; SVM	indoor	94.97
[28]	2021	fractal pattern, statistics; kNN, SVM	open space	96.10
Proposed method	2021	LFCC, TDF; ensemble NN	open space	95.02

V. CONCLUSIONS

This paper presents a set of new features that have been developed for reliable acoustic detection of gunshots from hunting weapons in the wild. Our contribution to novelty in the field lies both in finding new types of features and in optimally combining individual features into an efficient feature vector. Based on real acoustic signals, a gunshot detection rate of 95.02 % was achieved. Very useful, especially for practical use, is also the ability of the proposed method to ignore almost all non-gunshots, i.e., 98.16 % of other environmental sounds.

In the intended application for the protection of elephants, an important aspect is not only the high reliability of gunshot detection, but also the low energy consumption because the implemented system will be powered by a self-sustainable energy source. In our next work, we will optimize the feature calculations and the classification approach to reduce the power consumption of the processor [32]. For this purpose, the study in [33] compares computational costs versus classification performance for different approaches of sound recognition.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] S. Tangkawanit and S. Kanprachar, “Spectral vector design for gunfire sound classification system with a smartphone using ANN”, in *Proc. of the International Symposium on Wireless Personal Multimedia Communications*, Chiang Rai (Thailand), 2018, pp. 421–426. DOI: 10.1109/WPMC.2018.8712930.

[2] P.-C. Lin and C.-Y. Wen, “Gunshot detection by STE and ZCR”, *Forensic Science Journal*, vol. 18, no. 1, pp. 35–46, 2019. DOI: 10.6593/FSJ.201912_18(1).0004.

[3] M. Hrabina and M. Sigmund, “Acoustical detection of gunshots”, in *Proc. of 25th International Conference Radioelektronika*, Pardubice (Czech Republic), 2015, pp. 150–153. DOI: 10.1109/RADIOELEK.2015.7128993.

[4] D. Conka and A. Cizmar, “Acoustic events processing with deep neural network”, in *Proc. of 29th International Conference Radioelektronika*, Pardubice (Czech Republic), 2019, pp. 1–4. DOI: 10.1109/RADIOELEK.2019.8733502.

[5] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*. London: Prentice Hall, 2011, ch. 8, ch. 9.

[6] L. Grama and C. Rusu, “Extending assisted audio capabilities of TIAGo service robot”, in *Proc. of 10th International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Timisoara (Romania), 2019, pp. 1–8. DOI: 10.1109/SPED.2019.8906635.

[7] S. Ntalampiras, “A novel holistic modeling approach for generalized sound recognition”, *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 185–188, 2013. DOI: 10.1109/LSP.2013.2237902.

[8] J. Bajzik, J. Prinosil, and D. Koniar, “Gunshot detection using convolutional neural networks”, in *Proc. of 24th International Conference Electronics*, Palanga (Lithuania), 2020, pp. 1–5. DOI: 10.1109/IEEECONF49502.2020.9141621.

[9] Detection and Classification of Acoustic Scenes and Events (competition website). [Online]. Available: http://www.cs.tut.fi/sgn/arg/dcase2017/

[10] J. R. Aguilar, “Gunshot detection systems in civilian law enforcement”, *Journal of the Audio Engineering Society*, vol. 63, no. 4, pp. 280–291, 2015. DOI: 10.17743/jaes.2015.0020.

[11] S. U. Rahman, A. Khan, S. Abbas, F. Alam, and N. Rashid, “Hybrid system for automatic detection of gunshots in indoor environment”, *Multimedia Tools and Applications*, vol. 80, pp. 4143–4153, 2021. DOI: 10.1007/s11042-020-09936-w.

[12] D. Welsh and N. Roy, “Smartphone-based mobile gunshot detection”, in *Proc. of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops*, Kona (USA), 2017, pp. 244–249. DOI: 10.1109/PERCOMW.2017.7917566.

[13] J. A. Mazurek *et al.*, “Boomerang mobile counter shooter detection system”, in *Proc. of SPIE - The International Society for Optical Engineering*, vol. 5778 (PART I), Orlando (USA), 2005, pp. 264–282. DOI: 10.1117/12.607616.

[14] ShotSpotter FAQ (information sheet), Shotspotter Inc., Newark, USA. [Online]. Available: https://www.shotspotter.com/wp-content/uploads/2018/08/FAQ_Aug_2018.pdf

[15] N. G. La Vigne, P. S. Thompson, D. S. Lawrence, and M. Goff, “Implementing gunshot detection technology”, Urban Institute, Washington DC (USA), 2019.

[16] D. S. Lawrence, N. G. La Vigne, M. Goff, and P. S. Thompson, “Lessons learned implementing gunshot detection technology: Results of a process evaluation in three major cities”, *Justice Evaluation Journal*, vol. 1, no. 2, pp. 109–129, 2018. DOI: 10.1080/24751979.2018.1548254.

[17] G. Kalmar, G. Wittemyer, P. Völgyesi, H. B. Rasmussen, M. Maroti, and A. Ledeczi, “Animal-borne anti-poaching system”, in *Proc. of 17th Int. Conf. on Mobile Systems, Applications, and Services*, Seoul (South Korea), 2019, pp. 91–102. DOI: 10.1145/3307334.3326080.

[18] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Berlin: Springer-Verlag, 2013, ch. 12.

[19] G. Devkota, “The psychoacoustic properties of sound: An introduction”, *Indian Journal of Scientific Research*, vol. 10, no. 1, pp. 215–221, 2019.

[20] R. D. Kent and H. K. Vorperian, “Static measurements of vowel formant frequencies and bandwidths: A review”, *Journal of Communication Disorders*, vol. 74, pp. 74–97, 2018. DOI: 10.1016/j.jcomdis.2018.05.004.

[21] M. Stanek and M. Sigmund, “Finding the most uniform changes in vowel polygon caused by psychological stress”, *Radioengineering Journal*, vol. 24, no. 2, pp. 604–609, 2015. DOI: 10.13164/re.2015.0604.

[22] M. Hrabina and M. Sigmund, “Optimization of mel-frequency cepstral coefficients for automatic gunshot detection”, unpublished.

[23] H. Park and Ch. D. Yoo, “CNN-based learnable gammatone filter bank and equal-loudness normalization for environmental sound classification”, *IEEE Signal Processing Letters*, vol. 27, pp. 411–415, 2020. DOI: 10.1109/LSP.2020.2975422.

[24] M. Hrabina and M. Sigmund, “Audio event database collected for gunshot detection in open nature (GUDEON)”, *Journal of the Audio Engineering Society*, vol. 67, nos. 1/2, pp. 54–59, 2019. DOI: 10.17743/jaes.2018.0075.

[25] *Still North Media (firearm sound libraries)*. [Online]. Available: https://www.stillnorthmedia.com/libraries

[26] *Urban Sound Datasets (urban sound libraries)*. [Online]. Available: https://urbansounddataset.weebly.com/

[27] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research”, in *Proc. of the 2014 ACM International Conference on Multimedia*, Orlando (USA), 2014, pp. 1041–1044. DOI: 10.1145/2647868.2655045.

[28] S. Dogan, “A new fractal H-tree pattern based gun model identification method using gunshot audios”, *Applied Acoustics*, vol. 177, art. no. 107916, 2021. DOI: 10.1016/j.apacoust.2021.107916.

[29] A. Suliman, B. Omarov, and Z. Dosbayev, “Detection of impulsive sounds in stream of audio signals”, in *Proc. of 8th International Conference on Information Technology and Multimedia (ICIMU)*, Selangor (Malaysia), 2020, pp. 283–287. DOI: 10.1109/ICIMU49871.2020.9243540.

[30] I. Papadimitriou, A. Vafeiadis, A. Lalas, K. Votis, and D. Tzovaras, “Audio-based event detection at different SNR settings using two-dimensional spectrogram magnitude representations”, *Electronics*, vol. 9, no. 10, p. 1593, 2020. DOI: 10.3390/electronics9101593.

[31] S. Raponi, I. Ali, and G. Oligeri, “Sound of guns: Digital forensics of gun audio samples meets artificial intelligence”, *arXiv*, 2020. arXiv: 2004.07948.

[32] T. Fryza and R. Mego, “Power consumption of multicore digital signal processor: Theoretical analysis and real applications”, in *Proc. of 23rd IEEE International Symposium on Industrial Electronics*, Istanbul (Turkey), 2014, pp. 1894–1898. DOI: 10.1109/ISIE.2014.6864904.

[33] S. Sigtia, A. M. Stark, S. Krstulovic, and M. D. Plumbley, “Automatic environmental sound recognition: Performance versus computational cost”, *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 11, pp. 2096–2107, 2016. DOI: 10.1109/TASLP.2016.2592698.