The 7th International Telecommunications Symposium (ITS 2010)

Gunshot detection in noisy environments

Izabela L. Freire and José A. Apolinário Jr.
Program of Defense Engineering
Military Institute of Engineering (IME)
Rio de Janeiro, Brazil
e-mails: izabela.lyon.freire@gmail.com and apolin@ime.eb.br

***Abstract*— Gunshot detection finds application in the fields of law enforcement, forensic science, and defense (military applications). The first task of a sniper detector, aiming to estimate the direction of arrival of a given gunshot, is to detect automatically the presence of this audio event. Since it is a typical on-line application where a fast response is of paramount importance, a non (computationally) expensive procedure is needed. In a recent work, a simple procedure based on the correlation of the audio signal against a template has proved its efficiency as a gunshot detection algorithm. In this paper, we extend its evaluation to a noisy environment and assess its performance, in gunshot recognition and gunshot detection tasks, comparing it to other more complex methods.**

Keywords*— gunshot detection; gunshot classification; impulsive signals; Hidden Markov Models; Mel-frequency cepstral coefficients; stable distributions; linear predictive coding.

I. Introduction

The military application of DOA (Direction of Arrival) estimation of gunshot audio signals (sometimes referred to as *Sniper Detector*) requires an automatic on-line detection of a gunshot.

Gunshot sounds are made up of distinct components, namely, the *muzzle blast*, lasting about 3 milliseconds, and caused by the explosion of the charge that propels the bullet; the sounds related to mechanical actions on the gun, like the trigger and hammer mechanism, or the expulsion of used cartridges; in some cases a *shock wave* from supersonic projectiles; and sounds related to environmental perturbations that can generally be caused by impulsive sounds, as the sound wave hits the ground or other solid surfaces [1].

Gunshot and impulsive sound detection and classification has relied on methods from the area of speech processing and, recently, [2] has applied a variety of such methods to the problem of detection and classification of impulsive sounds. These methods may rely on intensive computations, and as such may not be fit for real-time operation in military and law enforcement applications [3].

The problem of gunshot detection has been explored in the context of audio streams from movies [4] using dynamic programming and another work evaluates different algorithms in the task of detecting firearms gunshots [5]; the later reminds the reader that, being a gunshot signal similar to an impulsive signal, its spectral characteristics shall most likely provide information of the acoustic surroundings. The correlation detection algorithm has presented prospective good results [5]. This work evaluates the performance of a correlation measure against a template, by comparing it to more sophisticated methods such as HMM [6] working on LPC [7], MFCC [7], or the impulsivity parameter of stable distributions [8].

This paper is organized as follows: Section II deals with the extraction of a number of features of the impulsive signal while Section III presents the results of tasks of gunshot detection and gunshot recognition. Finally, conclusions are summarized in Section IV.

II. Features extraction from impulsive signals

In this section, we discuss the impulsive characteristic of a gunshot signal and briefly detail the features used in the classification and detection tasks.

A. Impulsive sounds

Impulsive sounds are generated with the sudden appearance of an air pressure wave. This can occur, for example, in the explosion of an inflated balloon, handclaps, gunshots, and plosive consonants such as [p], [t], and [k]. Such sounds are characterized by a sharp attack phase and highly non-stationary properties.

Characteristics of reverberation of an impulsive sound reflect properties of the environment, more so than non-impulsive signals [9], but the main concern here is with sounds occurring in non-reverberating environments, such as open fields.

B. Features Extraction

From each audio file, the following features were extracted:

- Correlation against audio files in set of templates,
- 8th order Linear Predictive Coding (LPC) coefficients [7],
- The first 13 Mel-frequency cepstral coefficients (MFCC) [7], and

- The impulsivity parameter from stable distributions [8].

MFCC coefficients were calculated using the Auditory Toolbox [10], LPC coefficients were calculated by Matlab®'s native implementation, and the impulsivity parameter was calculated in Fraclab [11], by McCulloch's method [12].

The features with temporal developments — LPC, MFCC, and impulsivity measure — were computed for 25 millisecond windows, with overlaps of 8 milliseconds with each adjacent window.

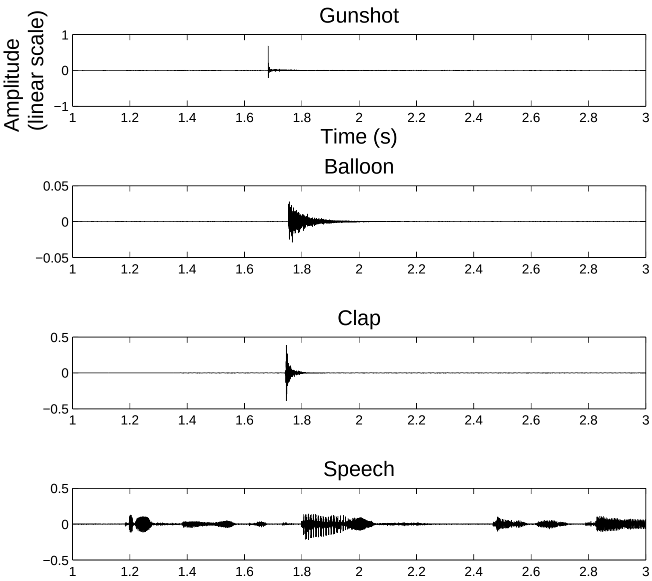

Fig. 1. Example waveforms in the database.

C. Classification Algorithms

While correlation is a single quantity relating two audio signals, the other features provide information about temporal variations in the sound.

Correlation against signals in a set of templates classifies a signal according to the class against which maximum correlation was obtained. Features that reflect temporal characteristics of the signal are fed to Hidden Markov Models (HMM) [6].

The HMM used 20 observable states and 8 hidden states. Probability distribution function of an observable state given a hidden state was modeled by a mixture of three Gaussians. Model creation and testing were performed in a cross-validation design. The HMM implementation is Kevin Murphy's *HMM Toolbox* [13].

D. Detection Algorithm

A *correlation threshold* method was used for detection of gunshots, according to a threshold for the correlation against a gunshot template. Correlations are calculated from z-score normalized signals.

III. Simulation Results

First we describe the audio database used in our experiments. Then we compare the correlations feature to other more sophisticated, albeit not exactly more efficient, features in clean and mildly noisy environments, in impulsive sound classification tasks. Based on the good results of the correlation feature, we decrease the SNR to more critical levels and asses its performance in this scenario. Then a correlation threshold method is proposed for gunshot detection.

A. The audio database

The database contains example sounds of handclapping, explosions of balloons, rifle and pistol shots, and speech. Figure 1 shows examples from our database.

All signals of our audio database were sampled at 44.1kHz. The length of each of them is 3 seconds. Gunshots were recorded at two open field sites, during Army training sessions carried out at CAEx (Centro de Avaliações do Exército) and CIAMPA (Centro de Instrução Almirante Milciades Portela Alves); speech was recorded in different environments; balloon explosions and handclaps were recorded in a laboratory environment (with medium level of reverberation).

In order to evaluate the performance of the classification algorithms at various SNR's, white noise was added to each audio signal. Measuring SNR for impulsive sounds can be done in various ways [2] so as to take into account the large and quick variations in signal potency, but here the most usual convention of defining SNR as $10\log(\sigma_s^2/\sigma_n^2)$ over the whole sound sample has been adopted, σ_s^2 and σ_n^2 being the signal and noise variances, respectively. White noise was added with the Matlab® function *awgn*.

B. Classification in clean and mildly noisy environments

In simulation environments without added noise, Hidden Markov Models working on LPC and MFCC coefficients perform with perfection over the database of choice. The impulsivity parameter from stable distributions, suggested in [8] as a good characterization of impulsive sounds, did not perform well in differentiating between the impulsive sound classes, but performed perfectly for speech versus impulsive sounds comparisons. This is an indication that it could be used to identify impulsive sounds in low-noise environments. Tables I and II show the results for all techniques in terms of confusion matrices for the original (clean) signals and for an SNR of 30dB, respectively.

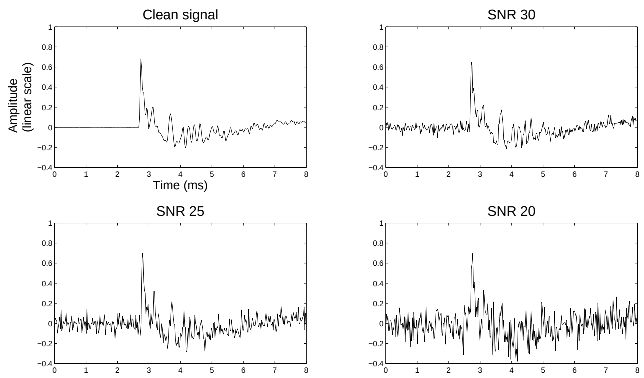

Fig. 2. Rifle shot template at various SNR's.

TABLE I
CONFUSION MATRICES FOR THE ORIGINAL (CLEAN) SIGNALS

Correlation	Gunshot	Balloon	Speech	Clap
Gunshot	**22**	0	0	0
Balloon	9	**13**	0	0
Speech	4	0	**18**	0
Clap	0	0	0	**22**

LPC	Gunshot	Balloon	Speech	Clap
Gunshot	**22**	0	0	0
Balloon	0	**22**	0	0
Speech	0	0	**22**	0
Clap	0	0	0	**22**

MFCC	Gunshot	Balloon	Speech	Clap
Gunshot	**22**	0	0	0
Balloon	0	**22**	0	0
Speech	0	0	**22**	0
Clap	0	0	0	**22**

Impulsivity	Gunshot	Balloon	Speech	Clap
Gunshot	**22**	0	0	0
Balloon	9	**5**	0	8
Speech	0	0	**22**	0
Clap	2	1	0	**19**

C. Classification in heavily noisy environments

Based on results from the 30dB SNR environment, the feature chosen for further experimentation in even noisier environments, with 20 dB SNR and 25 dB SNR, was the *correlation against templates from each class*. Results are reported as 2x2 *gunshot* x *non-gunshot* confusion matrices, in Table III. Experiments were run with the original four classes and the results are presented as 2 x 2 matrices.

At 25dB SNR, no gunshots were missed, though approximately one quarter of non-gunshots were false positives. At 20dB SNR, more elements from each class are classified as *non-gunshots*.

TABLE II
CONFUSION MATRICES FOR THE CASE OF $SNR = 30dB$

Correlation	Gunshot	Balloon	Speech	Clap
Gunshot	**22**	0	0	0
Balloon	9	**13**	0	0
Speech	4	0	**18**	0
Clap	3	0	0	**19**

LPC	Gunshot	Balloon	Speech	Clap
Gunshot	**13**	2	0	7
Balloon	1	**11**	0	10
Speech	0	0	**22**	0
Clap	3	3	0	**16**

MFCC	Gunshot	Balloon	Speech	Clap
Gunshot	**7**	14	1	0
Balloon	0	**21**	0	1
Speech	0	0	**22**	0
Clap	1	19	0	**2**

Impulsivity	Gunshot	Balloon	Speech	Clap
Gunshot	**14**	3	0	5
Balloon	0	**15**	0	7
Speech	1	6	**2**	13
Clap	1	6	0	**15**

TABLE III
CONFUSION MATRICES FOR THE CORRELATION METHOD IN HEAVILY NOISY SIGNALS

$20dB$	Gunshot	Not a gunshot
Gunshot	**20**	2
Not a gunshot	15	**51**

$25dB$	Gunshot	Not a gunshot
Gunshot	**22**	0
Not a gunshot	18	**48**

D. Gunshot detection

The good performance of the correlation measure in classification tasks, added to its low computational cost and robustness against noise suggests its use as a gunshot detection feature. The proposed gunshot detection method thus compares the signal only against a gunshot template and uses a threshold value to determine whether it is or not a gunshot. Figure 3 depicts histograms of correlation of each of the classes with a gunshot template (a rifle at a distance of 31m), at different SNR's. Varying the decision threshold, ROC curves are obtained and shown in Figure 4.

IV. CONCLUSIONS

The correlation against templates is not only a computationally cheap procedure but also displays comparable and even better performance than state-of-the-art algorithms adapted from the field of speech signal processing, especially in conditions of high environmental noise, in impulsive sound classification tasks. Based on these results,

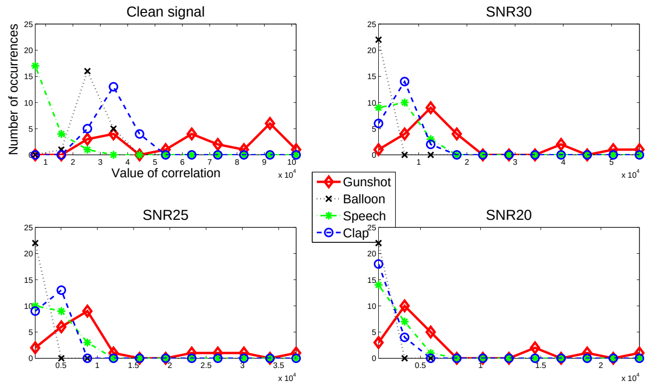

Fig. 3. Histogram of correlations between various classes and the rifle shot template, at different SNR's.

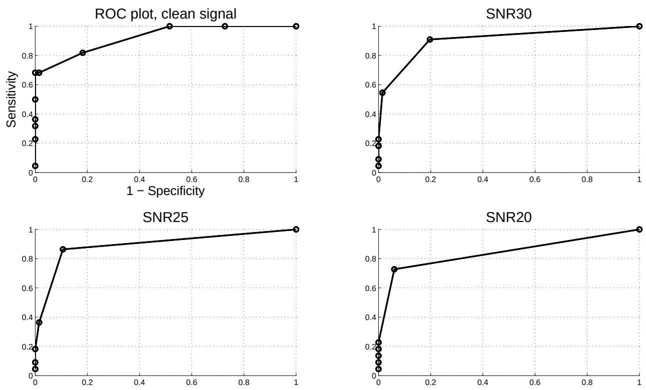

Fig. 4. ROC curves for the correlation threshold method with original (clean) signals and SNR's 30, 25 and 20 dB.

this paper proposes its use as a gunshot classification and detection feature.

Natural next steps are building a template database for gunshot detection and employing de-noising techniques prior to the correlation methods.

Acknowledgments

The authors want to thank the Brazilian agencies CAPES, CNPq, and FAPERJ for partial funding of this work.

References

[1] R. C. Maher, "Modeling and signal processing of acoustic gunshot recordings," in Proc. IEEE Signal Processing Society 12th DSP Workshop, Jackson Lake, USA, September 2006.

[2] A. Dufaux, "Detection and Recognition of Impulsive Sound Signals," D. Sc. thesis, Institute of Microtechnology. University of Neuchâtel, Neuchâtel, Switzerland, 2001.

[3] L. G. Mazerolle, "Random gunfire problems and gunshot detection systems," 1999 (Last updated: 8 June 2005).

[4] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis, "Gunshot detection in audio streams from movies by means of dynamic programming and Baysean networks," in Proc. ICASSP 2008, pp. 21–24, 2008.

[5] A. Chacón-Rodríguez and P. Julian, "Evaluation of gunshot detection algorithms," in Proc. of the Argentine School of Micro-Nanoelectronics, Technology and Applications (EAMTA), Buenso Aires, Argentina, September 2008.

[6] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall Signal Processing Series, 1993.

[7] X. Huang, A. Acero, and H-W. Hon, *Spoken Language Processing: a guide to theory, algorithms, and system development*, Prentice Hall, 2001.

[8] P. P. Pokharel, W. Liu, and J. C. Principe, "A low complexity robust detector in impulsive noise," *Signal Processing*, vol. 89, no. 10, pp. 1902–1909, October 2009. doi:10.1016/j.sigpro.2009.03.027.

[9] D. Sumarac-Pavlovic, et al., "A simple impulse sound source for measurements in room acoustics" *Applied Acoustics*, vol. 69, no. 4, pp. 378-383, April 2008.

[10] M. Slaney, "Auditory Toolbox," Version 2, Technical Report Nr 1998-010, Interval Research Corporation (USA). [Online] Available: http://cobweb.ecn.purdue.edu/∼malcolm/interval/1998-010/. [Accessed: May 2010].

[11] Software FRACLAB: a fractal analysis toolbox for signal and image processing. INRIA (France). [Online] Available: http://fraclab.saclay.inria.fr/. [Accessed: May 2010].

[12] S. Bates and S. F. McLaughlin, "The estimation of stable distribution parameters from teletrafic data," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 865–870, March 2000. doi:10.1109/HOST.1997.613553.

[13] K. Murphy, "Hidden Markov Model (HMM) Toolbox for Matlab," 1998 (Last updated: 8 June 2005). [Online] Available: http://www.cs.ubc.ca/∼murphyk/Software/HMM/hmm.html/. [Accessed: May 2010].