GUNSHOT SIGNAL ENHANCEMENT FOR DOA ESTIMATION AND WEAPON RECOGNITION

Ângelo M.C.R. Borzino†*, José A. Apolinário Jr.*†*, Marcello L.R. de Campos*‡*, and Carla L. Pagliari*†

† Military Institute of Engineering (IME)
Program of Defense Engineering
Rio de Janeiro, RJ – Brazil

‡ Federal University of Rio de Janeiro (UFRJ)
Program of Electrical Engineering (COPPE/PEE)
Rio de Janeiro, RJ – Brazil

E-mails: amborzino@hotmail.com, apolin@ime.eb.br, mcampos@ieee.org, and carla@ime.eb.br

ABSTRACT

This paper proposes a deconvolution technique for gunshot signals aiming at improving direction of arrival estimation and weapon recognition. When dealing with field recorded signals, reflections degrade the performance of these tasks and a signal enhancement technique is required. Our scheme improves a gunshot signal by delaying and summing its reflections. Conventional blind deconvolution schemes are not reliable when applied to impulsive signals. While other techniques impose restrictions on the signal in order to ensure stability, the one presented herein can be used without such limitations. The results of the proposed technique were tested with real gunshot signals and both applications performed well.

Index Terms— Signal deconvolution, gunshot signal, direction of arrival estimation, weapon recognition.

1. INTRODUCTION

The study of gunshot signals has become important as actions involving firearms are of increasing concern to police and defense forces, among other institutions. The first work related to the propagation of ballistic waves [1] dates from 1946. Issues related to the physics of the sound propagation [2] appeared in 1971. Two topics are of interest: weapon recognition and Direction of Arrival (DOA) estimation. The former can be a deciding factor in a criminal investigation, as knowing automatically which armament a shooter uses may give knowledge of its identity, help a forensic analyst solve a crime, or aid to find a weapon stolen from a government agency. The later may be useful to determine sniper localization.

A typical gunshot signal is impulsive and consists of two characteristic waves: the muzzle blast (MB) and the shockwave (SW). The first component is a consequence of the explosion of the charge in the gun barrel, lasts 3 to 5 milliseconds and propagates through the air at the speed of sound [3]. The later is due to the dispersion of air molecules caused by motion of the projectile when traveling at supersonic speed and usually arrives first in a microphone when it is located in the shockwave field of view. The shockwave lasts typically 0.3 to 0.5 milliseconds [3]. Fig. 1 shows a gunshot signal originated from an M964 Light Automatic rifle (FAL), around 300 meters far from the recording position. The high frequency portion of the signal on the left side of this figure refers to the shockwave and the following one (lower frequency) to the muzzle blast. The MB component is usually more attenuated whenever the microphone is closer to the bullet trajectory than to the shooter.

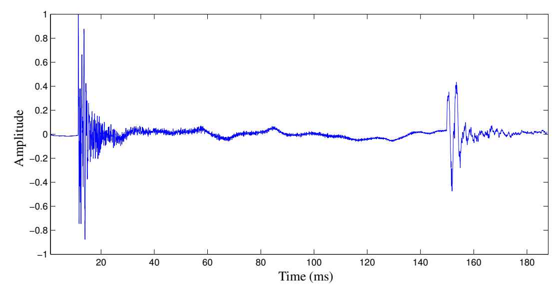


Fig. 1. SW and MB originated from a rifle.

Reflections on the ground surface or other obstacles are a common problem found in these signals (SW and MB components). They degrade both DOA estimation and weapon recognition, since the signals that reach the microphone (or microphone array) result from the convolution of the clean gunshot signal with the impulse response of the acoustic path, plus environmental noise.

There are some blind deconvolution techniques found in the technical literature; and many algorithms related to them (e.g. Bussgang-type ones) have been designed to extract digital communications signals corrupted by intersymbol interference. However, they generally fail when applied to impulsive signals [4]. For this class of signals, modified Bussgang algorithms were proposed in [4]. In [5], different objective functions (kurtosis and skewness) were compared for blind

Authors thank Brazilian Agencies CAPES (Project Prodefesa, Process no. 23038.009094/2013-83), CNPq, and FINEP for partial funding of this work.

deconvolution of impulsive signals. [6] uses the greatest common divisor (GCD) of the Z-transform of the measured signals, which, under some assumptions, is equal to the excitation impulsive signal. While [4] imposes statistical conditions to the signals in order to guarantee algorithm stability and the multichannel technique presented in [6] assumes that the Z-transforms of the channels do not share any root common to all of them, our scheme does not impose any restriction or statistical condition to the gunshot signal. It does not make any assumption other then the widely known information about maximum duration of gunshot components (SW and MB).

This work proposes a deconvolution scheme based on delaying and summing the reflections of a gunshot signal as in a rake receiver [7]. Similar technique was presented in [8] for improving gunshot detection. Our goal herein is to enhance the signal to improve DOA estimation and weapon recognition. After autocorrelation, or cross-correlation with a reference signal, peaks above a certain threshold are detected and the distances between them used to shift the gunshot signal, aligning its copies. The paper is organized as follows: Section 2 describes the proposed scheme for gunshot signal enhancement; Sections 3 and 4 explain the improvement obtained with the enhanced signal in DOA estimation and in weapon recognition, respectively; finally, Section 5 concludes the work. The proposed scheme does not require human interference.

2. PROPOSED GUNSHOT SIGNAL ENHANCEMENT

In our scheme, we first need to locate the reflections (copies) and this is carried out by means of correlation functions. Fig. 2 shows a SW signal originated from a .308 IMBEL AGLC (Sniper) Rifle, recorded around 300 meters far from the shooter position. In this figure, we also observe its autocorrelation function, a reference SW signal (obtained by averaging five AGLC shockwaves) and the cross-correlation between the original signal and this reference. As seen in this figure, SW signals resemble letter “N” and they are usually referred to as “N” waves. In this case, we can distinguish a peak on the left side of the original signal and its replica some samples later. This replica is certainly a reflection of the direct wave and we want to eliminate it as well as any other that might exist. The reference signal could be used if we knew that the shot came from a given rifle (AGLC, in this example).

From Fig. 2, we can observe that distances (in number of samples) between the two highest peaks (marked with circles) in the autocorrelation (AC) and cross-correlation (CC) are almost the same. Furthermore, these distances are nearly the same between the two peaks in the original signal. We assume that the armament is unknown and therefore focus our attention in the autocorrelation function.

After several tests, we have determined a threshold in the autocorrelation (red line in the Fig. 2) of 0.3 times the maxi-

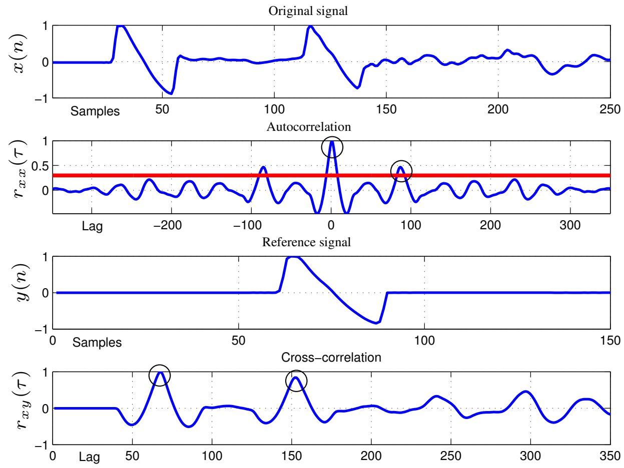


Fig. 2. SW signal $x(n)$ from an AGLC shot: autocorrelation $r_{xx}(\tau)$ and cross-correlation $r_{xy}(\tau)$ with a reference signal $y(n)$. The sample rate is 96 kHz.

mum peak (which, in our example, is normalized to one). The number N of peaks greater than 0.3 is the number of signals which will be shifted according to the distance between the AC peaks, cut on both sides, weighted according to the AC peaks, and added.

The following procedure was used to cut each of the N copies (direct wave plus $N-1$ reflections). Let $D_{SW} = 0.5$ ms be the maximum duration of SW signals, $D_{MB} = 5$ ms the maximum duration of MB signals, and D_i the distance between AC peaks $i+1$ and i.

1. If SW signal:
   (a) We cut $\frac{D_{SW}}{8}$ from the first peak to the left.
   (b) For each i^{th} peak, i from 1 to $N-1$:
      i. If $D_i < D_{SW}$, we cut $\frac{7D_i}{8}$ from peak i to the right to complete copy i and $\frac{D_i}{8}$ to the left of peak $i+1$ to form copy $i+1$.
      ii. If $D_i > D_{SW}$, we cut $\frac{7D_{SW}}{8}$ from peak i to the right to complete copy i and $\frac{D_{SW}}{8}$ to the left of peak $i+1$ to form copy $i+1$.
   (c) We cut $\frac{7D_{SW}}{8}$ from the last peak to the right.
2. If MB signal:
   (a) We cut $\frac{D_{MB}}{4}$ from the first peak to the left.
   (b) For each i^{th} peak, i from 1 to $N-1$:
      i. If $D_i < D_{MB}$, we cut $\frac{3D_i}{4}$ from peak i to the right to complete copy i and $\frac{D_i}{4}$ to the left of peak $i+1$ to form copy $i+1$.
      ii. If $D_i > D_{MB}$, we cut $\frac{3D_{MB}}{4}$ from peak i to the right to complete copy i and $\frac{D_{MB}}{4}$ to the left of peak $i+1$ to form copy $i+1$.

(c) We cut $\frac{3D_{MB}}{4}$ from the last peak to the right.

As seen from this procedure, the choice of how much signal we cut from the peak to the left or to the right depends on which component (SW or MB) we are dealing with, for they have different rise-times [9]. Note that the mentioned fractions, $\frac{1}{8}$ and $\frac{1}{4}$, can be used for any armament, since all SW and MB components we observed obey these proportions.

Fig. 3 shows the resulting (deconvolved) signal obtained from autocorrelation of the original signal in Fig. 2. It also shows, in bold, the two portions (direct wave and reflection, properly aligned with direct wave) that are summed. In this case, $D > D_{SW}$, since the reflection is far from the direct wave.

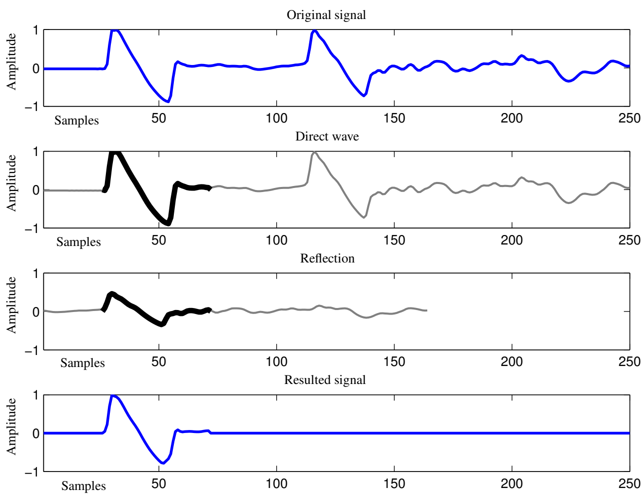
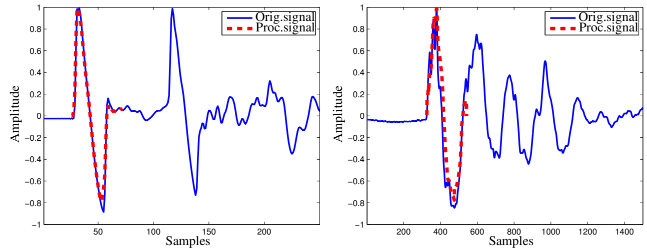


Fig. 3. Deconvolving a shockwave using autocorrelation. The sample rate is 96 kHz.

Fig. 4 (a) plots the original signal (in blue) and the deconvolved signal (in red). It is worth mentioning that all procedures are carried out without user interference and that there is no need to make assumptions in the signal as in [6]; there are also no restrictions as those imposed in [4]. Fig. 4 (b) shows the result when we applied the same technique (deconvolution with autocorrelation) to a 5.56 mm Tavor Assault Rifle (556) muzzle blast. We can observe that the resulting signal also resembles the direct wave of the original one. In the following two sections we shall use this signal enhancement method on two important applications. All test signals took place in open air shooting site.

3. IMPROVING GUNSHOT DOA ESTIMATION

There are several algorithms [10] that can be employed in DOA estimation including "Delay and Sum" (also known as *beamforming*), "Capon," "MUSIC," and "Generalized Cross Correlation (GCC)" [11]. The last one can be employed with wide-band signals (such as a gunshot signal), reason why it is briefly described in the following. A subclass of GCC,



(a) Shockwave signal (AGLC). (b) Muzzle blast signal (556).

Fig. 4. Deconvolving the components of a gunshot signal. The sample rate is 96 kHz.

used in this work, is the PHAT (Phase Transform), in which each component of the cross-spectrum phase is weighted equally [11].

The direction of arrival can be characterized by two angles [12]: ϕ, the azimuth, and θ, the complement of the elevation angle. Fig. 5 shows the microphone array used in our experiments, as well as angles ϕ and θ.

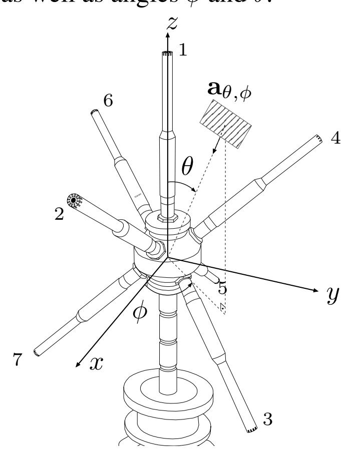


Fig. 5. Microphone array and angles of interest.

The unit vector in the direction of the wave propagation, $\mathbf{a}_{\theta,\phi}$, is given as

$$\mathbf{a}_{\theta,\phi} = \begin{bmatrix} -\sin\theta\cos\phi \\ -\sin\theta\sin\phi \\ -\cos\theta \end{bmatrix}. \tag{1}$$

Considering an array with M microphones, we obtain a total of $\frac{M(M-1)}{2}$ possible pairs of microphones and an equal number of cross-correlations which will be used in the DOA estimation procedure. Let τ_{ij} be the time difference of arrival (TDOA), in number of samples, between microphones i and j; it may be estimated from the cross-correlation between their signals as in

$$\tau_{ij} = arg\,max\,\hat{r}_{x_i x_j}(\tau), \tag{2}$$

where, $x_i(n)$ and $x_j(n)$ are the signals arriving at the i^{th} and j^{th} microphones, respectively, and $r_{x_i x_j}(\tau)$ their cross-correlation, estimated as $\hat{r}_{x_i x_j}(\tau) = \sum_{-\infty}^{\infty} x_i(n)x_j(n-\tau)$.

This cross-correlation corresponds to the convolution $\hat{r}_{x_i x_j}(\tau) = x_i(\tau) * x_j(-\tau)$, and is usually computed as the inverse Fourier transform of the cross-power spectrum density (CPSD) $X_i(e^{j\omega})X_j(e^{-j\omega})$. In the case of PHAT GCC, the CPSD is normalized by its absolute value prior to the inverse Fourier transform [11].

Defining the TDOA in time unit as $\overline{\tau}_{ij} = \frac{\tau_{ij}}{f_s}$, f_s being the sampling frequency, we know that it corresponds to the time the sound travels from microphone i to microfone j, a distance given by d_{ij} which also corresponds to

$$d_{ij} = \mathbf{a}_{\theta,\phi}^T \mathbf{p}_i - \mathbf{a}_{\theta,\phi}^T \mathbf{p}_j, \tag{3}$$

where $\mathbf{p}_i$ and $\mathbf{p}_j$ are the microphones coordinates. Therefore, we can also write

$$\overline{\tau}_{ij} = \frac{d_{ij}}{v_{sound}} = \frac{\mathbf{a}_{\theta,\phi}^T \mathbf{p}_i - \mathbf{a}_{\theta,\phi}^T \mathbf{p}_j}{v_{sound}} = \mathbf{a}_{\theta,\phi}^T \Delta\overline{\mathbf{p}}_{ij}, \tag{4}$$

where $\Delta\overline{\mathbf{p}}_{ij} = \frac{\mathbf{p}_i - \mathbf{p}_j}{v_{sound}}$. If we define the cost function

$$\xi = (\overline{\tau}_{12} - \Delta\overline{\mathbf{p}}_{12}^T \mathbf{a}_{\theta,\phi})^2 + ... + (\overline{\tau}_{(M\text{-}1)M} - \Delta\overline{\mathbf{p}}_{(M\text{-}1)M}^T \mathbf{a}_{\theta,\phi})^2, \tag{5}$$

we can find a DOA estimation by taking the gradient of this function with respect to $\mathbf{a}_{\theta,\phi}$ and equating to zero. The result is $\mathbf{a}_{DOA} = \mathbf{R}^{-1}\mathbf{p}$, where

$$\mathbf{R} = \Delta\overline{\mathbf{p}}_{12}\Delta\overline{\mathbf{p}}_{12}^T + ... + \Delta\overline{\mathbf{p}}_{(M\text{-}1)M}\Delta\overline{\mathbf{p}}_{(M\text{-}1)M}^T, \text{ and} \tag{6}$$

$$\mathbf{p} = \overline{\tau}_{12}\Delta\overline{\mathbf{p}}_{12} + ... + \overline{\tau}_{(M\text{-}1)M}\Delta\overline{\mathbf{p}}_{(M\text{-}1)M}. \tag{7}$$

Finally, taking the elements of $\mathbf{a}_{DOA} = [a_x \;\; a_y \;\; a_z]$, the horizontal angle (azimuth) is given by $\phi = -cos^{-1} a_z$ and the vertical angle (complement of the elevation) is given by $\theta = -tan^{-1} \frac{a_y}{a_x}$.

We applied this procedure to determine the DOA of 12 muzzle blast signals from FAL, AGLC, and .50 Browning Machine Gun (.50) originated all in the same position, approximately 300 m far from the microphone array, with and without the deconvolution technique. Fig. 6 shows the azimuth and elevation; it is clear from this figure that the elevation is less scattered with deconvolution, most probably due to the reflection on the ground surface, typical for a sniper scenario, being eliminated or drastically reduced.

We repeated the procedure to determine the DOA of 12 muzzle blast signals from AGLC, 556, and .50 originated from a distance around 400 m. Fig. 7 shows the results where it can be observed that, in this case, both elevations and azimuths are more concentrated after deconvolution. The different results in some DOA estimation without deconvolution is due not only to the reflections but also to the strong noise corrupting the signal. This occurs such that when cross-correlations are calculated, "false" peaks appear. When it happens, large errors may occur. When the deconvolution technique is applied, "false" peaks tend to disappear and the results become more concentrated. Note that azimuths vary from -180^o to 180^o; but, to improve visual information, we shifted some azimuths by -360^o (those below -180^o).

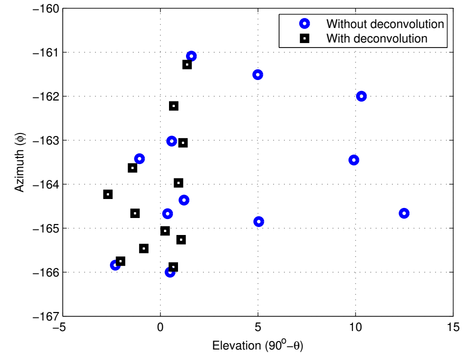


Fig. 6. DOA for gunshots from a distance of 300 m.

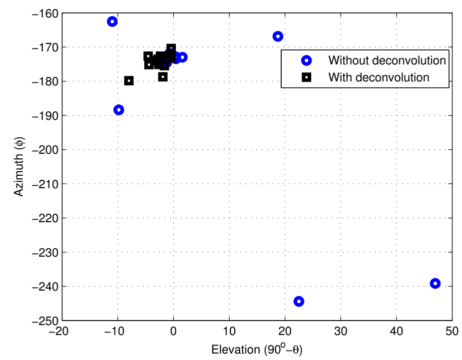


Fig. 7. DOA for gunshots from a distance of 400 m.

4. IMPROVING WEAPON RECOGNITION

It is known that well-trained ears are able to recognize the type of a weapon from listening to the sound of firing. Herein, we use a machine learning approach to armament classification. The weapon recognition technique described in the following is not intended to achieve definitive results but merely to show that deconvolution improves its performance.

We start with the spectrogram of a muzzle blast, quantized in gray levels and, using it as an image, and extracting textural features [13] as in a typical pattern recognition scenario. The features are fed to a Feedforward Neural Network classifier [14] to enable weapon recognition. We have used gunshot signals from .50, 556, AGLC, and FAL. Fig. 8 shows an example of the waveform and its corresponding spectrogram, in gray levels, of a FAL muzzle blast.

For each signal, a Hamming window with 256 samples (or 2.67 ms) was applied and the Fast Fourier Transform (FFT) with 1024 points calculated. The superposition of the frames was set to 96.875% (window skips 8 samples). From the resulting image, 12 attributes were obtained, all based on the image texture of the spectrogram, quantized with 256 levels of gray. From the 12 features, 7 (correlation, contrast, energy, entropy, inverse differential moment, homogeneity, and dissimilarity) were obtained from the *co-occurrency matrix* [15] and the other 5 (short run emphasis, long run emphasis, gray level distribution, run percentage, and long run low gray level

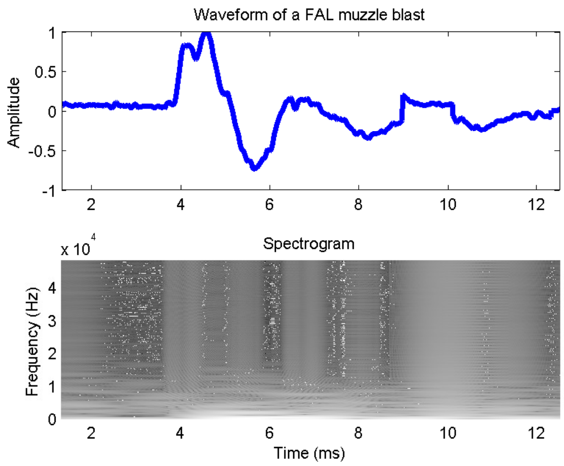


Fig. 8. Spectrogram of the muzzle blast from a FAL, an example of a image from which texture features are selected.

emphasis) from the *Run-Length Matrix* (RLM) [15], [13].

To classify the armament, we have used the Neural Pattern Recognition Tool, a Matlab® toolbox, with 2 layers: 20 neurons in the intermediate layer and 4 neurons (4 types of weapons or classes) in the last layer. 35 signals of each armament were used, from which 20 for training and 15 for testing. All gunshots were fired from the same position (300 m) for this experiment.

Table 1 shows the confusion matrix of the weapon classification carried out with signals without and with deconvolution. We can observe that the deconvolution scheme proposed in this work improved the recognition rate: the percentage of correct classification raised more than 10%. Signals originated from .50 Browning Machine Gun were all classified correctly in the two cases, which was expected, for it has a distinctive firing sound.

Table 1. Confusion matrices for the classification of four armaments without deconvolution (classification rate of 70%) and with the proposed scheme (81.7%)

WEAPON (CLASS)	Original signals				Deconvolved signals			
	FAL	**.50**	**556**	**AGLC**	**FAL**	**.50**	**556**	**AGLC**
FAL	12	0	7	5	13	0	6	1
.50	0	15	0	0	0	15	0	0
556	3	0	8	3	1	0	9	2
AGLC	0	0	0	7	1	0	0	12

5. CONCLUSION

A deconvolution scheme for gunshot signals based on a delay and sum approach was presented. The technique employs autocorrelation (or cross-correlation with a reference signal) and does not impose any restriction to the gunshot signal. The procedure reduces, or eliminates, any reflections and improved results were obtained in two applications: DOA estimation and weapon recognition.

It is expected that even better results are obtained with higher sampling rates or interpolating the autocorrelation when finding its peaks. For gunshots with low signal-to-noise ratio (long distances), the use of a noise reduction technique prior to deconvolution is topic of current research.

REFERENCES

[1] J. W. M. Dumond, E. R. Cohen, W. K. H. Panofsky, and E. Deeds, "A determination of the wave forms and laws of propagation and dissipation of ballistic shock waves," in *The Journal of the Acoustical Society of America*, 1946, vol. 18(1), pp. 93–118.

[2] A. D. Pierce, "Statistical theory of atmospheric turbulance effects on sonic-boom rise times," in *The Journal of the Acoustical Society of America*, 1971, vol. 49(3B), pp. 906–924.

[3] R. C. Maher, "Modeling and signal processing of acoustic gunshot recordings," in *Proc. IEEE Signal Processing Society 12th DSP Workshop*, 2006, vol. 59, pp. 257–261.

[4] H. Mathis and S. C. Douglas, "Bussgang blind deconvolution for impulsive signals," in *IEEE Transactions on Signal Processing*, 2003, vol. 51, pp. 1905–1915.

[5] P. Pääjärvi and LeBlanc J. P., "Skewness maximization for impulsive sources in blind deconvolution," in *Proceedings of the 6th Nordic Signal Processing Symposium*, 2004.

[6] P. Lecumberri, M. Gómes, and A. Carlosena, "Multichannel blind deconvolution of transient impulsive signals," in *Proc. of the 2006 IEEE Instrumentation and Measurement Technology Conference (IMTC06)*, 2006.

[7] R. Price and P. E. Green, "A communication technique for multipath channels," in *Proc. of the IRE*, 1958, vol. 46, pp. 555–570.

[8] M. Smith, S. Buscemi, and D. John Xu, "Gunshot detection system for JTRS radios," in *The 2010 Military Communications Conference*, 2010.

[9] W. S. Levine, *The Control Handbook*, CRC Press LLC - IEEE Press, 1996.

[10] H. L. V. Trees, *Optimum Array processing - Part IV - Detection, Estimation, and Modulation Theory*, Wiley-Interscience, 2002.

[11] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1976, vol. ASSP-24, pp. 320–327.

[12] I. L. Freire and J. A. Apolinário Jr., "DOA of gunshot signals in a spatial microphone array: performance of the interpolated generalized cross-correlation method," in *Proc. 5th Argentine School of Micro-Nanoelectronics, Technology and Applications*, 2011.

[13] M. M. Galloway, "Texture analysis using gray lavel run lengths," in *Computer Graphics and Image Processing*, 1975, vol. 4, pp. 172–179.

[14] S. Haykin, *Neural Networks A Comprehensive Foundation*, Prentice Hall, 1999.

[15] S. Selvarajah and S. R. Kodituwakku, "Analysis and comparison of texture features for content based image retrieval," in *International Journal of Latest Trends in Computing*, 2011, vol. 2.