

# Impulsive Sound Detection Directly in Sigma-Delta Domain

Igor D. dos S. MIRANDA, Antonio C. de C. LIMA

*Department of Industrial Engineering Postgraduate  
Federal University of Bahia*

2 Aristides Novis Street, 6th Floor, Federação, Salvador, Bahia 40.210-630, Brazil  
e-mail: engigordantas@gmail.com, acdcl@ufba.br

(received September 26, 2016; accepted February 1, 2017)

Recent implementations of Sigma-Delta ( $\Sigma\Delta$ ) converters have achieved low cost, low power consumption, and high integration while maintaining resolution as high as in Nyquist-rate converters. However, its usage implies demodulating the source signal delivered from  $\Sigma\Delta$  modulation to Pulse-Code Modulation (PCM) on a pre-processing stage. This work proposes an algorithm based on Discrete Cosine Transform for impulsive signal detection to be applied directly on a modulated  $\Sigma\Delta$  bitstream, targeting to reduce computational cost in acoustic event detection applications such as gunshot recognition systems. From pre-recorded impulsive sounds in  $\Sigma\Delta$  format, it has been shown that the new method presents a similar error rate in comparison with traditional energy-based approaches in PCM, meanwhile, it reduces significantly the number of operations per unit time.

**Keywords:** impulsive signal detection; sigma-delta modulation; discrete cosine transform.

## 1. Introduction

Impulsive sound signals is a category of audio signals characterised by an abrupt rise in amplitude, immediately followed by an under-damped motion or exponential decay, as shown in Fig. 1. Claps, glass breaking, hamming, and door slams are examples of everyday impulsive sound sources. When an impulsive signal

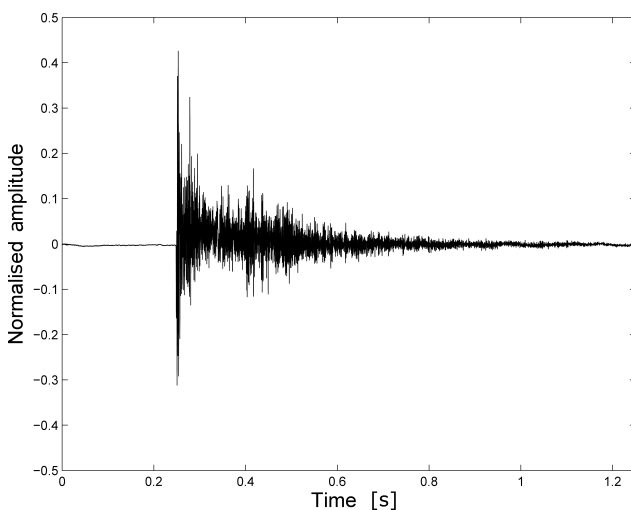


Fig. 1. Impulsive sound signal waveform of 0.40 semi-automatic pistol gunshot sampled at a 8 kHz rate.

reaches high power values, it may indicate abnormal events in several environments and human tasks, such as building collapses, gunfires, and explosions in factories. Systems to automatically detect these signals may be useful for monitoring emergency situations.

Impulsive signal detection has been addressed by several works based on a variety of approaches, including power (or amplitude) analysis, Wavelet Transform, correlation, and statistical methods (CHACÓN-RODRÍGUEZ *et al.*, 2011; KAUPPINEN, 2002). Techniques that rely on power assessment have presented lower computational cost at an acceptable performance as compared to statistical approaches.

These power-based methods are typically implemented by means of a power estimator followed by an onset detector, setting thresholds to activate the detector when limits for power and power slope are reached simultaneously (CHACÓN-RODRÍGUEZ *et al.*, 2011; DUFAUX, 2001). Better results were obtained by implementing adaptive thresholds (DUFAUX, 2001; SHARKEY *et al.*, 1996; SHOWEN, DUNHAM, 1999). Onset may be found through derivatives, median filters, and other methods applied to the power signal (DUFAUX, 2001; KAUPPINEN, 2002).

In general, impulse detection systems that address explosion and impact monitoring in large areas are composed of a central processing unit (CPU)

– a powerful server-like machine – and an array of acoustic sensors nodes, whose typical architecture contains a microphone, a microcontroller unit (MCU), and a transceiver (WESSELS, BASTEN, 2016). When an impulsive signal is detected in an acoustic sensor, the audio is transmitted to the CPU where a pattern classifier is run. If an emergency is detected, signals from different sensors are used to triangulate the source.

Despite the fact that CPU performs the most computationally demanding tasks and runs important decision algorithms, sensor nodes have bigger impact on system performance and cost. Sensors are spread throughout the monitored area such that there are intersection regions to allow triangulation. The overall system cost is then directly linked to the trade off between microphone quality, impulse detection efficiency, nodes separation, and MCU cost.

Categories of converters and microphones that present features compatible with those mentioned for sensor nodes are  $\Sigma\Delta$  ADCs and micro-electromechanical systems (MEMS) microphones, respectively. Both have been used in a broad range of applications due to their low cost, high quality, and high integration. Currently, semiconductor companies have manufactured MEMS microphone, pre-amplifier, and  $\Sigma\Delta$  ADC in a single integrated circuit.

$\Sigma\Delta$  modulation, also known as Pulse-Density Modulation (PDM), represents analog amplitude levels as a density of bits in a bitstream. Modulation process can be modelled as an oversampled ADC (at MHz range), single-bit quantiser and error feedback loop. This feedback implements the noise shaping technique that pushes the quantisation noise, generated by the low resolution, to higher frequencies and leaves the baseband almost intact. Therefore, conversion from  $\Sigma\Delta$  to PCM can be performed by means of a low pass filter followed by a downsampler.

An efficient method to filter and decimate with low computational cost, often used in  $\Sigma\Delta$  demodulation, is the Cascaded-Integrator Comb (CIC) filter. CIC decimators are implemented only with adders, decreasing the total number of operations substantially (PARK, 1991). However, their passbands are not flat, requiring an additional compensation FIR stage. Figure 2 illustrates a CIC decimator structure.

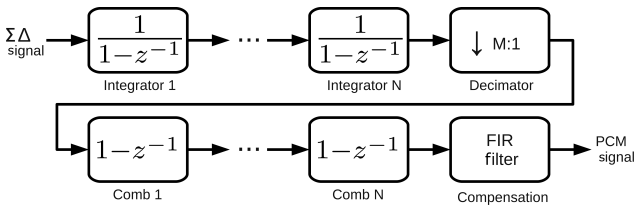


Fig. 2. CIC decimator structure for converting from  $\Sigma\Delta$  to PCM.

In this work, we propose an algorithm to detect impulsive sound signals directly in  $\Sigma\Delta$  format. It has

been shown that the proposed technique, which is based on DCT, reduces computational cost without increasing the error rate when compared to the reference architecture described in Sec. 2. The final goal is to enable sensor nodes to be implemented using MEMS microphones with built-in  $\Sigma\Delta$  converters, taking all of their advantages as well as relaxing MCU requirements.

## 2. Reference architecture for impulse detection in $\Sigma\Delta$ signals

Throughout this work, the system depicted in Fig. 3 has been chosen as the reference architecture for impulse detection in  $\Sigma\Delta$  modulated signals. The modules presented in Fig. 3 are described in this section and comprise three processing stages: a  $\Sigma\Delta$  demodulator, power estimator, and abrupt change detector.

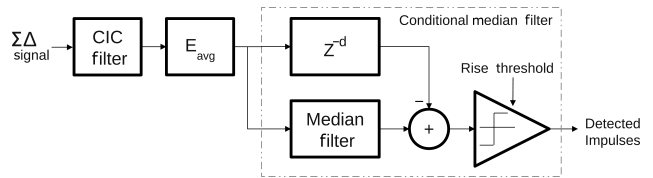


Fig. 3. Reference architecture for this work.

The  $\Sigma\Delta$  demodulator has been implemented using the CIC filter structure shown in Fig. 2. The number of integrators and comb pairs must be chosen according to demodulator requirements and signal characteristics, such as oversampling ratio, noise shape, output bits, and phase linearity. Fourth order CIC filters are suitable for most applications of  $\Sigma\Delta$  demodulator.

Since impulsive sounds are generally aperiodic, signal power has been estimated with successive computation of time average of the energy within  $L_P$  sample windows, as given by:

$$E_{avg}[k] = \frac{1}{L_P} \sum_{n=0}^{L_P-1} x^2[kL_P + n]. \quad (1)$$

The abrupt change detector has been implemented based on the Conditional Median Filter (CMF), introduced by KASPARIS *et al.* (1992). Firstly, in CMF, a median filter is applied to the power estimate sequence delivered by the previous stage, as in the expression below:

$$MF[k] = \text{median} \{E_{avg}[i] \mid i = k - L_M - 1, \dots, k\}. \quad (2)$$

Then, a threshold is applied to the median filter output, targeting to recover the power sequence but removing impulsive noise by detecting abrupt transitions, as follows:

$$\text{CMF}[k] = \begin{cases} \text{MF}[k], & \text{if } |\text{MF}[k] - E_{\text{avg}}[k-d]| < R_{\text{th}}, \\ E_{\text{avg}}[k-d], & \text{otherwise,} \end{cases} \quad (3)$$

where  $d = (L_m - 1)/2$  and  $R_{\text{th}}$  is a threshold for abrupt transitions. In case of adaptive threshold, this result might be used to estimate long term background noise power. For impulse detection, a binary version of CMF has been implemented as follows:

$$\text{BCMF}[k] = \begin{cases} 0, & \text{if } |\text{MF}[k] - E_{\text{avg}}[k-d]| < R_{\text{th}}, \\ 1, & \text{otherwise.} \end{cases} \quad (4)$$

Therefore the reference architecture has been implemented using (1), (2), and (4). In these equations,  $L_P$  and  $L_M$  determine the maximum impulse width detectable by the system, since the median filter only removes impulses with duration shorter than  $L_M/2$ .  $R_{\text{th}}$  must be chosen according to the average power of both impulses and background noise.

A simulation example for the reference architecture presented in this section is shown in Fig. 4. The generated input signal has two impulsive events and a sinusoid fragment between them (Fig. 4a). Note in Fig. 4b that although the impulses and the sinusoid have their power represented by the power estimator, impulses are removed from the median filter's output. Due to this, only two impulses are detected on the threshold module described by (4).

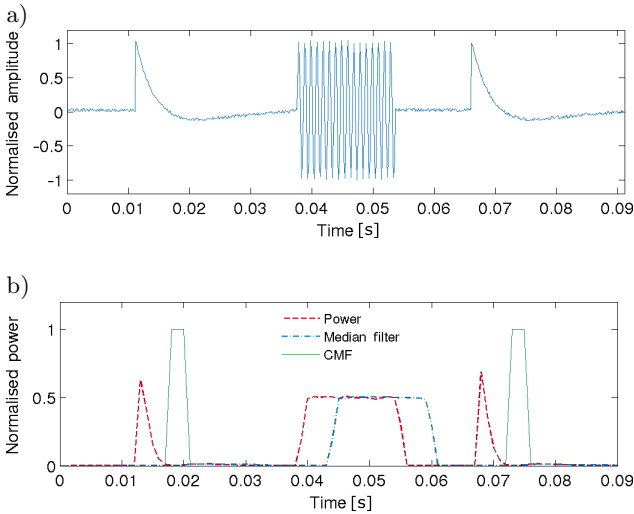


Fig. 4. Application example of the reference architecture: a) simulated input signal with two impulses and a sinusoid in between; b) outputs of power estimator, median filter, and CMF.

### 3. Proposed method

#### 3.1. Feature extraction

As mentioned before, the fast rise in energy is a very peculiar characteristic of impulsive signals. According

to Parseval theorem, this abrupt energy transition in time domain must be reflected by an abrupt energy transition in frequency domain, if Fourier transforms are taken immediately before and after the impulse onset.

In order to detect impulses in a  $\Sigma\Delta$  signal, only the energy in a narrow frequency range should be observed due to its high frequency quantisation noise. For example, most gunshots have their energy concentrated below 2 kHz, with peaks ranging from 500 to 600 Hz (GRAVES, 2012; MILLET, BALIGAND, 2006). This represents less than 1% of the spectrum in a  $\Sigma\Delta$  signal sampled at 512 kHz, making frequency selection a hard task to be performed in discrete time domain, as discussed in Sec. 1. In frequency domain, a straightforward strategy would be to use a Discrete Fourier Transform (DFT). Calculating a complete DFT for a wideband signal, with interest in only a few terms, is a very complex and wasteful operation.

Simpler and more efficient mathematical tools to represent the energy content of a signal are the Discrete Cosine Transform (DCT) and Discrete Sine Transform (DST) (AHMED *et al.*, 1974; BRITANAK *et al.*, 2010; RAO, YIP, 2014). Both transformations create a modified magnitude/frequency signal representation by correlating it, over time, with cosine and sine functions, respectively. DCT/DST has 16 different implementations, each one considering special boundary conditions on the finite input signal to form a periodic and symmetric sequence (BRITANAK *et al.*, 2010).

The DCT of type II (DCT-II) is the most used among DCTs and DSTs, mainly in audio and image applications due to its energy compaction property that allows signals to be represented using fewer components, if compared to DFT (KHAYAM, 2003). For an  $N$  point sequence  $x[n]$ , the direct form of DCT-II is defined as (OPPENHEIM *et al.*, 1989):

$$X^c[k] = 2 \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi k(2n+1)}{2N}\right), \quad (5)$$

$$0 \leq k \leq N-1.$$

The DCT-II transform produces coefficients that are related to  $2N$ -point DFT of a  $x_2[m]$  sequence, formed from zero padding  $x[n]$ . Being  $X_2[k]$  the DFT of  $x_2[m]$ ,  $X^c[k]$  can be expressed as (OPPENHEIM *et al.*, 1989):

$$X^c[k] = 2\text{Re}\left\{X_2[k]e^{\frac{-j\pi k}{2N}}\right\}, \quad k = 0, \dots, N-1. \quad (6)$$

This expression suggests that there is a close relationship between the DCT-II components and the signal spectrum. Squaring both sides of (6) and applying trigonometric identities, we receive:

$$\begin{aligned}
(X^{c2}[k])^2 &= 2 (\operatorname{Re}\{X_2[k]\})^2 \left(1 + \cos\left(\frac{\pi k}{N}\right)\right) \\
&+ 2 (\operatorname{Im}\{X_2[k]\})^2 \left(1 - \cos\left(\frac{\pi k}{N}\right)\right) \\
&+ 4 \operatorname{Re}\{X_2[k]\} \operatorname{Im}\{X_2[k]\} \sin\left(\frac{\pi k}{N}\right). \quad (7)
\end{aligned}$$

Similarly to DFT, a low pass filter may be implemented by computing only a few DCT components. Due to the narrow baseband of  $\Sigma\Delta$  signal and DCT energy compaction property, very few components may be computed for such a filter. This implies that  $\pi k \ll N$ , then  $\cos\left(\frac{\pi k}{N}\right) \simeq 1$  and  $\sin\left(\frac{\pi k}{N}\right) \approx 0$ . Thus, (7) may be reduced to:

$$(\operatorname{Re}\{X_2[k]\})^2 \approx \frac{(X^{c2}[k])^2}{4}. \quad (8)$$

Urban and countryside atmospheres are random medium for sound waves, due to the non-uniform presence of buildings, trees, mountains, and some moving objects that change wave characteristics through physical phenomena such as reflections, scattering, and dispersion. According to ISHIMARU (1978), these kinds of media vary in time and space, making sound amplitude and phase to fluctuate randomly. For a fixed source/receptor configuration, atmospheric turbulence may also cause signal variability between different emissions, as recently observed by CHEINET and BROGLIN (2015).

In this scenario, if the phase is considered to have approximately random uniform distribution through frequency domain and for different time frames, we can also consider that energy, for a given time window, is evenly distributed between real and imaginary components of DFT. Based on this assumption, energy may be estimated as:

$$\begin{aligned}
E_{x_2} &\approx \frac{1}{2N} \sum_{k=0}^{2N-1} 2 (\operatorname{Re}\{X_2[k]\})^2 \\
&\approx \frac{1}{2N} \sum_{k=0}^{2N-1} 2 (\operatorname{Im}\{X_2[k]\})^2. \quad (9)
\end{aligned}$$

Since  $x_2[m]$  and  $x[m]$  have the same energy and their spectra are symmetric, the estimator may be rewritten, arbitrary choosing the real term, as:

$$E_x \approx \frac{2}{N} \sum_{k=0}^{N-1} (\operatorname{Re}\{X_2[k]\})^2. \quad (10)$$

Replacing the result of (8) in (10), the DCT Energy Estimate (DEE) is introduced:

$$\text{DEE} = \frac{1}{2N} \sum_{k=k_{\min}}^{k_{\max}} (X^{c2}[k])^2, \quad (11)$$

where  $k_{\max}$  is the maximum selected component and  $k_{\min}$  is the minimum selected component.

DEE can be used as an power estimate for a selected range within a  $\Sigma\Delta$  signal. This estimator works over the following assumptions:

- $\pi k \ll N$ ;
- the baseband signal phase has an approximately random uniform distribution throughout frequency and time domains.

The first statement is generally true, once  $\Sigma\Delta$  signal is modulated using a high oversampling ratio. The second statement will be further analysed in Subsec. 4.2.

### 3.2. Impulse detector architecture

The proposal herein is to replace both the  $\Sigma\Delta$  demodulator and discrete power estimator of the reference architecture (Fig. 3) by DEE presented in (11). Figure 5 illustrates the proposed architecture for impulsive signal detection in  $\Sigma\Delta$  signals.

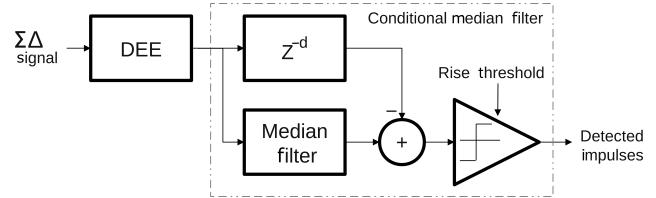


Fig. 5. Proposed impulsive signal detector architecture.

This replacement relies on the fact that DEE produces energy signals whose shapes are similar to those of the conventional method. If waveforms are preserved, impulses' characteristics in the energy signals are also preserved, allowing CMF (described in Sec. 2) to detect impulsive activities with a performance close to that obtained by the reference architecture. The main characteristics that must be preserved in the signal delivered by DEE are rise time, duration, and SNR.

The time frame length used in the demodulated signal must be also used when estimating energy with DEE in  $\Sigma\Delta$  domain. That means the number of samples analysed should increase by the oversampling ratio to keep the same frame duration.

Note that the proposed architecture does not convert  $\Sigma\Delta$  into PCM in any step of processing. Once detected an impulse, the audio signal may be transmitted to the CPU for further processing.

### 3.3. Computational budget decrease

To evaluate the advantage of the proposed detector against the reference architecture, the DEE module in Fig. 5 is compared to a CIC decimator, as depicted in Fig. 2, in conjunction with the discrete power estimator of (1).

A CIC decimator of  $L_{\text{CIC}}$  stages, as in Fig. 2, has  $L_{\text{CIC}}$  additions before decimation and  $L_{\text{CIC}}$  additions after it. For a  $L_{\text{comp}}$  th-order compensation filter, there are also  $(L_{\text{comp}} + 1)$  multiplications and  $L_{\text{comp}}$  additions after decimation. Therefore, for an oversampling ratio of  $M$ , it has  $L_{\text{CIC}} + \frac{L_{\text{CIC}} + L_{\text{comp}}}{M}$  additions and  $\frac{L_{\text{comp}} + 1}{M}$  multiplications per unit time (sampling period at Nyquist rate). The power estimator over  $L_p$  PCM samples requires  $\left(\frac{1}{M} + \frac{1}{L_p M}\right)$  multiplications and  $\left(\frac{1}{M} - \frac{1}{L_p M}\right)$  additions per unit time.

Each DCT component in DEE expression requires  $(k_{\text{max}} - k_{\text{min}})$  multiplication and  $(k_{\text{max}} - k_{\text{min}})$  additions per unit time. Thus, DEE requires  $(k_{\text{max}} - k_{\text{min}})(1 + \frac{1}{L_p M}) + \frac{1}{L_p M}$  multiplications and  $(k_{\text{max}} - k_{\text{min}})(1 + \frac{1}{L_p M}) - \frac{1}{L_p M}$  additions per unit time.

Multiplication on 1-bit streams can be replaced by conditional statements when computing DCT components. This way, multiplications per unit time reduce to  $\frac{k_{\text{max}} - k_{\text{min}}}{L_p M} + \frac{1}{L_p M}$  in the proposed one.

Figure 6 shows the number of operations as a function of the oversampling ratio for both reference architecture and proposed method. CIC and DEE parameters were assumed to be  $L_{\text{CIC}} = 8$ ,  $L_{\text{comp}} = 8$  and

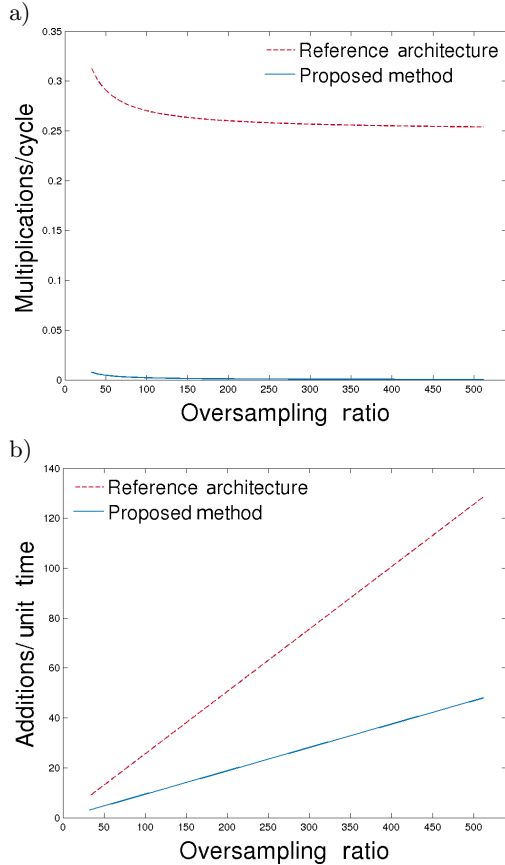


Fig. 6. Number of operations as a function of oversampling ratio  $M$ , assuming  $L_{\text{CIC}} = 8$ ,  $L_{\text{comp}} = 8$  and  $(k_{\text{max}} - k_{\text{min}}) = 3$  for  $M = 32$ . As  $M$  increases, these parameters have been increased in the same proportion.

A constant  $L_p$  of 32 was used.

$(k_{\text{max}} - k_{\text{min}}) = 3$  for  $M = 32$ . These parameters were increased at the same ratio as  $M$  in order to preserve the spectral characteristics for the two systems. The baseband processing frame,  $L_p$ , was fixed to 32. Note that the proposed method needs power operations in the considered range of  $M$  and the difference in the number of summations increases with  $M$ .

For instance, if  $M = 64$ ,  $L_{\text{CIC}} = 16$ , and  $L_{\text{comp}} = 16$  the power estimator using the CIC decimator requires 0.26 multiplications and 16.50 additions per unit time. For the same case, an energy estimator implemented using DEE, as in (11), with  $k_{\text{min}} = 1$  and  $k_{\text{max}} = 6$ , requires 0.01 multiplications and 6 additions per unit time. We can observe for this case that DEE requires 63.6% lower additions when compared to the reference architecture. Multiplications have a minor impact in both implementations.

## 4. Experimental results

### 4.1. Data collection

The dataset used in this work has 42 audio files containing a total of 407 gunshots, recorded in an outdoor shooting club. The recording was performed using a MEMS omni-directional digital microphone, MP45DT02, mounted on STM32F4 Discovery board that embeds an ARM based microcontroller, all three from ST Microelectronics.

The microphone was set to deliver 512 kHz PDM bit streams which were stored in a raw format for further processing. Besides the raw bit streams, a PCM version of these recordings was also generated by filtering it to 4 kHz and decimating it with a 64 down-sampling rate.

The firearms used for building this dataset were 0.38 revolvers and 0.40 pistols. Although the shooters were fixed in their positions, the sounds were collected from several distances and angles to provide, from one sample to another, diversity in propagation environment and Signal-to-Noise Ratio (SNR). The latter ranges from 36 to 54 dB.

### 4.2. Phase randomness and DEE analysis

An evenly distributed energy between real and imaginary through DFT components is a mandatory characteristic for the validity of (11). If that happens, the shape of the energy distribution is preserved even if only a real or imaginary component is used to compute its value.

This hypothesis has been evaluated assessing the similarity between the energy of  $X_2[k]$  with and without the imaginary component, using cross-correlation. Since the sound recordings are modulated in  $\Sigma\Delta$ , only the first 8 terms of a 2048-point DFT have been used, corresponding to 4 kHz in  $X_2[k]$  spectrum.

For the 42 audio files analysed, the cross-correlation average is 0.998 out of 1 and standard deviation of 0.0021. Thus, it is proper to say that the real portion of the income audio signal may be used to estimate its energy evolution.

In accordance with the result above, DEE has shown energy estimations similar to those obtained by the conventional process of (1), as observed in the plots of Fig. 7. This similarity indicates that an algorithm to detect impulses, like CMF, may have suitable performance when using DEE, since rise time, duration, and SNR of the energy pulses have remained nearly the same.

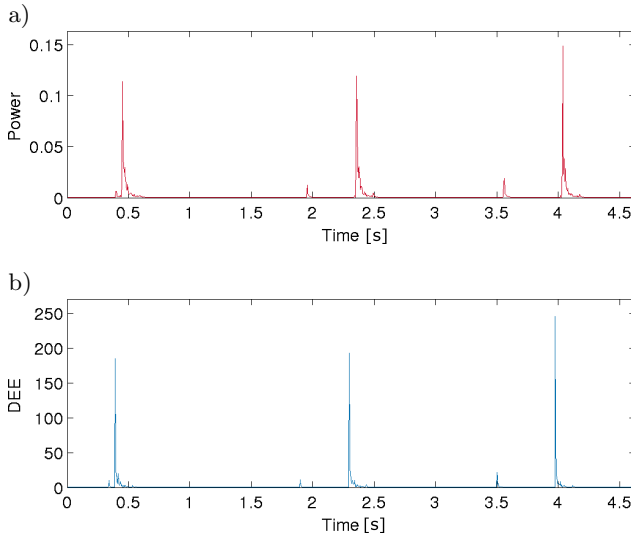


Fig. 7. Results for an audio fragment extracted from the dataset: a) power evolution calculated as in (1) from PCM signal; b) energy estimated with DEE from  $\Sigma\Delta$  bitstream, for  $k_{\min} = 1$  and  $k_{\max} = 6$ .

Note that although there are differences in energy peaks between the two wave shapes, the impulsive signature is preserved, allowing this signal to be used for detecting explosions. Additionally, a difference in scale was expected, since (1) is a measure of power, while (11) is an estimation of energy.

#### 4.3. Detection results

The proposed detector and the reference architecture were run over the dataset to detect impulsive activities. The power-based detector implemented assesses the signal using (1), while the proposed one uses the DEE. CMFs have been used in both detectors to highlight abrupt changes in energy.

Energy (or power) has been estimated on time windows of 4 ms, which corresponds to 32 samples for the power based detector and 2048 for the proposed one. The median filter length has been set to 11 in the two implementations. For both detectors, thresholds for energy/power and impulse measurer have been adjusted to decrease the False Positive and False Negative Rates

as much as possible for the audio sample with the lowest SNR.

As expected, the number of DCT components utilised in DEE has significant influence on the error rate. Figure 8 shows false positive and false negative errors as function of  $k_{\max}$ , reaching the minimum false negative error for  $k_{\max} = 6$ . If too few components are used, the spectrum is misrepresented, increasing the error. On the other hand, too many components may add  $\Sigma\Delta$  quantisation noise, affecting the detection.

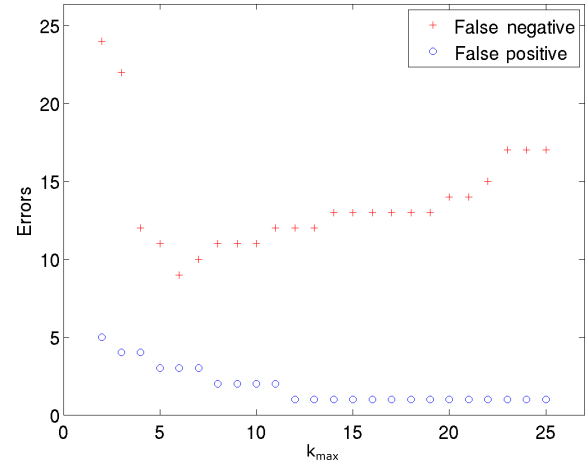


Fig. 8. False negative and false positive errors as a function of  $k_{\max}$  ( $k_{\min} = 1$ ).

Table 1 lists the results obtained by the two approaches. It can be seen that the proposed architecture has shown an acceptable performance (even better) as compared to the power-based implementation using fewer operation per cycle.

Table 1. Detection results for reference and proposed impulse detectors ( $k_{\min} = 1$  and  $k_{\max} = 6$ ).

	Reference architecture	Proposed method
Additions/Cycle	16.5	6
Detected	394 (96.8%)	398 (97.8%)
False Negative	13 (3.2%)	9 (2.2%)
False Positive	17	3

## 5. Conclusions

In this work, a method to detect impulsive sounds directly in  $\Sigma\Delta$  signals is presented. Experimental results performed on 42 recordings containing 407 gunshots have demonstrated that, in comparison with the traditional approach, a DCT-based energy estimator has presented an acceptable accuracy and lower computational cost. The proposed estimator has the only drawback of requiring to store tens of kilobytes for coefficients, which is reasonable for memory resources available in the today's microcontrollers technology. Evaluated with the same dataset, the proposed

impulse detector has shown a slightly better performance than the reference architecture, performing significantly fewer operations per unit time. Although, in this work, experimental results have been obtained for gunshot sounds, the proposed algorithm may be applicable to other impulsive events as long as they have similar waveforms and the parameters are adjusted to their characteristics.

## References

1. AHMED N., NATARAJAN T., RAO K.R. (1974), *Discrete cosine transform*, Computers, IEEE Transactions on, **100**, 1, 90–93.
2. BRITANAK V., YIP P.C., RAO K.R. (2010), *Discrete cosine and sine transforms: general properties, fast algorithms and integer approximations*, Academic Press.
3. CHACÓN-RODRÍGUEZ A., JULIÁN P., CASTRO L., ALVARADO P., HERNANDEZ N. (2011), *Evaluation of gunshot detection algorithms*, Circuits and Systems I: Regular Papers, IEEE Transactions on, **58**, 2, 363–373.
4. CHEINET S., BROGLIN T. (2015), *Sensitivity of shot detection and localization to environmental propagation*, Applied Acoustics, **93**, 97–105.
5. DUFAX A. (2001), *Detection and recognition of impulsive sounds signals*, Institute de Microtechnique Neuchatel, Switzerland.
6. GRAVES J.R. (2012), *Audio gunshot detection and localization systems: history, basic design, and future possibilities*, PhD thesis, University of Colorado.
7. ISHIMARU A. (1978), *Wave propagation and scattering in random media*, Vol. 2, Academic Press, New York.
8. KASPARIS T., TZANNES N.S., CHEN Q. (1992), *Detail-preserving adaptive conditional median filters*, Journal of Electronic Imaging, **1**, 4, 358–364.
9. KAUPPINEN I. (2002), *Methods for detecting impulsive noise in speech and audio signals*, [in:] Digital Signal Processing, DSP 2002, 14th International Conference on, Vol. 2, pp. 967–970, IEEE.
10. KHAYAM S.A. (2003), *The Discrete Cosine Transform (DCT): theory and application*, Michigan State University.
11. MILLET J., BALIGAND B. (2006), *Latest achievements in gunfire detection systems*, Technical report, DTIC Document.
12. OPPENHEIM A.V., SCHAFER R.W., BUCK J.R. (1989), *Discrete-time signal processing*, Vol. 2, Prentice Hall, Englewood Cliffs.
13. PARK S. (1991), *Principles of sigma-delta modulation for analog-to-digital converters*, Motorola Application Notes.
14. RAO K.R., YIP P. (2014), *Discrete cosine transform: algorithms, advantages, applications*, Academic press.
15. SHARKEY J.B., DOBLAR R.A., BOTHWELL F.E., BELT R.A., PAGE E.A. (1996), *System for effective control of urban environment security*, US Patent 5,504,717.
16. SHOWEN R.L., DUNHAM J.W. (1999), *Automatic real-time gunshot locator and display system*, US Patent 5,973,998.
17. WESSELS P.W., BASTEN T.G. (2016), *Design aspects of acoustic sensor networks for environmental noise monitoring*, Applied Acoustics, **110**, 227–234.