



# METHODS FOR EXTRACTION OF FEATURES AND DISCRIMINATION OF EMERGENCY SIRENS

Agustín Soto Otálora, David Eduardo Clavijo Osorio and Nicolás Charry Moreno

Ingeniero Electrónico Máster Ingeniería de Control Industrial Neiva, Huila, Colombia

E-Mail: [agussoto@usco.edu.co](mailto:agussoto@usco.edu.co)

## ABSTRACT

This article presents the analysis, study and tests, carried out to four (4) different methods combined of extraction of characteristics and discrimination or identification of these, in order to determine an adequate and effective method of identification and discrimination of Emergency Sirens (Police, Ambulance, etc). From the Cepstrum, MFCC's, FFT and Spectrogram as alternatives for extraction of characteristics of the signals to work, in addition Artificial Neural Networks (ANN) and Cross Correlation as discriminator and identifiers, we performed combined methods of these techniques to analyze their development and precision to the stated objectives. Finally a discussion of the methods worked on where we will present different guidelines and observations for the reader to keep in mind in his present and future works with related subjects.

**Keywords:** emergency sirens, tests, cepstrum, spectrogram, MFCC's, ANN, cross correlation, identifier, discriminator.

## 1. INTRODUCTION

Currently the identification and classification of sounds is a subject of technological development in progress and implementation. Due to its multiple functions in different contexts such as health, social, academic, labor and many others, are built, designed or implemented software and hardware with these integrated capabilities to meet different needs or problems that society has in Some of the above-mentioned contextual areas.

Emergency sirens (Police, Ambulance, Fire, etc.) are a subject of great importance and relevance for their representation and usefulness in situations that present with problems in some places such as cities, towns and other areas both urban and rural.

An autonomous classification system requires obtaining an input signal that represents the characteristics being evaluated (depending on the case may be, frequency components, spectral composition, power spectral density, etc.); A device that transforms that input signal into an electrical signal, and a digital processing system where all the operations and techniques necessary for the processes involved in the analysis of this type of signals are performed. Therefore, to perform the respective process to find an identifier and discriminator of emergency signals with good levels of efficiency and work, it is necessary to explore different techniques and alternatives to initiate the process of extraction of characteristics of the sound signals, for this process in For example, when identifying voice signals, Cepstrum and MFCC's (Cepstral Coefficients of Mel) are two very efficient work alternatives that we will test in this work, complementing also the basic and indispensable Work with the FFT and the spectrograms of the main signals; Knowing that the spectrogram is that representation in three dimensions, temporal, frequency and amplitude of the energy distribution of a signal that allows to better understand its content, than with a representation in the temporal domain. It should be mentioned that when it comes to classification of sound signals, the input signal may have variations and be directly affected by external factors such as noise,

making an exact classification become complicated, however a method such as artificial neural networks present Characteristics suitable for this problem as they are their flexibility and their tolerance to variations. The cross-correlation on its part is another technique for the relation of patterns or characteristics and to be able to reach that process of identification by alternate paths to that of the neural networks.

In this research we seek to conjecture different combined methods of these previously mentioned techniques to find different items to work like efficiency, computational cost, precision, etc. Four methods combined (extraction of characteristics and identification) will be presented in which they will briefly explain their development and tests carried out in each one with their respective analysis and relevant questions focused and oriented to the items that we put forward as bases.

The article has a sequential order by continuing with this introduction as follows. A small analysis and count of previous works that have served as guides in some aspect to the present article (II), A Background of the thematic of the different techniques to work and its individual process in the work from the process of acquisition of the signals, Extraction of characteristics, among others (III), an approach of the different combined methods to test and how they will be used with their respective tests (IV), ending with a space of analysis and discussion of everything proposed and concluding the best for different scenarios To work (V). It is also important to mention that the working environment of most of these tests was performed in MATLAB software.

## 2. REVIEW PREVIOUS WORK

The antecedents and previous investigations, are a fundamental base for the work that we realize; Specifically in the area of sound identification are presented different items to work to fulfill this main objective, mainly because there are differences between types of audio like speech (speech) or music versus some like this type of sirens, because of their different structure,



Frequency and periodic characteristics, among other peculiarities. There are applications or devices that identify by group sections of sounds, at different distances or with alternatives of similar effectiveness, however in specific cases of precision in identifying exactly the objective other measures are taken, and these measures have not been used in Great measure in this type of siren sounds of emergencies, at least in latitudes close to our country. An algorithm that is truly enviable because of its accuracy when it comes to identification and its range of reach is that of the popular application of millions of SHAZAM downloads [1], its extraction of characteristics of signals using spectrograms is excellent because of its wealth of information, data Which we took as examples even though audios of a different type were identified.

In other cases, situations of identification of environmental or environmental sounds are presented [2], special cases where the extraction of characteristics of the signals focuses on time and frequency aspect, working algorithms forged empirically and with large bases that find high efficiencies In the identification of the characteristics of the signals using also the MFCC's. Different sounds like rain, flowing rivers, insects, etc. are differentiated with great precision and a high average of effectiveness. For situations such as speech identification or word pronouncement, there is an investigation of an inexpensive device for identifying isolated words [3], its main tools in the analysis, extraction of characteristics and identification are the Neural Networks, MFCC's, DTW (Dynamic Time Warping) and PCA (Principal Component Analysis), where the combination of neural networks with PCA presented the best results in terms of efficiency and precision, adding that 3 different scenarios were worked out between these combined techniques.

Using the Fast Fourier transform twice in the detection of ambulance signals, to try to convert the data into numerical values that facilitate future processes for detection [4], an algorithm in more rustic conditions such as greater ambient noise Or low volume and still detect it. Developed by Japanese Colleagues, they emphasize that the ambulance siren (very specific case for us) presents two important characteristics to analyze, one is that the sound has two different frequencies and the other is that each tone periodically sounds and goes off, through the use Of the FFT is looking for that particularity of converting to numeric value those two characteristics, working first with a FFT of 64 points, and then one of 16 points. The method of analysis consists of two steps, the first is to analyze the frequency tones, and the second is to analyze the amplitude changes of each pitch frequency, in both steps the FFT are used to analyze the frequencies. A research with similar aims to ours but rather indirect methods program an algorithm in an embedded system able to identify steels by acoustic methods using Artificial Neural Networks [5]. For the identification the sound emitted by a bar of the material when it is struck As a bell, the identified sound is transformed to the frequency domain by applying an FFT, from the frequency spectrum the parameters used to train the network are obtained: energies per frequency band, fundamental components,

among others. Finally, a Back propagation trained network with multilayer structure is trained.

As an example of a final research, an excellent project of the Worcester Polytechnic Institute in the USA, an identifier of the song of the wild birds [6] was studied, the main objective of the project is that it is able to identify these songs of many species of species of Birds and that is worked for that range is much greater. The basic operation of the algorithm uses the MFCCs and DFT for its process of extraction of characteristics and Cross correlation in its process of identification.

Finally as researchers between our main objective we highlight the analysis of the different techniques or methods previously exposed and others that can be found to be able to pose different work scenarios and locate the most appropriate methods according to what we obtained in the different tests performed, one idea, but produce several.

### 3. BACKGROUND

In the first instance we will define the characterization of the emergency signals and the main techniques to work for the extraction of the main characteristics of these.

The different emergency signals were simulated by a professional percussion siren acquired for this work. The recording was performed with the different Matlab commands for this type of functions. Recordings were made with a time of five seconds, with a Sampling Frequency of 8000 Hz. We present some of the graphs obtained from the emergency signals. Before we concentrate on the extraction of signal characteristics, it is important to know that the numerical values contained in these signals are not adequate to be processed by the algorithms that are used for the recognition of the signals, that is why we perform a pre- Signal processing mainly composed By the design of a pre-emphasis filter [G], an IIR filter used in the recognition systems to eliminate signal noise, reduce volume effects; And a hamming window.

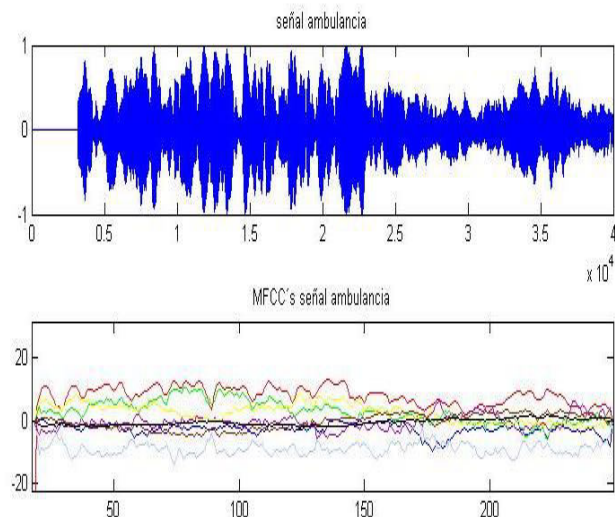


Figure-1. Ambulance signal.

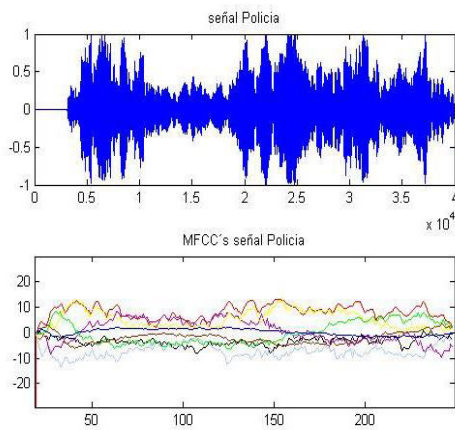


Figure-2. Police signal.

### $H(z) = (1 - az^{-1})$ Ec 1. Pre-emphasis Filter

In the first instance we will define the characterization of the emergency signals and the main techniques to work for the extraction of the main characteristics of these.

The different emergency signals were simulated by a professional percussion siren acquired for this work. The recording was performed with the different Matlab commands for this type of functions. Recordings were made with a time of five seconds, with a Sampling Frequency of 8000 Hz. We present some of the graphs obtained from the emergency signals. Before we concentrate on the extraction of signal characteristics, it is important to know that the numerical values contained in these signals are not adequate to be processed by the algorithms that are used for the recognition of the signals, that is why we perform a pre- Signal processing mainly composed By the design of a pre-emphasis filter [G], an IIR filter used in the recognition systems to eliminate signal noise, reduce volume effects; And a hamming window.

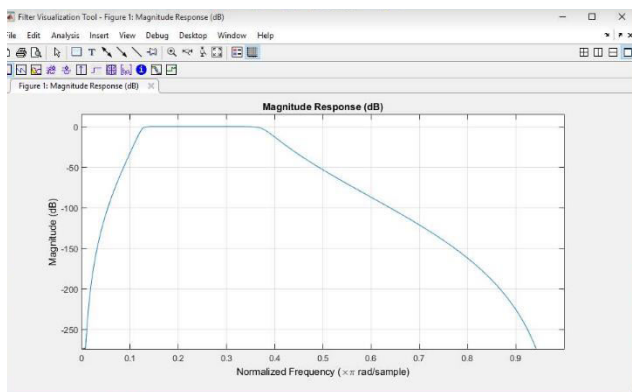


Figure-3. Filter magnitude response.

FFT (Fast Fourier Transform): [H] The direct implementation of this equation (Eq. 2) is very inefficient, especially when the sequence of length  $N$  is very long since to obtain a complete set of DFT coefficients, Complex and  $N(N-1)$  complex sums, which is a

drawback. Therefore, the Fast Fourier Transform (FFT) is used, an algorithm that allows the rapid and efficient implementation of the DFT. In the Elaboration of the FFT, the properties of symmetry and periodicity of the phase factor  $w_N$  are exploited, based on the use of the "divide and conquer" strategy.

$$x(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}}, k = 0, 1, 2 \dots N-1$$

### Ec 2. Fourier Fast Transform

Basically it takes a Buffer of 4096 corresponding to a little more than 0.5 seconds of signal; this Buffer is composed of 8 Frames of 512 samples. Each Frame passes through a Filtering process, corresponding to the previously designed Passband Filter, after which a tapering process is performed through a Hamming sale and finally the 512 Point FFT is applied. Use of the fft and hamming functions in MATLAB was performed for the tests.

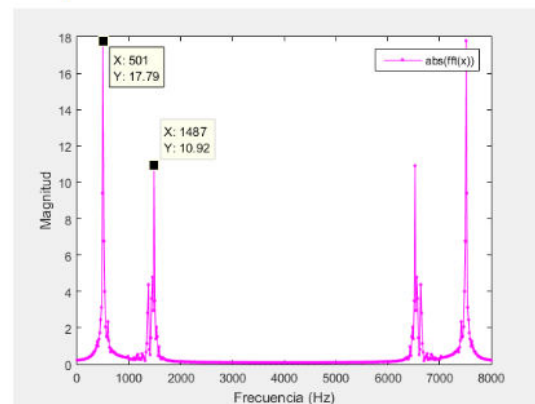
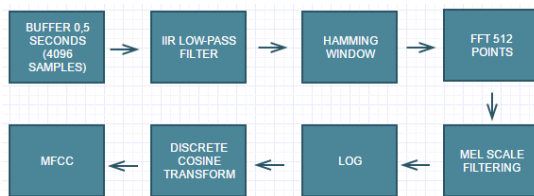


Figure-4. 512 pts FFT of ambulance signal.

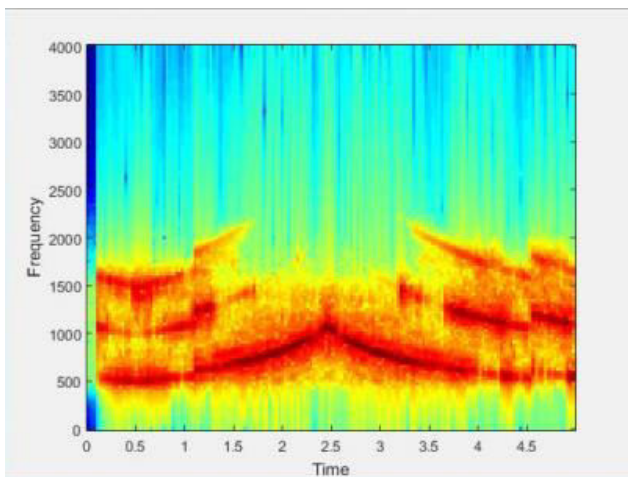
MFCC's (Mel Frequency Cepstral Coefficients): [I] The MFCC algorithm is used to extract the features. The functions used for feature extraction [ $x_{cep}$ ,  $x_E$ ,  $x_{delta}$ ,  $x_{acc}$ ]. MFCC are the most important features, which are required among various kinds of speech applications; it gives high accuracy results for clean speech. And MFCC can be regarded as the "standard" features in speaker as well as speech recognition. . Signal analysis using the cepstral coefficients has been commonly used for voice recognition with good results. By applying this technique, the samples of the speech signal are transformed to a set of coefficients that efficiently represent the spectral properties and concentrations of Energy of the Voice Signal, trying to emulate the type of processing that our auditory system performs. When taking into account the characteristics of the ear, it is a question of resembling the system to the recognition made by a person. This analysis is based on the use of the Mel frequency scale, which is a linear spacing of the frequency below 1000Hz and logarithm spacing above 1000Hz.



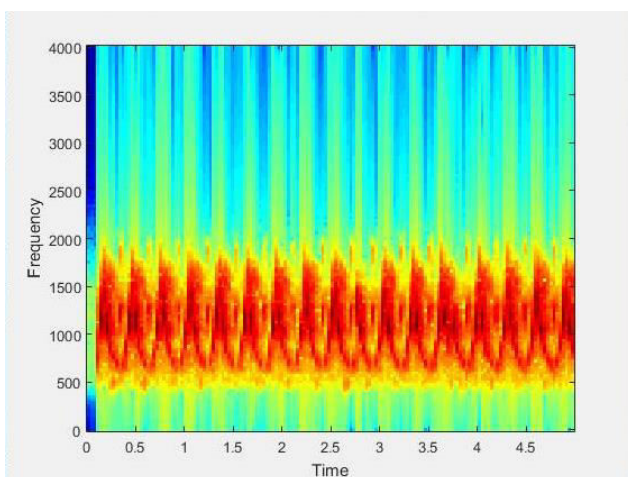


**Figure-5.** MFCC's feature extraction.

**Spectrogram:** Through visual inspection of the spectrograms of typical sound events, it is clear that a large amount of information is contained in the joint time-frequency representation. With careful analysis, it is possible to recognize similar sound events based on this visual information information [J]. In the initial phase of development all the spectral processing described is performed by the MATLAB spectrogram function, which returns the spectrogram of the signals under consideration, and whose operating parameters are the signal to be processed, the number of FFT points per Window, sample rate of signal, window function and number of overlapping points between windows.

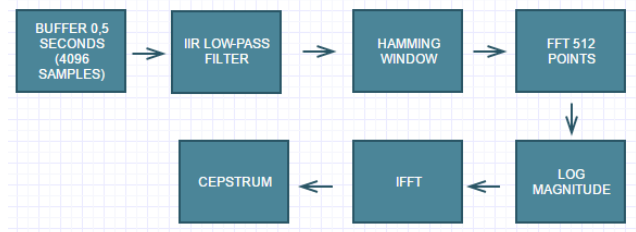


**Figure-6.** Spectrum representation of Siren's ambulance.



**Figure-7.** Spectrum representation of Siren's police.

**Cepstrum [K]:** The cepstrum allows distinguishing between vocalized and non-vocalized sounds. The tone can be determined from the cepstrum. Regarding the implementation of Cepstrum as a characteristic extraction algorithm in MATLAB, the following procedure was elaborated:



**Figure-8.** Cepstrum feature extraction.

Again, a 4096 sample buffer, composed of 8 frames of 512 samples, each of which has been filtered and wrapped, is extracted to obtain the FFT of 512 points and complete the cepstrum algorithm by the logarithm of the inverse transform of the Logarithm of fft. For its respective implementation the functions of MATLAB hamming, fft, log, and ifft were used.

To conclude with this section, we also refer to the methods that we will use in the classification area of these characteristics for the final methods that we implement. In this case we work with two important issues:

**Artificial Neural Networks (ANNs):** A rather clear and concise concept of a neural network is the one proposed by Hech-Nielsen who preaches the following: "An artificial Neuronal network is a computer system, made by a large number of simple elements Highly interconnected process elements, which process information by means of its dynamic state as a response to external inputs "[L]. For the automatic detection as to the determination of which type of Emergency signal has been captured, the use of Neural Networks is chosen, for this particular case, a Feedforward network in one of its specialized versions known as Network of Pattern Recognition (Patternet), which is basically a network that is used to classify the entries according to the target of different classes.

The activation functions for this type of network are: Tansig for the softmax for the output layer. In MATLAB we have the function patterned through which we will perform the respective tests and tests.

In addition to neural networks by pattern recognition we also work with self-organized maps (SOM). This is a special form of ANN trained with an unsupervised paradigm of learning. It follows the competitive method of learning that enables it to work effectively as a feature map, classifier, and, at times, filter. SOM has a special property of effectively creating spatially organized "internal representations" of various features of input signals and their abstractions [M].

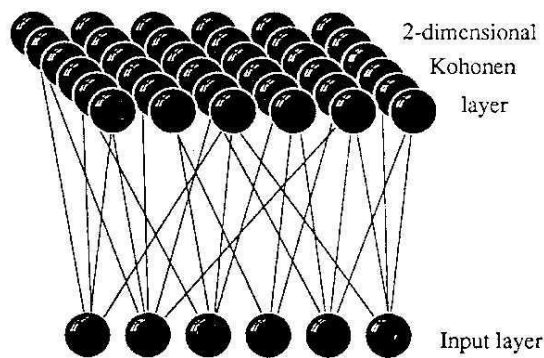


Figure-9. A simple SOM.

Cross-correlation: As previously mentioned, one of the main purposes of this research is to establish a comparative analysis of different techniques and methods to design a system of extraction of characteristics and identification, that is why this search came to study the algorithms of correlation [O], due to its great potential in real-time operations and its moderate requirements in terms of hardware and software resources in signal processing, the cross-correlation [N] (or sometimes referred to as "cross-covariance") is Measure of similarity between two signals, often used to find relevant characteristics in an unknown signal by comparison with another that is known. It is a function of the relative time between signals, sometimes also called displaced scalar product, and has applications in pattern recognition and cryptanalysis.

Given two discrete functions  $f_i$  and  $g_i$  the cross-correlation is defined as:

$$(f \star g)_i \stackrel{\text{def}}{=} \sum_j f_j^* g_{i+j}$$

#### Ec 3. Correlation of two functions discrete

Where summation is performed on integer values of appropriate  $j$ ; and the asterisk is indicating the conjugate. For the case of two continuous functions  $f(x)$  and  $g(x)$  the cross correlation is defined as:

$$(f \star g)(x) \stackrel{\text{def}}{=} \int f^*(t)g(x+t) dt$$

#### Ec 4. Correlation of two continuous functions

The cross-correlation has a similar nature to the convolution of two functions. It differs in that the correlation does not involve a signal inversion as it does in the convolution.

### 4. IMPLEMENTED METHODS

In most modern research projects engineers, researchers, scientists, etc. ... seek to take a step forward in each of the paths that have been developed or explored previously. That is our main objective as researchers in this project; to find and materialize new steps in the development of the aforementioned techniques and

methods in the specific area of the identification and classification of Emergency Sirens Sounds.

After reviewing and explaining the techniques and methods for the extraction of characteristics and classification respectively, we proceed to make the combination of these and to propose the different methods that we decided to work for this research. In this section we are going to mention each one of them and what was the process that we carried out developing them.

#### A. FFT - ANN

After obtaining the Buffer resulting from the extraction of characteristics by means of FFT, this matrix is Normalized by the normalization method min-max [-1, 1], obtaining a training matrix. As an example of an illustrative training the following situation is proposed in which a patterned network will be trained from 3 samples of 0.5 seconds captured from each signal. We perform the methodology instructed in the block diagram for a sound capture of each of the target signals, that is, we obtain three training vectors, one for each objective signal, in this way we form the Vector Input:

Table-1. Input vector 1 x 12288.

Main Vector 1 X 4096 Elements AMBULANCE	Main Vector 1 X 4096 Elements POLICE	Main Vector 1 X 4096 Elements FIREMAN
---	--	---

Now with the compound input vector, the classification matrix is created; this matrix contains the targets belonging to each of the classes (Ambulance, Police, and Fire). This matrix must correspond to the size of the Input Vector, that is, it must have a size of 3x12288 for this case. The composition of this matrix can be illustrated as follows:

Table-2. Matrix Classification 3 x 12288.

4096 Ones	4096 Zeros	4096 Zeros
4096 Zeros	4096 Ones	4096 Zeros
4096 Zeros	4096 Zeros	4096 Ones

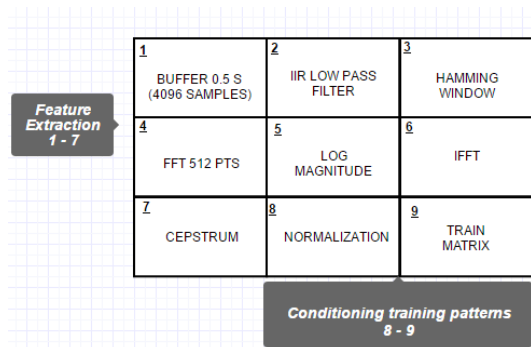
Each row of the classification matrix has 1/3 elements formed by some (1), this corresponds to the elements of the Input Vector, in this way the target is assigned to each class of the Input Vector to perform the training.

For the neural network we worked on an input matrix, which would correspond to the normalized patterns obtained by the characteristic extraction algorithm. An occult layer composed of 150 neurons and an exit layer

Composed of three neurons corresponding to each emergency signal to be discriminated.



### A. Cepstrum - ANN



**Figure-10.** Performance methodology Cepstrum – ANN.

We return to the Cepstrum algorithm mentioned in the previous section. The conformation of the Input and Classification vectors for the Patterned network are performed identically to the one mentioned in the last method. The network topology is preserved in the same way.

To carry out the training of the Neuronal Network patterned from the patterns obtained by algorithms of extraction of characteristics it is necessary to carry out a process of Normalization of data, since it is the optimum and therefore what is intended with this, is that both Inputs and outputs are concentrated in an acceptable length range and are scattered within that range for simpler data processing.

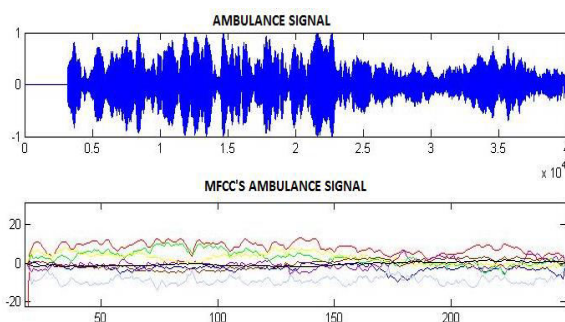
For our case a normalization min-max is applied that looks for that the values of the training matrix are between [-1, 1]

$$x'_i = \left( \max_{\text{objetivo}} - \min_{\text{objetivo}} \right) \times \left[ \frac{(x_i - \min_{\text{valor}})}{(\max_{\text{valor}} - \min_{\text{valor}})} \right] + \min_{\text{objetivo}}$$

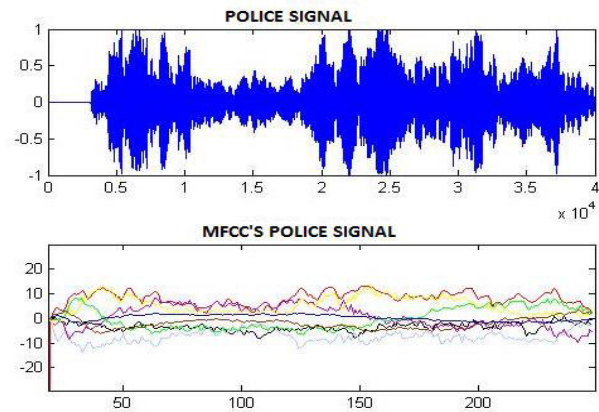
### Ec 5. Normalization

#### B. MFCC - Cross correlation

For the design of the algorithm that was worked for this method, in the first instance as it is due to the extraction of the characteristics of the emergency signals through the MFCC's as explained in the previous section (Background).



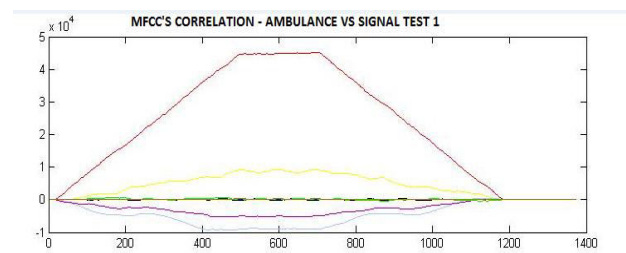
**Figure-11.** Extraction MFCC's Ambulance Signal.



**Figure-12.** Extraction MFCC's police signal.

Having already obtained the MFCC's of the emergency signals with which it is going to work, we proceed to realize what is in itself the cornerstone of the algorithm of this method, the database. A database is set up to store the MFCC's of the signals with which it was worked, to have them as points of comparison and analysis.

After having this ready, it is where the use of cross-correlation begins to work, because what is done in this method is that different test signals are entered into the algorithm so that it also obtains its respective MFCC's and finally the Correlation of the MFCC's obtained from the test signal X that enters in its moment with the MFCC's of the emergency signals that are housed in the database that was mentioned previously.



**Figure-13.** MFCC's Correlation of Signal Ambulance vs Signal Test 1.

#### C. Spectrogram - ANN (SOM)

This is the last method proposed to develop in this project. In this method is used Spectrogram next to the networks SOM. This method has a more complex algorithm both in feature extraction and in the discrimination method.

For this method a frequency subband compression algorithm is applied, the resulting matrix of the spectrogram for a 0.5 second signal buffer has a size of 121x31 elements, which is to reduce the size of that matrix for Optimize the training process. The compression algorithm basically works as follows, the 121 columns correspond to the frequencies of 0-4000 Hz, therefore, they divide three sub-bands, one of 0-500 Hz, another of 500-1500 Hz (Band of Interest) and finally 1500-4000 Hz, in this way the matrix is reduced to a size of 17x31.





The topology of the proposed SOM Network is as follows:

There will be 31 SOM networks corresponding to the 31 windows used in the Spectrogram, in addition to another final SOM Network which will be trained with the answers of the 31 SOMs above, in order to achieve the discrimination of the alarm signals. Each SOM has a 3x3 dimension.

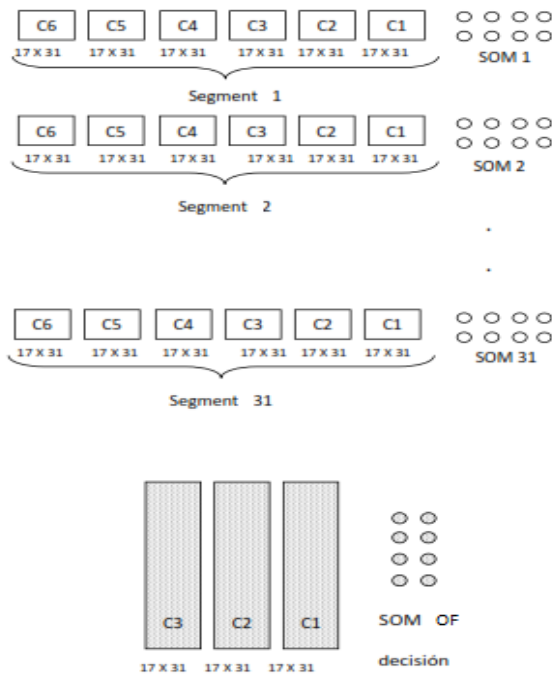


Figure-14. Neuronal classification system.

In the decision SOM C1, C2, C3 are the classes that identify each command, where its 17 rows are given by each SOM of the classification neural system and its 31 columns by the number of times a command is repeated for training.

#### 4. DISCUSSIONS

	FFT+ANN (%)				CEPSTRUM +ANN (%)				MFCC'S + CORRELACION CRUZADA (%)				ESPECTROGRAMA+ ANN (%)			
	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
1	42	45	42	40	59	57	60	62	78	77	77	80	71	73	76	75
2	39	46	40	41	58	62	60	63	78	83	82	80	76	75	77	75
3	40	45	43	41	61	60	57	55	80	80	82	85	75	75	76	76
4	47	43	41	41	61	56	58	63	82	82	76	78	77	76	75	74
5	42	37	32	40	60	61	61	58	80	80	78	76	76	71	77	73
M	42	43	40	41	60	59	59	60	79	80	79	80	75	74	76	75

#### Results of testing methods for extraction of characteristics and identification

In the competent to Method 1 (FFT + ANN), we have that it is a simple method as far as the algorithm used in the extraction of characteristics, due to the fact that

when processing 4096 samples, in the end we only use 2048 per analyzed buffer as effect to use Only the unilateral spectrum, which is a significant reduction in terms of memory in order to implement said algorithm in an embedded system. As for the implementation of the Neural System, the program memory costs are to save the bias and weights of each of the layers of the neuronal system. For the topology used we have two layers, in layer 1 two matrices of dimensions 150x1 corresponding to Bias and Synaptic Weights. Finally for the layer 2 corresponding to the output layer there is a matrix of dimension 3x1 for Bias and for weights one of dimension 3x150. In synthesis we have an algorithm of identification and classification that in view of being implemented in an embedded system would have a low computational cost in terms of the use of RAM and of program, reason why would be optimum for the operation in real time. Unfortunately as seen in the results table, after the tests performed the efficiency of such algorithms is not optimal for our purpose. The algorithm of extraction of characteristics to be a simple frequencies analysis is not enough to obtain a set of significant samples that allow a correct classification, mainly because the bandwidths of each one of the emergency signals share frequencies components very common among them Which makes its classification difficult. To try to improve the efficiency of this method we tried to increase the Transform points without any success, however, it was observed that reducing even 1/3 the neurons of the Classification System obtained the same results.

For Method 2 (Cepstrum + ANN), a more complex method of extraction of features was set up in order to detect changes in tonalities by means of Cepstrum. For the implementation of such an algorithm as mentioned above, the Fourier Transform, the Fourier Reverse Transformation and Logarithm should be used. This algorithm unlike the previous one will occupy many more variables in memory unlike Method 1, occupying four matrices of dimensions 2048x1 to perform said feature extraction processing. For the implementation of the Network we have Bias and Weights matrices with the same dimensions of method 1 for each layer of the neuronal system. As we see the implementation of this method is much heavier for the implementation in embedded system due to the large cost of RAM and program that entails said feature extraction algorithm. As for its efficiency, in the tests carried out initially it was the indicated method because it reached discrimination thresholds of each class above 85% by increasing the transformation points between 512 and 1024, these results were obtained when training the Network With a single possible scenario, when training with different scenarios efficiency decreased by about 60%. When implementing this method we find a new problem when observing the decreasing efficiency of the classification algorithm, which when trained in several scenarios, the time in which each change of tone occurs is different for each recording scenario and sometimes coincide between different classes, which severely reduces the classification efficiency.



In relation to Method 3 (MFCC's + Correlation) a more robust extraction algorithm was implemented than the two previous methods, unlike Cepstrum's method, a filter bank is used and the Discrete Cosine Transformation as was seen in the Background. As for this method its critical part is the implementation of the database for the correlation of the MFCC, since it is confirmed by 5 scenarios for each of the three classes, which would occupy enough program memory. However, the use of DCT will increase the use of RAM, added to the correlation function. Comparing method 1 and 2, this would be the most complex to implement and the most inconvenient to operate in real time would have, its counterpart is that it has a great efficiency over 80%, a very efficient result for our purpose.

Finally, Method 4 (Spectrogram + ANN), the implementation of the feature extraction algorithm is just as simple as method 1 since it is based on the FFT. Its counterpart is in the execution of the classification algorithm, the proposed neuronal system is much more complex as mentioned previously, being 32 Neuronal networks of the SOM type, the program memory expense to save the weights of the SOM Network is quite Considerable, we also have the implementation of the compression algorithm which will consume a significant space in RAM. This method obtained an efficiency of 75%, but unlike the previous method, it has more possibility to function optimally in real time, which is why it is the method selected for future implementation.

## REFERENCES

- [1] Chun W. Avery Li. An Industrial – Strength Audio Search Algorithm. Shazam Entertainmetn Ltda. Available: <https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf>
- [2] Chu S., Narayanan N., Kuo J. 2009. Environmental Sound Recognition with Time-Frequency Audio Features. in IEEE Transactions on Audio, Speech, And Language Processing. 17(6).
- [3] González. C.C, Murrugura L. N. 2013. Isolated word recognition using a Low Cost Microcontroller. In III Brazilian Symposium on Computing Systems Engineering).
- [4] Miyazaki. T. -KitazonoYuhki. - Shimakawa M. 2013. Ambulance Siren Detector using FFT on dsPIC. in III Proceedings of the 1<sup>st</sup> IEE/IIAAE International Conference on Intelligent Systems and Image Processing.
- [5] Ramírez M. I., Duque P. E., Aedo C. J, Aristizabal R. Sistema Embebido para la Identificación de Aceros por Métodos Acústicos, Utilizando Redes Neuronales. Univesidad de Antioquia.
- [6] Carroll T., Colangelo R., Strott T., 2010. Bird Call Identifier, Identifying Songs of Bird Species through Digital Processing Techniques. Worcester Polytechnic Institute.
- [7] Sánchez M., Doubert G. 2004. Segmentación y Realce de Señales de Voz Usando la Transformada de Wavelet y DSP's. Universidad del Quindío.
- [8] Sek M. Frequency Analysis Fast Fourier Transform, Frenquency Spectrum. Vicoria University. Available: <http://www.staff.vu.edu.au/msek/frequency%20analysis%20-%20fft.pdf>.
- [9] Dhingra S., Nijhawan G., Pandit P. 2013. Isolated Speech Recognition Using MFCC and DTW. In International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2(8).
- [10] Dennis J. William. 2014. Sound Event Recognition in Unstructured Environments Using Spectrogram Image Processing. Nanyang Technological University.
- [11] Pino Raúl; Gómez Alberto; De Abajo Nicolás. 2001. Introducción a la Inteligencia Artificial: Sistemas Experto, Redes Neuronales Artificiales y Computación Evolutiva. Oviedo, España. Servicios de publicaciones Universidad de Oviedo(Ed).
- [12] Randall B. Robert. A History of Cepstrum Analysis and its Application to Mechanical Problems. School of Mechanical and Manufacturing Engineering University of New South Wales.
- [13] Sarma M., Sarma K. K. 2013. Vowel Phoneme Segmentation for Speaker Identification Using an ANN - Based Framework. In Journal of Intelligent Systems. 2: 111-130.
- [14] Ibarra Manzano Oscar. 2007. Notas Académicas Curso Procesamiento Digital de Señales. Mexico. Universidad de Guanajuato(Ed).
- [15] Roma N., José Santos V., Tomé J. A Comparative Analysis of Cross- Correlation Matching Algorithms Using a Pyramidal Resolution Approach. Instituto Superior Técnico.