Detection of acoustic events with application to environment monitoring

J. Moragues, A. Serrano, G. Lara, J. Gosálbez and L. Vergara

Instituto de Telecomunicaciones y Aplicaciones Multimedia,
Universitat Politècnica de València,
8G Building - access D - Camino de Vera s/n - 46022 Valencia (Spain)
Corresponding author: jormoes@upvnet.upv.es

Abstract

The goal of this work is to present different detection techniques and its feasibility for detecting unknown acoustic signals with general applicability to different noise conditions. These conditions replicate those commonly found in real-world acoustic scenarios where information about the noise and signal characteristics is frequently lacking. For this purpose, different extensions of the energy detector and even new structures for improving the robustness in detection are considered and explained. Furthermore, three different research lines of application are presented in which the energy detector and its extensions are used to improve the localization accuracy and the classification rates of acoustic sounds.

Keywords: Acoustic scene analysis, energy detector, non-Gaussian background noise, non-independence.

1. Introduction

The simplest problem in detection is to decide whether the observation vectors are formed by a known signal in the presence of noise, or just noise. However, this requires knowing the characteristics of the signal to be detected and the noise in which it is found. Therefore, the degree of difficulty of a detector is inversely related to the degree of knowledge about the signal and the background noise, in terms of probability density function (PDF). The ideal detection case occurs when the PDF of the signal and noise are fully known, since this situation offers the possibility of obtaining the optimum detector for both Bayes and Neyman Pearson criteria [1]. When the characteristics of the signals are not entirely known, some other detection solutions, though probably not optimal, can be attempted in order to obtain a suitable detector.

Gaussian distribution is frequently considered because of its widespread theory and practical applications. Detection of unknown deterministic signals in background Gaussian noise is a classic detection problem. The energy detector (ED) implements a generalized likelihood ratio test (GLRT) when the noise is Gaussian [1]; but when the noise is non-Gaussian and non-independent, the performance decreases. The Gaussian and independent noise assumptions are typically used in various scenarios for mathematical simplicity when studying the behavior of a detector. In numerous applications and real-world problems, utilizing this approach is very useful as it makes the implementation of simpler detectors feasible. However, not all scenarios can be characterized as independent Gaussian noise due to the unpredictable characteristics of each particular case. In these situations, it is necessary to study the detectors in the presence of noise with a non-Gaussian distribution and with dependence between the samples. Therefore, it is possible to apply several extensions to the classical ED to improve the behavior of the detector as it will be presented through this work.

It is also important to note a highly-relevant issue affecting detection, yet which is often overlooked: the actual duration of the signals to be detected. In some applications such as echo detection in radar, sonar or acoustics, an approximate idea of the signal length may be available. However, in the context of novelty or event detection, where the characteristics of the signals are unknown, no information about the duration is available since any type of signal may appear in the environment under study. Selecting the temporal duration of the observation vector is very challenging and may significantly affect the detection, since the observation duration could be too long or too short for the actual signal duration, thereby producing a significant loss in the probability of detection. A similar problem is observed in the frequency domain, where signal bandwidth (instead of signal duration) is difficult to determine. Hence, signal duration introduces a degree of uncertainty to the detection problem. A detailed study is thus necessitated to determine how the behavior of the detector used will be affected.

Finally, this work will also focus on the description of new technologies in three growing and well differentiated application areas that incorporate techniques to automatically detect events or novelties in a monitored acoustic environment: surveillance systems, man-machine interaction and cultural heritage applications. Hence, sound-based systems are good candidates for verifying and evaluating the behavior of an energy detector; they define scenarios where both background noise and events may have variable and unpredictable characteristics. These systems are currently being assessed by the "Grupo de Tratamiento de Señal (GTS)" in the "Universitat Politècnica de València (UPV)" for incorporation within the framework of different projects:

- The first one falls within the field of security in public places and it has been supported under a CENIT project called HESPERIA [2].
- The second one lies within the framework of a cooperative exchange program of Ph.D. students with the "Acoustic scene analysis group" of "Universität Karlsruhe (TH)" (Germany). The cooperation is part of a project conducted by the *Humanoid Robots* research group.
- The third application is currently being developed inside a European Project under the FP7 program called ARtSENSE [3], which main goal is to provide a more adaptive and attractive artwork experience in museum scenarios.

Therefore, Section 2 presents the energy detector as the optimum solution to detect unknown signals in presence of Gaussian and uncorrelated samples. Then, Section 3 describes the extensions for the non-Gaussian and non-independent case, as well as, a multiple energy detector structure to deal with the uncertainty of the signal duration. Section 4 presents the aforementioned projects which provide real-world scenarios to test the new energy detector algorithms proposed herein, and finally the conclusions of our work are summarized in Section 5.

2. Energy detector and the matched subspace filter

The detection theory is essential for the design of systems which implement an automatic processing of the signal for both decision making and information extraction. Examples of this kind of systems appear in communications, radar, biomedicine, image processing, etc. All of them share the common aim of being able to decide when an event of interest occurs and to determine as much information as possible about it. The detection theory is based on making a decision between two options from the available measures. In many of the typical applications mentioned, detection algorithms must decide between "noise only" or "signal masked with noise".

The problem resides in defining a decision rule that indicates which of two hypotheses should be chosen: hypothesis $\boldsymbol{H_0}$, where only noise is present, or hypothesis $\boldsymbol{H_1}$, which indicates the presence of a signal and noise. The decision rule can be represented by the following expression:

$$\Lambda(\mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [1]$$

where λ is the threshold and $\Lambda(\mathbf{y})$ is a function that depends on the measurements. If it exceeds the threshold, then $\boldsymbol{H_1}$ is selected; otherwise, $\boldsymbol{H_0}$ is decided. The aim of the detection theory is, hence, to design the most appropriate detector by defining $\Lambda(\mathbf{y})$ and λ. Taking into account the approach of Neyman Pearson, the search for the optimal detector is based on maximizing the probability of detection (PD) for a given probability of false alarm (PFA).

When the background noise is Gaussian distributed, several scenarios can be differentiated, each determined by the degree of uncertainty related to the knowledge of the signal characteristics. On one hand, based on the simplest example, where the desired signal is known and the noise is white and Gaussian, the optimal detector for this case is called the matched filter (MF). On the other hand, energy detectors are employed in automatically detecting signals in the presence of background noise when there is no exact knowledge of the signal waveform. Energy detectors have a very simple structure and are easily implementable. The ED is optimal when both signal and background noise are modeled as random uncorrelated Gaussian; or, it is at least a GLRT when the signal is deterministic and completely unknown [1].

Taking the above considerations into account, the hypotheses $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ can be expressed as follows:

$$\begin{aligned} &H_0\text{: } \mathbf{y} = \mathbf{w} && \mathbf{w}\text{: } N(0,\sigma_w^2\,\mathbf{I}) \\ &H_1\text{: } \mathbf{y} = \mathbf{x} + \mathbf{w} && \mathbf{x}\text{: } N(0,\sigma_w^2\mathbf{I}) \end{aligned} \qquad [2]$$

where $\mathbf{y}$ is the observation data vector. The signal vector is given by $\mathbf{x} = [x_1,\ldots,x_N]^T$, and it is modeled as a zero mean, white Gaussian random process. The vector $\mathbf{w} = [w_1,\ldots,w_N]^T$, is the white Gaussian noise with variance σ_w^2, and the test becomes:

$$\frac{\mathbf{y}^T\mathbf{y}}{\sigma_w^2} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [3]$$

which is known as the energy detector (ED), since it ultimately compares the energies of the measurements with a certain threshold λ. Notice that for notation simplicity λ is used in (3) as a general threshold and it will also be used in the following tests although it is calculated in a different manner in each case. It is intuitively deduced that when the signal is present, the energy of the received data $\mathbf{y}$ will increase. In fact, the test of the expression (3) can be considered as an estimator of the data variance and, after comparing it with a threshold, it can be decided whether it has a variance under hypothesis $\boldsymbol{H_0}\,(\sigma_w^2)$, or under hypothesis $\boldsymbol{H_1}\,(\sigma_w^2+\sigma_w^2)$.

In the previous cases, it was assumed that the noise was white. But sometimes this is not the case, and the implications of this fact must necessarily be examined. In this section, the energy detector mentioned above is studied when the noise is not white and the test is reformulated. The hypotheses are similar to those presented in (2). This time, however, the noise is characterized by $\mathbf{w}$: $N(0,\mathbf{R_w})$, where $\mathbf{R_w}= E[\mathbf{w}\mathbf{w}^T]$ is a general covariance matrix [4]. For the Gaussian case, independence and uncorrelation are equivalent, hence simple pre-whitening is enough. The original observation vector $\mathbf{y}$ is transformed into a pre-whitened observation vector $\mathbf{y}_p = \mathbf{R_w}^{-1/2}\,\mathbf{y}$. The ED is then applied to the pre-processed observation vectors, obtaining the following test denoted as pre-processed energy detector (PED):

$$\frac{\mathbf{y}^T\mathbf{y}_p}{\sigma_{w_p}^2} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad \longleftrightarrow \quad \mathbf{y}^T\mathbf{R_w}^{-1}\,\mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [4]$$

Notice that $\mathbf{R}_{\mathbf{w}_p}= E[\mathbf{w}_p\mathbf{w}_p^T]$ and, hence, σ_{wp}^2. So, we can verify that the transformation used in the pre-whitening yields independent samples and also normalizes the variance of the original noise observation vector.

Finally, the linear model will be considered here so as to introduce the special detection problem when the signal is unknown, but can be assumed to be included in a subspace. To briefly review the classical linear model, consider the following detection problem:

$$\begin{aligned} &H_0\!: \mathbf{y} = \mathbf{w} && \mathbf{w}\!: N(0,\sigma_w^2\,\mathbf{I}) \\ &H_1\!: \mathbf{y} = \mathbf{s} + \mathbf{w} && \mathbf{s}\!: \mathbf{H}\theta \end{aligned} \qquad [5]$$

where $\mathbf{y}$ is the observation vector in each hypothesis (dimension N), $\mathbf{w}$ is the noise background vector with zero-mean and Gaussian distribution, and, finally, $\mathbf{s}$ is the signal vector. It is assumed that $\mathbf{s}$ is defined in a subspace formed by $p<N$ columns of the known matrix $\mathbf{H}(p\times N)$, termed the observation matrix, and modeled by a vector of unknown parameters θ, as follows $\mathbf{s}\!: \mathbf{H}\theta$. As the signal

The energy detector (ED) implements a generalized likelihood ratio test (GLRT) when the noise is Gaussian; but when the noise is non-Gaussian and non-independent, the performance decreases.

is deterministic with unknown parameters, the detector used in this case is the matched subspace filter (MSF). This kind of detector is based on the estimated energy of the observation vector contained in the signal subspace by implementing the following test [1]:

$$\frac{\mathbf{y}^T\mathbf{H}^T(\mathbf{H}^T\mathbf{H})^{-1}\,\mathbf{H}^T\mathbf{y}}{\sigma_w^2} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [6]$$

where we can define $\mathbf{P}=\mathbf{H}^T(\mathbf{H}^T\mathbf{H})^{-1}\,\mathbf{H}^T$ as the projection matrix onto the subspace signal $\mathbf{H}$.

3. Extensions of the energy detector

The EDs are used to detect random signals in the presence of background noise when there is no information about the waveform to be detected. EDs provide optimal solutions to the detection problem when the signal and noise follow a Gaussian distribution with zero-mean, and the samples are uncorrelated. However, when the noise is not only non-Gaussian, but also non-independent, further extensions of the ED and the MSF are required. In addition, when there is no information about the signal duration, an important uncertainty is introduced. Therefore, novel detector solutions must be studied to face these problems as described below.

3.1. Non-Gaussian noise

When the noise characteristics differ from the assumption of Gaussianity, the ED is neither optimum nor does it implement a generalized likelihood ratio test (GLRT). Let us first consider the simplest case in which the components of the noise vector $\mathbf{w} = [w_0, w_1,...,w_{N-1}]^T$ are independent and identically distributed (i.i.d.), and sampled from a non-Gaussian distribution. As the ED is either optimum or GLRT when the noise is Gaussian, the use of a non-linear function $g(.)$ is proposed to convert a random variable having arbitrary distribution function $F_w(w)$ to a zero-mean and unit-variance Gaussian random variable. Thus, an extended energy detector (EED) is obtained by applying this transformation, denoted by $g(.)$, to every component of the observation vector $\mathbf{y}$ before computing the energy, and then by subsequently implementing the following test:

$$\frac{g(\mathbf{y})^T\,g(\mathbf{y})}{P_{g(w)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [7]$$

This test resembles the particularization of the Rao test for detecting unknown signals. $P_{g(w)}$ is the mean power of the transformed background noise and λ is the detec-

In the context of novelty or event detection, no information about the signal duration is available and, therefore, selecting the length of the observation vector is a very challenging task that may significantly affect the detection.

tion threshold fixed by the PFA selected. The non-linear transformation $g(.)$ in (7) can be implemented using different techniques. One of these is the well-known parametric Box-Cox transformation [5]; followed by a new non-linear transformation based on the estimation of the noise PDF which has to be transformed into Gaussian. In both cases, the Gaussianization process consists of performing a non-linear transformation of the data without memory to ensure that the distribution of values is as close as possible to a Gaussian function [6].

3.2. Non-independent noise

Both the ED and the EED assume i.i.d. components of the noise vector $\mathbf{w}$. When this is not the case, and the samples exhibit some form of dependence, it is necessary to apply additional pre-processing. As mentioned, independence and uncorrelation are equivalent for the Gaussian case; hence simple pre-whitening is sufficient and the pre-processed energy detector (PED) already presented can be utilized.

However, the statistical dependence problem for the non-Gaussian case is not so easily solved. The problem of calculating the PDF of a vector with correlated non-Gaussian samples has not been solved yet for many cases of study, so analytical expressions of the likelihood functions are not available for calculating the likelihood ratio (LR) or formulating Averaged LRs of GLRTs. One of the techniques used to reach vectors with independent components is to apply an independent component analysis (ICA) [7,8]. Essentially, ICA may be applied to yield an observation vector $\mathbf{y}_p$ with independent components by means of a linear transformation $\mathbf{U}$ of vector $\mathbf{y}$ as follows $\mathbf{y}_p=\mathbf{U}\mathbf{y}$, where matrix $\mathbf{U}$ not only pre-whitens, but also achieves statistical independence. Actually, the estimation of $\mathbf{U}$ is usually divided into two steps: the first decorrelates the elements of vector $\mathbf{y}$ by means of matrix $\mathbf{R}_\mathbf{w}^{-1/2}$, and the second yields the desired independence by means of a unitary transformation matrix named $\mathbf{Q}$ (equivalent to a rotation). As a result, the ED for non-Gaussian, non-independent noise is defined as follows:

$$g\left(\mathbf{Q}\mathbf{R}_\mathbf{w}^{-\frac{1}{2}}\mathbf{y}\right)^T g\left(\mathbf{Q}\mathbf{R}_\mathbf{w}^{-\frac{1}{2}}\mathbf{y}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [8]$$

It will be referred to hereafter as the pre-processed extended energy detector (PEED) [6]. Notice that, normalization by the noise mean-power is implied, as the non-linear transformation generates random zero-mean unit-variance Gaussian variables.

3.3. Generalization of the matched subspace filter

In this section, a further generalization of the MSF is proposed in order to consider the most general case of non-Gaussian and non-independent noise. First, our attention turns to the Rao test, since it is rather simple to adapt the general form of this test to the signal subspace detection problem in non-Gaussian noise assuming that the components of the noise vector $\mathbf{w}$ are i.i.d. random variables. Therefore, a non-linear transformation $g(.)$ is applied to the original observation vector prior to computing the normalized subspace energy. When the components of $\mathbf{w}$ are not i.i.d., the non-independent observation vector could be transformed into a new one having independent components. This can be done again by means of ICA with an appropriate linear transformation leading to a new generalization of the MSF that will be termed hereafter as generalized matched subspace filter (GMSF) [9]. Therefore, using $\mathbf{U}$ as the linear transformation to obtain i.i.d. vector noise samples with $\mathbf{u} = \mathbf{U}\mathbf{w}$, the GMSF is proposed:

$$\frac{g(\mathbf{U}\mathbf{y})^T\ \mathbf{P}_\mathrm{U}\, g(\mathbf{U}\mathbf{y})}{P_{g(u)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad [9]$$

where $P_{g(u)}=E[g^2(u)]$, and the pre-processed subspace matrix can be expressed as $\mathbf{P}_U=\mathbf{H}_U(\mathbf{H}_U^T\mathbf{H}_U)^{-1}\mathbf{H}_U^T$ with $\mathbf{H}_U=\mathbf{U}\mathbf{H}$.

3.4. Unknown signal duration

It is of particular importance to examine the relationship between the dimensions of the observation vector N and the behavior of the ED. To this end, when N is large and assuming white Gaussian noise, the chi-square distribution can be approximated by a Gaussian distribution having mean N and variance $2N$. Similarly, the non-central chi-square distribution can be approximated by a Gaussian one with a mean $SNRN+N$ and a variance $4SNRN+2N$. In consequence, the expression of the ROC for the ED when N is too large can be easily obtained, as in [1]:

$$PD \approx Q(Q^{-1}(PFA)\text{-}SNRN) \qquad [10]$$

where Q is the error function.

The term $SNRN=SNR/\sqrt{2N}$ is defined as a normalized SNR, and, taking into account (10), it can be clearly observed that, for a given PFA, the PD not only depends on the SNR, but also on the dimension N of the observation vector. Hence, for a specific SNR, if the signal duration is significantly smaller than N, the $SNRN$ (and in consequence, the PD) will be much lower than it would be with a choice of N that matched the actual signal duration.

To address this problem, a multiple energy detector structure (MED) is presented and analyzed in detailed in [10]. The MED was formed by multiple EDs matched to different signal durations and bandwidths, and the segmentation strategy followed in this case was based on several layers. *Figure 1* shows the proposed layered structure of

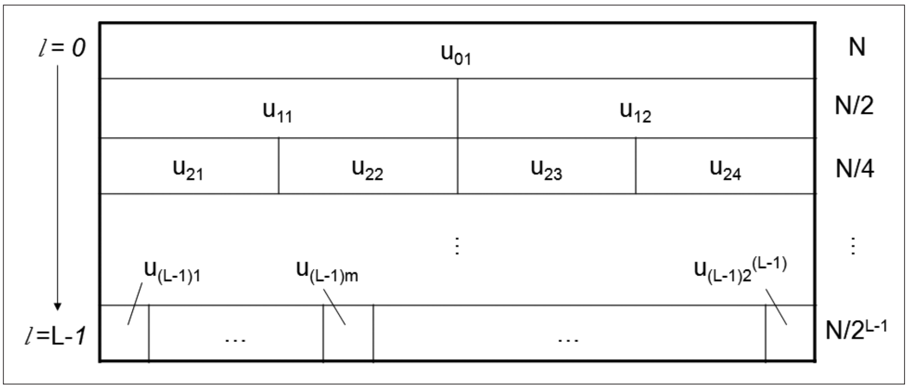


Figure 1. *Layered MED structure of L levels.*

the detectors where each layer was formed by successive segmentations (of a factor of 2) of the original observation vector. The presence of signal was determined if at least one of the ED decides it.

However, the analysis of this multiple energy detector structure is complicated by the fact that the individual decisions of each ED used in this structure ($\mathbf{u}_i$) are statistically dependent. This fact complicates the derivation of the overall PFA and PD of the MED structure; hence, it is necessary to carry out a theoretical study in order to obtain these expressions, as addresses in [10].

4. Use of energy detectors in acoustic scene analysis applications

In previous sections, the different extensions of the ED were presented; however, it is necessary to extend this study to the case of real-world signals and noises that may be found in different applications. There are a great number of areas in which the detection of unknown events is required. One of the most interesting fields of research is acoustic scene analysis, in which the signals recorded by a set of microphones are processed to extract as much information as possible about the environment.

4.1. HESPERIA Project

Part of the present work was developed within the framework of a collaborative research project between various companies and Spanish universities called HESPERIA (Homeland Security: Technologies for the Security in Public Spaces and Infrastructures) [2]. The objective of this consortium was to develop technologies that allow the creation of innovative security systems, video surveillance and operation control in private building and public places. The project sought to substantially increase the security of strategic infrastructures (electrical substations, water deposits, telecommunications centers) and of public places (airports, railway stations, ports, and urban environments like pedestrian areas, shopping centers, etc.). The "Grupo de Tramiento de Señal (GTS)" was tasked with acoustic monitoring of the environment with the aim of preventing dangerous situations by detecting, classifying and localizing suspicious events that might be enshrouded in a background noise.

The detection of such events has traditionally been attempted with techniques of video processing, but not with acoustics. However, the video systems observe the information in a certain direction for a specific instant of time, while acoustic processing allows one to listen in any direction at any instant of time. Therefore, by using the detection of acoustic events it is possible to overcome some of the technical deficiencies of the video systems. Some advantages are related to the immunity to lighting conditions, the ability to adapt to noisy and changing environments and, possibly the most relevant, the ability to detect events that take place in hidden areas outside the view of the security cameras. It is of paramount interest, therefore, to apply all the detection techniques described in the previous sections to the acoustic monitoring of surveillance environments. This will provide a cognitive audio system with the capability of automatically detecting unusual events in different scenarios, as shown in *Figure 2*.

Within the framework of this project, the detection and identification of possible dangerous events that might occur in the scene under analysis is of particular importance. Most of the earlier studies in the literature present the mel-frequency cepstral coefficients (MFCC) to be the most suitable features for speech and sound sources identification [11]. MFCC usually offers good performance, but the vulnerability of these features to noise degrades their recognition performance. Better features are generally desired for noisy environments and for that reason, we presented in [12] a noise robust feature extraction method to deal with this problem. The MED is capable of determining the presence of an acoustic event within a background noise; however, it can also provide information about it, which can then be employed for classification. Therefore, taking advantage of the MED, a modified MFCC extraction method was presented and some appropriate novel features were extracted by using the signature an event produced when it was processed by the MED. Furthermore, we also presented in [13] a detailed evaluation where the multiple energy detector was shown to provide significant improvement in the detection of

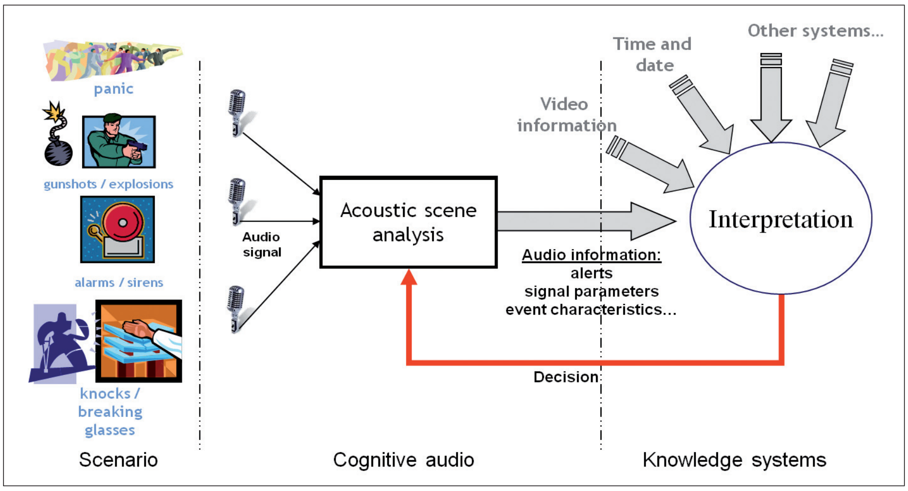


Figure 2. *General description of the surveillance system.*

Sound based systems are good candidates for evaluating the energy detector extensions presented; they define scenarios where both background noise and events may have variable and unpredictable characteristics.

acoustic events. In particular, the performance of the MED was studied in the time and frequency domains since detection of signals with unknown bandwidth is also of particular interest. The MED structure was also evaluated with real signals of unknown duration and bandwidth. As expected, the MED offered the best detection results in comparison to the single ED, when the acoustic signal duration or bandwidth was smaller than the original observation vector length. In addition, it was also demonstrated how this improvement was worthwhile, despite of the possible degradation resulting from the opposite case of using unnecessary layers of the MED.

4.2. HUMANOIDS robot

One of the most important areas in which the acoustic scene analysis is required is in the interaction between man and machine. Appropriate situations occur in scenarios where a human cooperates with a *humanoid robot*, or is assisted by one [14]. In this case, several active sound sources can exist in the robot's proximity, for example in a kitchen, which contains many different acoustically observable appliances. However, the localization of these sound sources is highly influenced by the presence of background noise and the unpredictable waveforms of the signals.

Thus, to resolve the two aforementioned issues, the "Grupo de Tratamiento de Señal (GTS)" in cooperation with the "Acoustic scene analysis group" of "Universität Karlsruhe" presented two novel localization approaches in [15] and [16]. They combine the information provided by an energy detector with the well-known localization method SRP-PHAT [17]. In [15], a microphone array was built according to the head geometry of a humanoid robot (*Figure 3*) and it is employed to localize dominant acoustic sources in a given noisy environment. This capability is successfully used in good SNR conditions, but its accuracy decreases considerably in the presence of other background noise sources. In order to counteract this effect, a novel approach is used: it implements a background noise suppression algorithm based on the ED to improve the localization method SRP- PHAT.

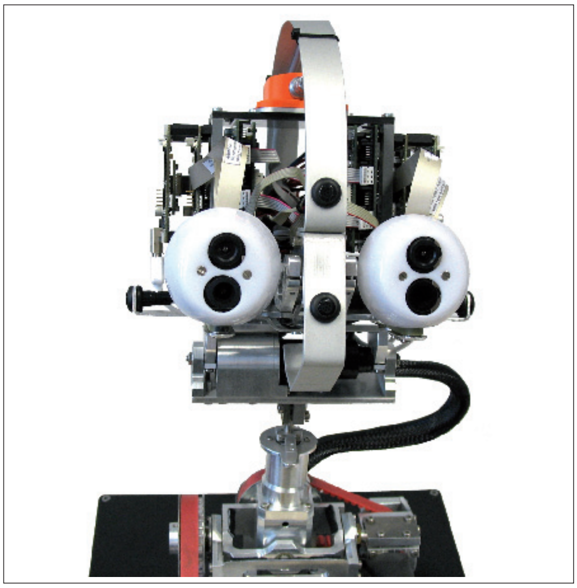


Figure 3. *Head of the humanoid robot ARMAR III*

The work developed in [16] lies within a humanoid robot application where a complete acoustic scene analysis is necessary to localize and detect all types of sound events which can occur in the proximity. Basically, two types of sound sources were differentiated: impulsive and non-impulsive. In many cases, only the non-impulsive ones, mainly *speech*, are taken into account. But, especially for the detection of dangerous or unusual situations, it is often necessary to localize and detect also impulsive sound sources like *slamming doors* or *breaking glass*. In order to be able to properly localize both types of events, a modification of the standard SRP-PHAT algorithm was

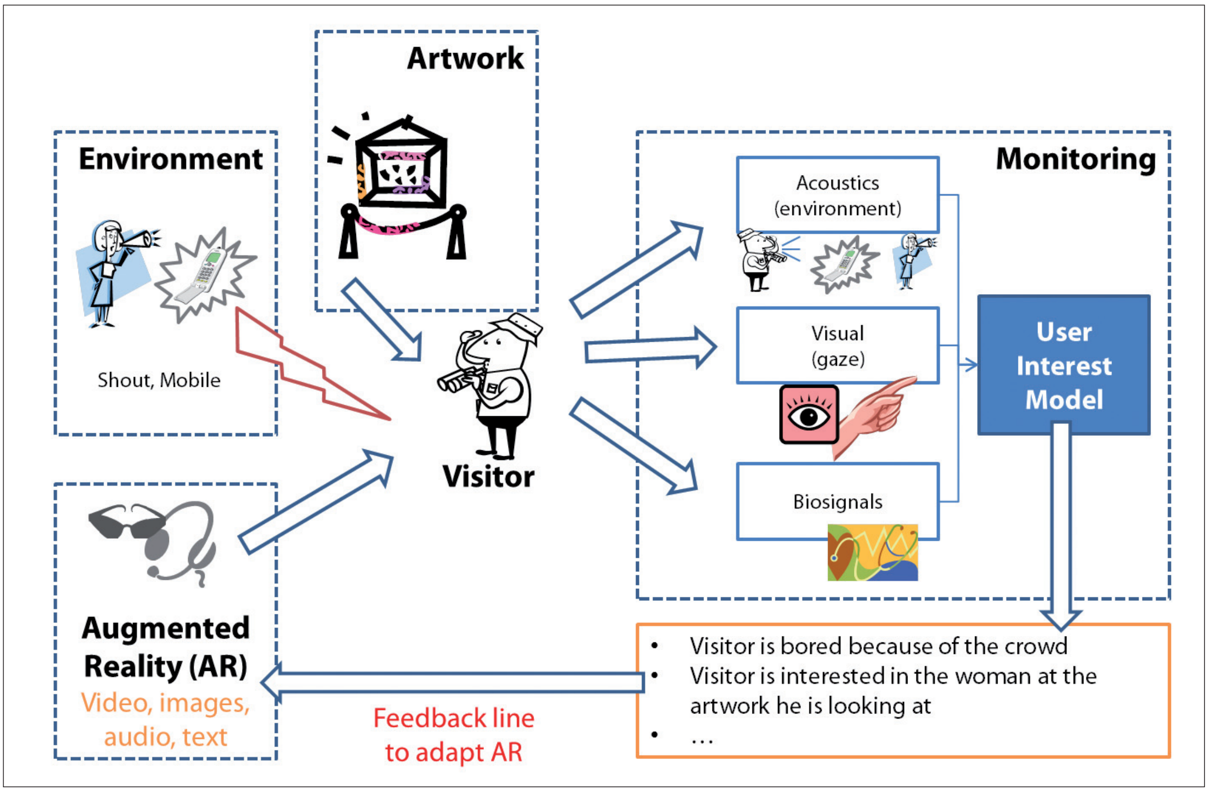


Figure 4. *Overview of ARtSENSE system.*

presented and a novel approach using an ED in a temporal event alignment and a pre-classification was proposed.

4.3. ARtSENSE Project

ARtSENSE [1] (Augmented RealiTy Supported adaptive and personalized Experience in a museum based oN processing real-time Sensor Events) is a European research project that involves nine European institutions, from technological centers and universities to private companies and cultural heritage institutions (CH) such as museums. The main objective of the project is to provide a personalized experience for all visitors by adapting the content presented by an augmented reality (AR) system taking into account the psychological state of the visitor. The consortium is formed by 3 European museums (end users) and 6 technical partners from different European countries: France, Spain, Germany, United Kingdom, Serbia and Hungary. The project therefore features an interdisciplinary and participatory design approach that leverages the synergy among CH professional and technology partners in a highly inclusive way.

Consequently, one of the objectives consists of determining the level of interest when the visitor is in front of an artwork. To reach this milestone, the system structure of *Figure 4* is presented. In order to obtain the user interest it is necessary to model the visitor and the environment in which he is immersed. This model considers different information sources such as gaze and gesture, acoustics and bio-signals in order to obtain the user attention. The acquisition of the different signals is carried out by three different technologies involved: acoustics, video and biosensing. In that sense, it is crucial to define a structure capable of relating the sensors and the information extracted from them in a suitable manner. Therefore, if the visitor is interested on the artwork, the AR system will provide more information and more details about it. On the contrary, if the visitor seems to be lacking interest, other artworks will be suggested and/or different content will be provided.

One of the main roles of the GTS in the ARtSENSE project is devoted to the acoustic scene analysis of the visitor's environment and the generation of the corresponding audio contents. To determine the user level of interest, it is important to know what is happening in his acoustic proximity since different kinds of sounds can occur: *people talking*, *mobiles ringing*, *crowd,* etc. These sounds can affect and disturb the visitor and therefore a set of omni-directional electret condenser microphones and a multichannel audio data acquisition unit worn by the visitor are used to capture the acoustic events and background noise. The digitalized signals are transferred in real time to the processing unit, which is in charge of applying the corresponding algorithms to extract valuable information about the event. These algorithms are divided in three stages. The first one consists of detecting the presence of any possible acoustic event that has occurred over the noise level. When a sound source is detected, the second stage is responsible of extracting more information about the event and to determine its level of disturbance. Finally, the third stage determines if the visitor has turned towards the acoustic event or not.

5. Conclusions

We have introduced the detection problem related to the amount of available knowledge about a signal and the background noise in which it is immersed. We reviewed the possible detection solutions which depended on the degree of knowledge available about the signal to be detected. The MF and the ED were presented as the optimal solutions for the detection of known deterministic and unknown signals, respectively. Furthermore, it was demonstrated that when the signal is unknown, but is present in a subspace, the MSF is used. In all these cases, the noise was assumed to be uncorrelated and Gaussian distributed.

The ED has been studied in three particular situations dealing with the non-Gaussianity of the noise, dependence of the noise sample and with the unknown duration of the event of interest. Therefore, the particular generalizations of the ED were presented to improve its performance under these conditions. When the noise is non-Gaussian, the Rao test was examined as a suboptimum solution to the detection problem but alternative non-linear functions were proposed, leading to an extended version of the ED, termed EED. Then, when the noise is not only non-Gaussian, but also non-independent, further extensions of the ED and the MSF were presented, leading to novel detector solutions termed PEED and GMSF, respectively. In both cases, ICA was applied to estimate the matrix transformation that makes as much i.i.d. as possible the components of the data vectors. And finally, it has been demonstrated that the signal duration incorporates an important uncertainty that influences the detection problem. Therefore, to address the problem of detecting unknown signals with undetermined duration, a novel approach based on the implementation of a MED structure was reviewed.

Finally, three real acoustic applications developed in the framework of some research projects have been presented where sound sources are not entirely known and thus, the design of an appropriate detector is more difficult. One of these projects is focus on surveillance systems by means of acoustic and image processing signals. This project is called HESPERIA and it was funded under a national CENIT program. The second project is the fruit borne of a cooperative exchange program with the "Universität Karlsruhe (TH)" under the HUMANOIDS robots project, where it has been shown the promising results achieved by combining the localization and classification of acoustic sounds with energy detectors. And finally, the last project lies within the European 7th framework Program and is called ARtSENSE, which main goal is to provide a personalized experience for visitors by adapting the content presented by an augmented reality system taking into account the state of the visitor. Consequently, it has been demonstrated the necessity of applying modern detection theory in order to design an efficient and robust acoustic detector capable of determining the presence of an event within a background noise. Furthermore, it has been shown the promising results achieved by combining the localization and classification of acoustic sounds with energy detectors.

References

[1] S. M. Kay, "Fundamentals of Statistical Signal Processing: Detection Theory", 1st Edition, HJ: Prentice-Hall, 1998.

[2] "Homeland sEcurity: teconologíaS Para la sEguridad integRal en espacIos públicos e infraestructuras", CENIT-2005.

[3] ARTSENSE (2011): http://www.artsense.eu

[4] Hippenstiel, R.D., "Detection theory: Applications and digital signal processing", Academic Press, 2001.

[5] G.E.P. Box and D.R. Cox, "An analysis of transformations", Journal of the Royal Statistical Society. Series B.General, vol. 26, n. 2, pp. 211-252, 1964.

[6] Moragues, J. and Vergara, L. and Gosálbez, J. and Bosch, I., "An extended energy detector for non-Gaussian and non-independent noise", Signal Processing, vol. 89, n. 4, pp. 565-661, 2009.

[7] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications", Neural Networks, vol. 13, n. 4-5, pp. 411-430, 2000.

[8] T.-W. Lee, "Independent component analysis: theory and applications", Kluwer Academic Publishers, 1st edition, 1998.

[9] Moragues, J. and Vergara, L. and Gosálbez, J., "Generalized matched subspace filter for non-independent noise based on ICA", IEEE Transactions on Signal Processing, vol. 59, n. 7, pp. 3430-3434, 2011.

[10] L. Vergara, J. Moragues, J. Gosalbez and A. Salazar, "Detection of signals of unknown duration by multiple energy detectors", Signal Processing, vol. 90, no 2, pp. 719-726, 2010.

[11] A. Dufaux, "Detection and recognition of impulsive sounds signals", Ph.D. dissertation, Faculté des sciences de l'Université de Neachâtel, Neuschatel, Switzerland, 2001.

[12] J. Moragues, A. Serrano, L. Vergara and J.Gosálbez, "Acoustic detection and classification using temporal and frequency multiple energy detector features", ICASSP 2011, Prague, (Czech Republic).

[13] Moragues, J. and Serrano, A. and Vergara, L. and Gosálbez, "Improving detection of acoustic signals by means of a time and frequency multiple energy detector", IEEE Signal Processing Letters, vol. 18, n. 8, pp. 458-461, 2011.

[14] Asfour, T. and Regenstein, K. and Azad, P. and Schröder, J. and Bierbaum, A. and Vahrenkamp, N. and Dillmann, R., ARMAR-III: An integrated humanoid platform for sensory-motor control, IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS), Genoa, Italy, 2006.

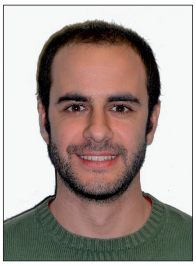
[15] Moragues, J. and Machmer, T. and Swerdlow, A. and Vergara, L. and Gosálbez, J. and Kroschel, K., "Background Noise Suppression for Acoustic Localization by Means of an Adaptive Energy Detection Approach", Proceedings of 33th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, USA, 2008.

[16] Machmer, T. and Moragues, J. and Swerdlow, A. and Serrano, A. and Vergara, L. and Kroschel, K., "Robust Impulsive Sound Source Localization by means of an

Energy Detector for Temporal Alignment and Pre-classification", EUSIPCO 2009, Glasgow (UK), 2009.
[17] DiBiase, J. H. and Silverman, H. F. and Brandstein, M. S., "Robust localization in reverberant rooms", Springer, pp. 157-180, Berlin, 2001.

Biographies

Dr. J. Moragues received his B.S. degree in Telecommunication Engineering and his M.S. degree in Technologies, Systems and Communications Networks, both from the Universitat Politècnica de València (UPVLC) in 2006 and 2008 respectively. In 2011 he received his Ph.D. in Telecommunications for his work in the development of new energy detector extensions with application in acoustic surveillance applications. He conducted his final project work at EADS-Ewation GmbH, Ulm (Germany) and he developed part of his Ph.D research studies within the acoustic scene analysis research group of the Universität Karlsruhe (TH) and the Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) of Karlsruhe, Germany. He is currently working within the Signal Processing Group (GTS) of the Institute of Telecommunications and Multimedia Applications (iTEAM), at UPVLC. His research interests focus on the theory and application of statistical signal processing mainly related to signal detection and array processing. Currently, all these advanced techniques are being used in several applications, particularly surveillance and monitoring systems based on audio signal processing.



Ing. A. Serrano was born in Cox, Alicante (Spain) in 1980. He received the Ingeniero de Telecomunicación degree and the M.S. degree in Technologies, Systems and Communications Networks from the Universitat Politècnica de València (UPV) in 2006 and 2008 respectively. He is currently member of the Institute of Telecommunication and Multimedia Applications (iTEAM) of UPV within the Signal Processing Group (GTS). His main interests are classification and image processing, more specifically in the field of change detection.

Ing. G. Lara was born in Valencia. He has recieved the Telecommunication Engineering degree from the Universidad Politécnica de Valencia (UPV) Spain in 2010. He is currently a Ph.D. student in the Institute of Telecommunication and Multimedia Applications (I-TEAM) of UPV. His research interest is focused in pattern recognition and statistical processing applied at the submarine acoustic. Currently he is envolved in the development of a submarine buoy capable of recording cetaceans sounds without losing samples, programming the electronics and the internal hardware.

Dr. J. Gosálbez. was born in Valencia (Spain) in 1975. He received Telecomunications Engineering and PhD degrees from the Universidad Politécnica de Valencia (UPV) in 2000 and 2004 respectively. He is Associate Professor at Departamento de Comunicaciones (UPV) and member of the Signal Processing Group of the Institute of Telecommunication and Multimedia Applications (I-TEAM) of UPV. His research concentrates in the statistical signal processing area, where he has worked in different theoretical an applied problems. His theoretical aspects of interest are time-frequency analysis, signal detection and array processing. Currently he is involved in ultrasound signal processing for non-destructive evaluation of materials, in surveillance systems based on acoustic information and in acoustic source location and tracking based on sensor and array signal processing. He has published more than 50 papers including journals and conference contributions.

Prof. L. Vergara. was born in Madrid (Spain) in 1956. He received the Ingeniero de Telecomunicación and the Doctor Ingeniero de Telecomunicación degrees from the Universidad Politécnica de Madrid (UPM) in 1980 and 1983 respectively. Until 1992 he worked at the Departamento de Señales, Sistemas y Radiocomunicaciones (UPM) as an Associate Professor. In 1992 he joined the Departamento de Comunicaciones (Universidad Politécnica de Valencia UPV, Spain), where he became Professor and where it was Department Head until April 2004. From April 2004 to April 2005 he was Vicerector of New Technologies at the UPV. He is now responsible of the Signal Processing Group of the UPV, a member group of the Institute of Telecommunication and Multimedia Applications (I-TEAM) of UPV. His research concentrates in the statistical signal processing area, where he has worked in different theoretical an applied problems, many of them under contract with the industry. His theoretical aspects of interest are signal detection and classification, independent component analysis and spectral analysis. Currently he is involved in ultrasound signal processing for non-destructive evaluation, in infrared signal processing for fire detection and in cognitive audio for surveillance applications. He has published more than 150 papers including journals and conference contributions.