Universitat Politècnica de València
Departamento de Comunicaciones

New energy detector extensions with application in sound based surveillance systems

DOCTORAL THESIS

Jorge Moragues Escrivá

Directed by

Dr. Luis Vergara Domínguez

Dr. Jorge Gosálbez Castillo

July 2011

New energy detector extensions with application in sound based surveillance systems

A dissertation presented to the "Departamento de Comunicaciones" to obtain the degree of "Doctor en Telecomunicación" by

Jorge Moragues Escrivá

Directed by

Dr. Luis Vergara Domínguez

Dr. Jorge Gosálbez Castillo

Universitat Politècnica de València

Departamento de Comunicaciones

July 2011

To Sandra.

All our dreams can come true,
if we have the courage to pursue them.
Walt Disney

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my director, Prof. Dr. Luis Vergara, for his continuous support of my Ph.D. study and research as well as for his patience, motivation, enthusiasm, and immense knowledge. His understanding and personal guidance have been of great value for me. For his constructive comments and for his important support lent to this work, I also wish to warmly thank my director, Dr. Jorge Gosálbez, the person who introduced me to the world of research.

I wish to express my sincere thanks to Dr. Ramón Miralles and Dr. Ignacio Bosch for their trust and generous support. I would also like to thank Prof. Dr. Kristian Kroschel for offering me the internship opportunities in his group at "Universität Karlsruhe" (Germany) where I participated in diverse exciting projects together with other Ph.D. students.

I am deeply grateful to Arturo for his valuable advice and friendly help. He has made his support available to me in a number of ways during the past years, and this thesis would not be the same without his comments and suggestions. I am also grateful to all my colleagues at "Grupo de Tratamiento de Señal" with whom I have shared many experiences– especially, Gonzalo and Mª Angeles. With their company and friendship, my days were made more enjoyable.

My deepest gratitude goes to my family and friends for their unflagging love and support throughout my life; this dissertation would be simply impossible without them. I am indebted to my parents, Pepe and Mila, who spared no effort to provide the best loving environment for me to grow up in. My sincerest thanks goes to my sister Patricia, for her constant support and love during all these years, and for the detailed review, constructive criticism and excellent advice during the preparation of this thesis. I also wish to thank my aunt, Rosana; she always took care of me like a son and her advice was consistently timely and useful. My warm gratitude to my grandfather, Casimiro, for his prays and for encouraging me to do my best in all matters of life.

Last, but not the least, I owe my loving thanks to my wife Sandra. She encouraged, supported, understood, and loved me at every moment. None of this could have happened without her. This thesis is dedicated to her.

Resumen

Esta tesis está dedicada al desarrollo de nuevos detectores de energía para la detección de señales desconocidas en presencia de ruido no gaussiano con muestras no independientes. Para ello, se ha llevado a cabo un amplio estudio de las diferentes estructuras de detección existentes basadas en energía y se han propuesto nuevas técnicas capaces de resolver este tipo de situaciones.

El detector de energía se presenta como la solución óptima para la detección de señales gaussianas no correladas, o como un test de razón de verosimilitud generalizado cuando las señales son completamente desconocidas. En ambos casos, el ruido de fondo debe ser gaussiano e incorrelado. Sin embargo, su comportamiento se degrada cuando el ruido de fondo no cumple estas características. En primer lugar se proponen dos extensiones, por un lado el detector de energía extendido cuando el ruido es no gaussiano, y por otro el detector con preprocesado de energía extendido cuando además es no independiente. En segundo lugar, se presenta una generalización del filtro adaptado a subespacio basada en una modificación del test de Rao. Para evaluar el comportamiento de estas extensiones con respecto al detector de energía clásico, se define un factor de ganancia que ilustra las mejoras logradas en detección.

A continuación, se demuestra como el desconocimiento de la duración de la señal puede deteriorar el comportamiento del detector. Para hacer frente a este problema, se presenta una novedosa estructura de múltiples detectores de energía basada en la subdivisión sucesiva del intervalo de observación original. Esta técnica de detección nos conduce a una estructura en capas con varios detectores de energía cuyos vectores de observación se ajustan a diferentes intervalos de duración de la señal. Se ha desarrollado todo el análisis requerido para el cálculo de las correspondientes probabilidades de falsa alarma y de detección para una estrategia de subdivisiones particular, estableciendo los procedimientos necesarios para su generalización a otros casos. Las simulaciones realizadas muestran las ventajas de utilizar la nueva estructura con respecto a un solo detector cuando la duración real de la señal y el intervalo de observación original son diferentes.

Los buenos resultados alcanzados en las simulaciones han permitido el empleo de dichos detectores en aplicaciones reales de vigilancia basada en

sonido. Estos sistemas presentan un interesante ámbito de aplicación donde es posible comprobar la robustez de los detectores analizados en los desarrollos teóricos de esta tesis. Para ello, se han analizado varias fuentes acústicas en diferentes escenarios de ruido de fondo, tanto reales como simulados, presentándose además dos enfoques novedosos. En el primero de ellos, la información proporcionada por el detector de energía se combina con un método de localización sonora de fuentes. Gracias a la utilización de esta nueva técnica, se demuestra como la localización de sonidos en presencia de ruido de fondo mejora considerablemente. Por último, un nuevo conjunto de características extraídas a partir de la estructura de múltiples detectores son evaluadas y comparadas con otras características comúnmente utilizadas en el reconocimiento de sonidos acústicos. De nuevo, los resultados obtenidos con las nuevas características ofrecen mejores probabilidades de acierto en la clasificación, especialmente en baja relación señal a ruido.

Resum

Aquesta tesi està dedicada al desenvolupament de nous detectors d'energia per a la detecció de senyals desconeguts en presència de soroll no gaussià i les mostres del qual són no independents. Per a això, s'ha dut a terme un ampli estudi de les diferents estructures existents en detecció basada en energia i s'han proposat noves tècniques capaces de resoldre aquest tipus de situacions.

El detector d'energia es presenta com la solució òptima per a la detecció de senyals gaussians no correlats o com un test de raó de versemblança generalitzat quan els senyals són completament desconeguts. En ambdós casos, el soroll de fons ha de ser gaussià i incorrelat. No obstant això, el seu comportament es degrada quan el soroll de fons no complix aquestes característiques. En primer lloc es proposen dues extensions, per un costat el detector d'energia estés quan el soroll és no gaussià i per un altre el detector amb preprocessat d'energia estés quan a més és no independent. En segon lloc, es proposa una generalització del filtre adaptat a subespai basada en una modificació del test de Rao. Per a avaluar el comportament d'aquestes extensions respecte al detector d'energia clàssic, es definix un factor de guany que il·lustra les millores en detecció aconseguides.

D'altra banda, es demostra com el desconeixement de la duració del senyal pot deteriorar el comportament del detector. En aquest sentit, per a millorar aquest comportament, es presenta una estructura de múltiples detectors d'energia basada en la subdivisió successiva de l'interval d'observació original. Aquesta nova tècnica de detecció ens conduïx a una estructura en capes amb diversos detectors d'energia els vectors d'observació de la qual s'ajusten a diferents intervals de duració del senyal. Les corresponents probabilitats de falsa alarma i de detecció es calculen per a una estratègia de subdivisions particular establint els procediments necessaris per a la seua generalització a altres casos. Les simulacions realitzades mostren els avantatges d'utilitzar la nova estructura respecte a un sol detector quan la duració real del senyal i l'interval d'observació original són diferents.

Els resultats bons aconseguits en les simulacions han permés la utilització dels esmentats detectors a aplicacions reals de vigilància basada en so. Aquests sistemes presenten un interessant àmbit d'aplicació on és possible com-

provar la robustesa dels detectors analitzats en els desenvolupaments teòrics d'aquesta tesi. Per a això, diverses fonts acústiques s'han analitzat en presència de diferents escenaris de soroll de fons, tant reals com simulats, presentant a més a més dos enfocaments novedosos. En el primer d'ells, la informació proporcionada pel detector d'energia es combina amb un mètode de localització sonora de fonts. Gràcies a la utilització d'aquesta nova tècnica, es demostra com la localització de sons en presència de soroll de fons millora considerablement. Per últim, un nou conjunt de característiques extretes a partir de l'estructura de múltiples detectors són avaluades i comparades amb altres característiques comunament utilitzades en el reconeixement de sons acústics. De nou els resultats obtinguts amb les noves característiques oferixen millors probabilitats d'encert en la classificació, especialment en baixa relació senyal a soroll.

Abstract

This thesis is dedicated to the development of new energy detectors employed in the detection of unknown signals in the presence of non-Gaussian and non-independent noise samples. To this end, an extensive study has been conducted on different energy detection structures, and novel techniques have been proposed which are capable of dealing with these problematic situations.

The energy detector is proposed as an optimum solution to detect uncorrelated Gaussian signals, or as a generalized likelihood ratio test to detect entirely unknown signals. In both cases, the background noise must be uncorrelated Gaussian. However, energy detectors degrade when the noise does not fulfill these characteristics. Therefore, two extensions are proposed. The first is the extended energy detector, which deals with the problem of non-Gaussian noise; and the second is the preprocessed extended energy detector, used when the noise also possesses non-independent samples. A generalization of the matched subspace filter is likewise proposed based on a modification of the Rao test. In order to evaluate the expected improvement of these extensions with respect to the classical energy detector, a signal-to-noise ratio enhancement factor is defined and employed to illustrate the improvement achieved in detection.

Furthermore, we demonstrate how the uncertainty introduced by the unknown signal duration can decrease the performance of the energy detector. In order to improve this behavior, a multiple energy detector, based on successive subdivisions of the original observation interval, is presented. This novel detection technique leads to a layered structure of energy detectors whose observation vectors are matched to different intervals of signal duration. The corresponding probabilities of false alarm and detection are derived for a particular subdivision strategy, and the required procedures for their general application to other possible cases are indicated. The experiments reveal the advantages derived from utilizing this novel structure, making it a worthwhile alternative to the single detector when a significant mismatch is present between the original observation length and the actual duration of the signal.

The important simulation results yielded by the new energy detectors offer promising opportunities for real-world applications, such as surveillance systems based on sound analysis. These systems present a suitable scope for verifying the robustness of the theoretical detectors presented in this thesis. Thus, several acoustic sources and a variety of real and simulated noise scenarios were tested and two novel approaches were presented. The first combines the information provided by an adaptive energy detector with the standard localization method. The localization rates are considerably improved with this original technique, mainly when the sound source is in presence of a background noise. Finally, a unique set of features are extracted from the multiple energy detector structure, evaluated, and compared with other common features used for the recognition of acoustic sounds. The results obtained with the new features considerably improve the classification accuracy, especially in low signal-to-noise ratios.

Contents

List of Figures

List of Tables

Part I

Preamble

Chapter 1

Introduction

Only those who will risk going too far
can possibly find out how far one can go.

T.S. Eliot

The goal of this thesis is to study the feasibility of detecting unknown 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 techniques, extensions, and even new structures for improving the robustness in detection are considered and explained.

1.1 Motivation and objectives

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 [43]. 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 theoretical 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; 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, the Gaussian model cannot be considered. Moreover, it is important to model the noise more accurately, as failure to do so can lead to deterioration in the performance of the detector. For this reason, it is also advantageous to study the detectors in the presence of noise with a non-Gaussian distribution. Similarly, another problem that may often arise is the dependence of the samples, which is particularly complex to deal with scenarios where the noise is not Gaussian. In this case, it is possible to apply several extensions to the classical ED to improve the behavior of the detector. Thus, **the first objective of this thesis is to obtain extensions of the energy detector appropriate to this general case of non-Gaussian and non-independent noise samples.**

It is 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. In these cases, for example, the size of the observation interval is defined by the practical implementation constraints related to computational requirements, the permitted delay before deciding or the non-stationary behavior of the background noise. Consequently, 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. **The second objective of this thesis addresses this issue by deriving procedures and structures focused on the use of energy detectors, and capable of dealing, in an optimal manner, with signals of unknown duration and bandwidth.**

Finally, this thesis will also focus on the development of new technologies in a growing application area: security and surveillance systems, particularly those systems which incorporate techniques to automatically detect events or novelties in a monitored environment. Recently, surveillance based on sound analysis has become increasingly important and has been proposed for use in a variety of contexts. 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 Tramiento de Señal (GTS)" in the "Universitat Politècnica de València (UPV)" for incorporation within the framework of different projects. Sound offers surveillance in situations where cameras are blind (e.g., hidden people, low lightning conditions and blind spots), and in situations where the images are apparently normal, but the sounds could be abnormal. Moreover, sound may be used in combination with video to improve general performance through information fusion, or to index the video stream. Also, when compared to video, sound captures different levels of privacy when surveillance has to be conducted in public sites. In this thesis, we will consider two related applications currently under development by the GTS:

- The first falls within the field of security in public places. Video information is complemented by simultaneous audio information gathered by distributed arrays of microphones. The acoustic events present in the acoustic scene are detected, classified and, finally, localized. The detection phase in this process is fundamental to achieving higher quality results in the latter two phases.

- The second application lies within the framework of a cooperation program from the "Grupo de Tratamiento de Señal (GTS)" of "Universitat Politècnica de València (UPV)" and the "Acoustic scene analysis group" of "Universität Karlsruhe (TH)" (Germany). The cooperation is part of a project conducted by the *Humanoid Robots* research group. As the goal of the project is to design service robots to assist humans, analysis of the environment scene is required. The system is based on cameras and microphones which attempt to emulate human abilities by implementing subsequent signal processing stages.

The two aforementioned projects provide real-world scenarios to test the new energy detector algorithms proposed herein. Hence, **the third objective of the thesis is the application of the new extended energy detectors to surveillance and sound-based monitoring systems.**

1.2 State of the art

Energy detectors are employed in automatically detecting signals in the presence of background noise when there is no exact knowledge of the signal waveform [43]. In a more general context, energy detection is of interest in detecting departures from a known background due to imprecisely defined changes (event or novelty detection) [56]. The energy detector 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 [43].

The original energy detector proposed in [86] dealt with the detection of unknown deterministic signals in Gaussian noise. This detector was later extended in [47] to detect random signals corrupted by Gaussian noise. In [15], an improved energy detector for random signals in Gaussian noise is presented. A great deal of research has been devoted to obtaining generalizations about the energy detector when the background noise is Gaussian and non-independent (colored), or non-Gaussian and independent. The former is usually solved by means of pre-whitening transformations [67], while the latter is addressed by a scalar non-linear function applied to every component of the observation vector, subsequently followed by summation of all the components. This second solution is most germane to the present thesis. To wit, a number of alternative non-linear functions have been proposed [40, 77], and have lead to the creation of different detectors commonly termed generalized energy detector (GED). In light of this previous work, a new scalar non-linear transformation is herein proposed and can be thus considered a new variation of the GED. However, the statistical dependence problem present in the non-Gaussian case requires greater attention. Moreover, the literature reveals a paucity of efforts directed at uncovering a solution. In contrast, the search for linear transformations to achieve vectors with independent components has received much greater attention in recent years and has yielded various techniques [36, 50]. In fact, independent component analysis (ICA) is currently being applied to multiple types of problems, such as blind source separation (BSS); it is, therefore, a natural candidate for pre-processing the original observation vectors in a detection context where the original noise samples are dependent. In other words, ICA may be applied to obtain vectors with independent components by means of a linear transformation.

The problem of detecting unknown signals of unknown duration has been previously considered in [45] using a GLRT approach in which the starting sample of the signal is known, but the signal length represents an unknown parameter. It is shown that, assuming white Gaussian noise, the maximum likelihood estimate of the signal duration always coincides with the whole duration of the original observation interval. Therefore, this new approach

does not offer any solution to the detection problem of unknown signal duration; thus, the ED still remains the GLRT solution. If the starting time of the signal is set as an unknown parameter, a straightforward extension to the previously-posed problem is generated; however the logical conclusion remains unavailing. In [45], a new modification of the energy detector is proposed, based on the theory of embedded probability density functions [44]. The method consists of applying the energy detector to every possible interval of the original observation vector and, subsequently, the corresponding statistic (the normalized energy) is transformed using the non-linear function proposed. The maximum of the transformed statistic is then compared with a threshold. When the assumed signal duration is increasing, the non-linear function avoids the systematic growth of the likelihood. This aspect points to the earlier conclusion regarding the GLRT solution. Unfortunately, the non-linear function complicates the task of deriving closed expressions for the probability of detection (PD) and the probability of false alarm (PFA). A different approach is thus proposed in this thesis to overcome this complexity problem brought on by the non-linear function. Instead of employing only a single detector, whose results are complex statistics based on the non-linear transformed energy for every possible signal duration, the implementation of multiple energy detectors matched to different possible signal durations is proposed.

Numerous areas require the detection of unknown events. One of the most interesting areas being research today is acoustic scene analysis, devoted to surveillance applications in which the signals recorded by a set of microphones are processed to extract as much information as possible from the environment. In [3], processing audio signals yields the detection of different human activities, such as shouting, talking, walking, and crying. In [17] and [68], public spaces and metro stations were selected to test automatic detection of potentially dangerous acoustic events. Moreover, under particularly adverse lighting conditions, with hidden objects and blind spots, for example, it is possible to use acoustic sensors to cull information not captured by video sensors [58, 91]. Other related studies have utilized acoustic event detection to determine the presence of sounds in real life scenarios [1]. Furthermore, earlier studies were conducted to evaluate the detection performance in the frequency domain [13], as well as in applications (e.g., echo detection in radar, sonar, and acoustics) where an approximate idea of the signal is obtainable, assuming knowledge of its bandwidth [16]. Despite these advances, much of this research does not take into account the noise characteristics and the actual duration of the signal, in turn, the segment length of the observation vector).

The localization and classification of sound sources has been recently applied to surveillance applications in which the use of audio sensors is becoming increasingly decisive [72, 87]. However, much of the recent work

in these research areas disregards the presence of background noise. This critical factor normally decreases the performance of the localization and classification phases, thereby leading to an increase in false localizations and to a poor recognition rate.

With regards to the classification phase, most of the earlier studies are based on supervised classifiers such as Gaussian mixture model (GMM) and support vector machine (SVM). Some examples can be found in home applications [35] and public transport scenarios [73], as well as in surveillance environments, as detailed in [29]. Other unsupervised classifications methods, such as the classical k-means, may likewise be employed [33]. Turning to the features selection, previous studies have presented the mel-frequency cesptral coefficient (MFCC) as the most suitable features for speech and sound source identification [27]. MFCC typically offers good performance, but the noise vulnerability of these features degrades the recognition performance; thus, more appropriate features for noisy environments are generally desired. In [92], for example, an attempt was made to address this difficulty with a noise-robust feature extraction method. However, in overcoming this hurdle, this thesis has turned to a new direction; and thus will present a new set of features extracted from a multiple detection structure.

The localization of acoustic sounds is widely employed in surveillance applications for intrusion detection [93], and different techniques have been presented in the literature [11], most of them based on the estimation of time difference of arrival (TDOA) for each microphone pair [46]. When several microphones pairs are available, the SRP-PHAT algorithm is used [24], but the background noise can considerably decrease the localization results. In the present work, an adaptive ED is employed in combination with this last technique to suppress the background noise effect and, at the same time, to take into account events of a varying nature.

1.3 Organization of this research

1.3.1 Structure

In this thesis, emphasis is placed on the detection problem of unknown signals in the presence of background noise of unknown characteristics. The following chapters will review the most pertinent principles and theories involved, and then will proceed with a detailed description of the numerous experiments realized using simulated and real signals.

Chapter 2 reviews the possible detection methods available for use when the signal to be detected is known, but complete knowledge about its characteristics is not at hand. Two different detectors are presented, the matched filter and the energy detector, and they are analyzed in the presence of Gaussian uncorrelated and correlated noise.

In Chapter 3, the problem of detecting unknown signals in non-Gaussian noise conditions is analyzed, with special interest given to the case of dependent noise samples. The use of non-linear functions and independent component analysis is studied, and alternative extensions of the energy detector and the matched subspace filter are presented. The performance of these extensions is analyzed by means of the receiver operating characteristic (ROC) curves. Following this, the problem of detecting signals of unknown duration is studied in Chapter 4. A novel structure of multiple detectors is presented and analyzed, including ROC curves for different signal durations in time and frequency.

Acoustic event detection is the subject of Chapter 5, where the robustness of a complete detection system is evaluated. The behavior of the proposed energy detector and its extensions is studied to detect sound sources in the presence of real background noise whose characteristics differ from Gaussian and independent assumptions. Chapter 6 studies diverse possibilities for combining single and multiple energy detectors with other techniques, such as the classification and localization of sound sources. The improvements achieved by coupling the aforementioned techniques with the energy detector are shown in different acoustic applications and scenarios.

Finally, Chapter 7 presents the principal conclusions reached, along with future lines of research.

1.3.2 Main contributions

The chief scientific contributions are found in the innovative detection techniques presented and their combination with well-known classification and localization algorithms for subsequent application to real acoustic scenarios. They are listed below:

- Energy detectors degrade when the noise is non-independent and non-Gaussian. Therefore, two extensions, the extended energy detector (EED) and the pre-processed extended energy detector (PEED), are presented in this work to deal with these situations. Independence is achieved by means of a linear matrix transformation derived from ICA. Non-Gaussianity is avoided by applying a scalar non-linear function to every element of the linearly transformed observation vector. Practical procedures for estimating the linear and nonlinear transformations are given in this work. A signal-to-noise ratio (SNR) enhancement factor has been defined for the weak signal case, which is indicative of the expected improvement of the proposed extension of the energy detectors.

- A generalization of the matched subspace filter (MSF) termed as GMSF is proposed for the detection of unknown signals in a background of non-Gaussian and non-independent noise. The generalization is based

on a modification of the Rao test by including a linear transformation derived from ICA. The ROC curves computed for simulated examples illustrate the significant improvements achieved with the generalized solution.

- An extension of the classical ED is proposed to deal with the case of unknown signal duration. Multiple energy detectors (MED) are applied to subintervals of the original observation interval, and a global criterion is establish to discern the signal presence. A specific strategy of employing successive half segmentations of the original interval is applied to obtain a layered structure of energy detectors. Moreover, the corresponding probabilities of false alarm and detection are derived. The novel structure shows significant improvements in detectability in comparison with the ED when there is large mismatch between the original observation length and the actual duration of the signal.

- The EED and the PEED are evaluated in real acoustic scenarios. These findings are further extended by considering a new generalization based on the rejection of the false alarms induced by the difficulty of estimating the noise statistic in non-stationary conditions. The evaluation of the different detectors has been conducted by measuring their performance in terms of the ROC curves. Several acoustic sources and a variety of real and simulated noise scenarios were tested. It is demonstrated that the generalization approach works significantly better than the ED in low SNR, but in real scenarios there is a loss in detectability. In such non-Gaussian and non-independent noise conditions, the proposed generalization of the EED and the PEED improve performance, leading to a significant enhancement in the event detection.

- A MED structure is applied to both the time and frequency domains, and is subsequently evaluated in real acoustic scenarios. The results obtained demonstrate the robustness of the MED structure and the improvements in performance reached versus the ED.

- Two novel approaches are presented for the detection and localization of impulsive and non-impulsive sound sources in the presence of non-stationary background noise. Both approaches combine the information provided by an adaptive ED with the standard SRP-PHAT localization method. The first modified approach uses the ED to calculate the noise correlation; the second distinguishes between impulsive and non-impulsive sound sources and, additionally, aligns the detection window to the event. In both cases, the localization rates are considerably improved with the novel techniques.

- A MED structure is applied to extract a new set of features for classification: time MED (TMED) and frequency MED (FMED). The

combination of these two types of features leads to the combined MED (CMED) features. These novel feature sets are compared with the commonly used MFCC features, and their performance is evaluated in a general sound classification task with different acoustic sources and adverse noise conditions. It is shown that, in low SNR, the proposed CMED features work significantly better than the MFCC.

Chapter 2

Energy detector

A man can succeed at almost anything
for which he has unlimited enthusiasm.

Charles Schwab

This chapter offers a review of the possible detectors that can be used when the signal to be detected is known and when complete knowledge about its characteristics remains unavailable. In both cases, the probability density function (PDF) of the background noise, in which the signals are immersed, is assumed to be Gaussian distributed. Furthermore, the statistical dependence problem of the noise samples is also studied within this framework and the performance of the energy detector under these scenarios is discussed.

2.1 Detectors design

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 are communication systems, 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, such as radar systems, 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 H_0, where only noise is present,

or hypothesis H_1, which indicates the presence of a signal and noise. The decision rule can be represented by the following expression:

$$\Lambda(\mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{2.1}$$

where λ is the threshold and $\Lambda(\mathbf{y})$ is a function that depends on the measurements. If it exceeds the threshold, then H_1 is selected; otherwise, H_0 is decided. The aim of the detection theory is, hence, to design the most efficacious detector by defining $\Lambda(\mathbf{y})$ and λ. For that purpose, two approaches are described below.

2.1.1 Bayes detector

The aim of the Bayes detector is to minimize the mean cost, whose expression can be calculated as:

$$C = C_{10}P(C_{10}) + C_{01}P(C_{01}) + C_{11}P(C_{11}) + C_{00}P(C_{00}) \tag{2.2}$$

where C_{ij} is the cost of deciding hypothesis i, when hypothesis j is the correct option. From this expression it is possible to derive the decision rule, expressed as [43]:

$$\frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{C_{10}P_0}{C_{01}P_1} \tag{2.3}$$

where λ takes the value $\frac{C_{10}P_0}{C_{01}P_1}$; the function $\Lambda(\mathbf{y}) = \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})}$ is called the likelihood ratio; and $f_i(\mathbf{y})$ is the probability density function (PDF) of $\mathbf{y}$ under hypothesis H_i. The logarithm of this expression is often used, obtaining $\ln(\Lambda(\mathbf{y}))$. Thus, it is necessary to know the PDF of $\mathbf{y}$ under the hypotheses H_0 and H_1, and then to compare their ratio with the threshold in order to implement the optimal Bayes detector.

Some extensions of optimal Bayes signal detection with known prior statistics are given in [69], where the likelihood function is extended to include overlapping signal classes, which occur in adaptive detection and in stochastic signal detection.

2.1.2 Neyman Pearson detector

The Neyman-Pearson detector [66] follows a different philosophy than that of the Bayes detector. In this kind of detector, first, a probability of false alarm (PFA) is fixed and then, the probability of detection (PD) is maximized. Following a similar procedure to the one performed for the Bayes detector, we can reach the following equation:

$$\frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{2.4}$$

where $\Lambda(\mathbf{y})$ is, again, the likelihood ratio. But, to obtain the threshold λ, it is first necessary to apply the restriction imposed by the required maximum PFA [43]:

$$PFA = \int_{\lambda}^{\infty} f_0(\Lambda) d\Lambda = \alpha. \tag{2.5}$$

Therefore, with this detector we can maximize the PD for a given PFA by selecting the detection threshold obtained from the restricted $PFA = \alpha$.

2.2 Signal detection in presence of Gaussian noise

When the background noise is Gaussian distributed (see Appendix A.1), several scenarios can be differentiated, each determined by the degree of uncertainty related to the knowledge of the signal characteristics. First, a case is considered in which the signal to be detected is perfectly known and deterministic in the presence of white Gaussian noise (WGN). Then, the detection problem is studied with signals that are totally unknown and modeled as a random process. The detection of both known and unknown signals is considered in the presence of Gaussian noise, but with correlated components. Finally, the detection of signals in a subspace is studied.

2.2.1 Deterministic signal

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). By applying the criterion of Neyman-Pearson (NP), the resulting test will consist of differentiating between the following hypotheses [43]:

$$\begin{array}{ll} H_0 : \mathbf{y} = \mathbf{w} & \mathbf{w} : N(0, \sigma_{\mathbf{w}}^2 \mathbf{I}) \\ H_1 : \mathbf{y} = a\mathbf{s} + \mathbf{w} & a > 0, \mathbf{s}^T\mathbf{s} = 1 \end{array} \tag{2.6}$$

where $\mathbf{y}$ is the observation data vector. The signal vector is given by $\mathbf{x} = [x_1, \ldots, x_N]$, and it can be expressed as $\mathbf{x} = a\mathbf{s}$; so it is known, except for its level. The vector $\mathbf{w} = [w_1, \ldots, w_N]$ is the WGN with variance $\sigma_{\mathbf{w}}^2$, and the NP detector will decide the hypothesis H_1 if the logarithm likelihood ratio (log-likelihood) exceeds a certain threshold:

$$\ln(\Lambda(\mathbf{y})) = \ln\left(\frac{f_1(\mathbf{y})}{f_0(\mathbf{y})}\right) = \frac{\mathbf{y}^T\mathbf{s}}{\sigma_{\mathbf{w}}^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\lambda) = \lambda_0. \tag{2.7}$$

This is a uniformly most powerful (UMP) test, since the statistic compared with the threshold does not depend on the unknown parameter a [51]. The threshold λ_0 will be fixed by imposing a certain PFA:

$$PFA = 1 - erf(\lambda_0). \tag{2.8}$$

Attention should be paid to the limitations of this detector at the time of its implementation with real signals. On the one hand, the background noise should be white and Gaussian, and it is necessary to adaptively estimate the parameter $\sigma_{\mathbf{w}}^2$. On the other hand, in order to reach proper detector performance, the desired signal $\mathbf{s}$ must be known, but is often quite difficult to discern.

2.2.2 Random signal

This section will first focus on the optimal detector when the desired signal is modeled as a random signal, which follows a zero mean and Gaussian distribution with an arbitrary covariance matrix ($\mathbf{C_x}$). Then, we will particularize for the case in which the signal has a known covariance. In both cases the WGN is also considered with a known variance ($\sigma_{\mathbf{w}}^2$) and totally independent of the signal.

Estimator correlator

Taking the above considerations into account, the hypotheses H_0 and H_1 can be expressed as follows:

$$\begin{array}{ll} H_0 : \mathbf{y} = \mathbf{w} & \mathbf{w} : N(0, \sigma_{\mathbf{w}}^2 \mathbf{I}) \\ H_1 : \mathbf{y} = \mathbf{x} + \mathbf{w} & \mathbf{x} : N(a\mathbf{s}, \mathbf{C_x}). \end{array} \tag{2.9}$$

The covariance matrix of the signal is:

$$\mathbf{C_x} = E[(\mathbf{x} - a\mathbf{s})(\mathbf{x} - a\mathbf{s})^T] \tag{2.10}$$

where $\mathbf{s}$ is normalized ($\mathbf{s}^T\mathbf{s} = 1$) and the signal is also uncorrelated with the noise $E[\mathbf{w}\mathbf{x}^T] = 0$. After considering the PDF of $\mathbf{y}$ under the two hypotheses, it is possible to arrive at the expression of the likelihood ratio:

$$-\frac{1}{2}(\mathbf{y} - a\mathbf{s})^T\mathbf{A}^{-1}(\mathbf{y} - a\mathbf{s}) + \frac{1}{2\sigma_{\mathbf{w}}^2}\mathbf{y}^T\mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} \ln\lambda - \ln\frac{\sigma^N}{(det\mathbf{A})^{1/2}} = \lambda_1 \tag{2.11}$$

where $\mathbf{A}$ is defined as $\mathbf{A} = \mathbf{C_x} + \sigma^2\mathbf{I}$. In this case, it is not a UMP test, since the detector decision is not independent of the parameter a. It is possible to modify (2.11), obtaining:

$$\frac{1}{2\sigma_{\mathbf{w}}^2}\mathbf{y}^T\mathbf{y} - \frac{1}{2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} + a\mathbf{y}^T\mathbf{A}^{-1}\mathbf{s} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_1 + \frac{a^2}{2}\mathbf{s}^T\mathbf{A}^{-1}\mathbf{s} = \lambda_2. \tag{2.12}$$

After defining $\hat{\mathbf{x}} = \frac{1}{2\sigma_{\mathbf{w}}}\mathbf{y} - \frac{\sigma_{\mathbf{w}}}{2}\mathbf{A}^{-1}\mathbf{y} + a\sigma\mathbf{A}^{-1}\mathbf{s}$, this detector can thus be expressed in the form:

$$\frac{\mathbf{y}^T\hat{\mathbf{x}}}{\sigma_{\mathbf{w}}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_2. \tag{2.13}$$

This expression is similar to the one derived in (2.7) for the matched filter, but in this case, the detector correlates the received data $\mathbf{y}$ with an estimate of the signal $\hat{\mathbf{x}}$ and not with the known signal $\mathbf{s}$ as before. This detector is therefore termed the *estimator-correlator* [85].

Energy detector (ED)

One particular case of (2.11) is obtained when the signal is modeled as a zero mean, white Gaussian random process. This implies that $a = 0$ and, therefore, from (2.10), we can express the covariance matrix as $\mathbf{C_x} = E[\mathbf{x}\mathbf{x}^T]$. If the signal is also white, $\mathbf{C_x} = \sigma^2_{\mathbf{x}}\mathbf{I}$ is fulfilled, and the previous test becomes:

$$\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2_{\mathbf{w}}} \underset{H_0}{\overset{H_1}{\gtrless}} 2\frac{\sigma^2_{\mathbf{x}} + \sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{w}}}\lambda = \lambda_3 \tag{2.14}$$

which is known as the energy detector (ED), since it ultimately compares the energies of the measurements with a certain threshold λ_3.

Energy detectors are used in the automatic detection of signals in the presence of a background noise when the signal waveform is not precisely known. In a more general context, energy detection is of interest for detecting departures from a known background (event or novelty detection [56]). Energy detectors have a very simple structure and are easily implementable. The ED is optimum when the signal is independent of the noise and it is possible to model it as a random uncorrelated Gaussian process [43]. Otherwise, when the signal is completely unknown, the ED can only be considered a generalized likelihood ratio test (GLRT) [41]. The GLRT yields the likelihood ratio test statistic, obtained by replacing the unknown parameter under each hypothesis with their maximum likelihood estimate (MLE) [55]. It is intuitively deduced that when the signal $\mathbf{s}$ is present, the energy of the received data $\mathbf{y}$ will increase. In fact, the test of the expression (2.14) 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 H_0 ($\sigma^2_{\mathbf{w}}$), or under hypothesis H_1 ($\sigma^2_{\mathbf{w}} + \sigma^2_{\mathbf{x}}$). That said, it is possible to express the test (2.14) under each of the hypotheses as:

$$\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2_{\mathbf{w}}} \sim \chi^2_N \qquad \text{under} \quad H_0 \tag{2.15}$$

and

$$\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2_{\mathbf{x}} + \sigma^2_{\mathbf{w}}} \sim \chi^2_N \quad \text{under} \quad H_1. \tag{2.16}$$

The expression of the energy calculation follows a known distribution, χ^2_N chi-square with N degrees of freedom (see Appendix A.2). Therefore, it is possible to find the PFA and the PD, as well as fix the threshold of this

detector for a given PFA from the following expressions [43]:

$$PFA = Pr\left\{\frac{\mathbf{y}^T\mathbf{y}}{\sigma_{\mathbf{w}}^2} > \frac{\gamma'}{\sigma_{\mathbf{w}}^2}; H_0\right\} = Q_{\chi_N^2}\left(\frac{\gamma'}{\sigma_{\mathbf{w}}^2}\right) \tag{2.17}$$

and

$$PD = Pr\left\{\frac{\mathbf{y}^T\mathbf{y}}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{w}}^2} > \frac{\gamma'}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{w}}^2}; H_1\right\} = Q_{\chi_N^2}\left(\frac{\gamma'}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{w}}^2}\right) \tag{2.18}$$

where $Q_{\chi_\nu^2}$ is the right-tail probability for a χ_ν^2 random variable, given by:

$$Q_{\chi_\nu^2}(x) = \begin{cases} 2Q(\sqrt{x}) & \nu = 1 \\ 2Q(\sqrt{x}) + \dfrac{exp(-\frac{1}{2}x)}{\sqrt{\pi}} \displaystyle\sum_{k=1}^{\frac{\nu-1}{2}} \frac{(k-1)!(2x)^{k-\frac{1}{2}}}{(2k-1)!} & \nu > 1 \text{ and } \nu \text{ odd} \\ exp(-\dfrac{1}{2}x) \displaystyle\sum_{k=1}^{\frac{\nu-1}{2}} \frac{(\frac{x}{2})^k}{k!} & \nu \text{ even} \end{cases} \tag{2.19}$$

where $Q(x)$ is the probability that a Gaussian random variable with zero mean and unit variance exceeds x, as shown in Appendix A.5.

In summary, the ED is optimum for detecting zero-mean uncorrelated Gaussian signals, and is a GLRT detector in the case of an unknown deterministic signal. In both cases, the background noise must be zero-mean uncorrelated Gaussian.

2.2.3 Detection in correlated Gaussian noise

In the two previous cases, i.e., the matched filter and the energy detector, 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 two types of detectors mentioned above are studied when the noise is not white and the tests for each case are reformulated [34].

The hypotheses are similar to those presented in (2.6) and (2.9). This time, however, the noise is characterized by $\mathbf{w} : N(0, \mathbf{R_w})$, where the covariance matrix of the noise is not defined by the expression $\mathbf{R_w} = \sigma_{\mathbf{w}}^2\mathbf{I}$, as in the case of white noise. Now there is some correlation between the noise samples that must be taken into account, and therefore the new detection model can be expressed as follows:

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{w} \qquad \mathbf{w} : N(0, \mathbf{R_w}) \\ H_1 &: \mathbf{y} = \mathbf{s} + \mathbf{w} \quad \mathbf{s} : N(0, \sigma_{\mathbf{s}}^2\mathbf{I}). \end{aligned} \tag{2.20}$$

We first study the resulting test of the energy detector when the signal is in presence of correlated noise and the covariance matrix is consequently

not equivalent to the identity, as assumed in the previous models. 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 $\mathbf{y}_p$:

$$\mathbf{y}_p = \mathbf{R}_{\mathbf{w}}^{-1/2}\mathbf{y} \tag{2.21}$$

where $\mathbf{R}_{\mathbf{w}} = E[\mathbf{w}\mathbf{w}^T]$ represents the noise covariance matrix. The ED is then applied to the pre-processed observation vectors, obtaining the following test:

$$\frac{\mathbf{y}_p^T\mathbf{y}_p}{\sigma^2_{\mathbf{w}_p}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \qquad \Leftrightarrow \qquad \mathbf{y}^T\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{2.22}$$

Notice that $\mathbf{R}_{\mathbf{w}_p} = E[\mathbf{w}_p\mathbf{w}_p^T]$ and, hence, $\sigma^2_{\mathbf{w}_p} = 1$. 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. This test (2.22) will be henceforth referred to as pre-processed energy detector (PED). This modified test is optimum in case of assuming low signal-to-noise ratio (SNR), that is when $\sigma^2_{\mathbf{s}} \ll \sigma^2_{\mathbf{w}}$, the most difficult detection case.

For the matched filter case, a similar procedure can be employed. Using test (2.7), it is possible to obtain a modified version in which it also becomes necessary to apply the pre-whitening transformation by means of the inverse matrix $\mathbf{R}_{\mathbf{w}}^{-1}$:

$$\mathbf{y}^T\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{s} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_2. \tag{2.23}$$

In this case, it is also possible to separate the first term in $\mathbf{y}^T\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{s} = \mathbf{y}_p^T\mathbf{s}_p$, and thereby define $\mathbf{s}_p = \mathbf{R}_{\mathbf{w}}^{-1/2}\mathbf{s}$. The signals $\mathbf{y}_p^T$ and $\mathbf{s}_p$ are the observed samples pre-processed by means of a transformation matrix $\mathbf{R}_{\mathbf{w}}^{-1/2}$. This matrix can be estimated from a training set of K noise vectors $\mathbf{w}_k$, with $k = 1, \ldots, K$, using the sample estimate:

$$\hat{\mathbf{R}}_{\mathbf{w}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{w}_k\mathbf{w}_k^T \tag{2.24}$$

which is the noise covariance matrix known as the pre-whitening matrix. Note that in (2.22) and (2.23), it is not necessary to divide by $\sigma^2_{\mathbf{w}}$ since the pre-whitening also represents a normalization of the original observation noise vector.

2.2.4 Signal in subspace

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:

$$\begin{array}{ll} H_0 : \mathbf{y} = \mathbf{w} & \mathbf{w} : N(0, \sigma_{\mathbf{w}}^2 \mathbf{I}) \\ H_1 : \mathbf{y} = \mathbf{s} + \mathbf{w} & \mathbf{s} : \mathbf{H}\boldsymbol{\theta} \end{array} \tag{2.25}$$

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 $\mathbf{H}$ $(N \times p)$, termed the observation matrix, and modeled by a vector of unknown parameters $\boldsymbol{\theta}$, as follows $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$.

As the signal is deterministic signal 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 [43]:

$$T(\mathbf{y}) = \mathbf{y}^T \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{H} \hat{\boldsymbol{\theta}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{2.26}$$

According to the hypothesis described in (2.25), the covariance matrix of the noise can be expressed as $\mathbf{C}_{\mathbf{w}} = \sigma_{\mathbf{w}}^2 \mathbf{I}$, and the vector of unknown parameters can be estimated as $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_{\mathbf{w}}^{-1} \mathbf{y}$. This leads to the following test:

$$\frac{\mathbf{y}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}}{\sigma_{\mathbf{w}}^2} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{2.27}$$

where we can define $\mathbf{P} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ as the projection matrix onto the subspace signal $\mathbf{H}$. The statistic $\mathbf{y}^T \mathbf{P} \mathbf{y} / \sigma_{\mathbf{w}}^2$ follows a chi-square distribution with p degrees of freedom (χ_p^2) and, therefore, the threshold λ is easily calculated for a given PFA. For the particular case of $p = N$, the matrix $\mathbf{P}$ becomes the identity matrix $\mathbf{I}$ because we shall consider the entire observation space and (2.27) becomes the classical energy detector.

Current work on matched subspace and filter detection has been conducted in [75] and [76], where Gaussian noise and signal interference is assumed. Other variations of this problem include adaptive subspace detection and are mainly studied in [48].

2.3 Performance of the ED and the PED

2.3.1 Detector evaluation: ROC performance factor

Detector evaluation can be conducted by studying the performance of the receiver operating characteristic (ROC) curves, which represent the PD of the detector as a function of the PFA. It must be noted that the PD and the

PFA depend on the threshold λ and are calculated according to a specific threshold range. In addition, the PD not only depends on the threshold, but also on the SNR. Therefore, the ROC curves can be considered indicative of the quality of a detector, as will be seen below.

In general terms, the performance of a detector is improved when it presents a ROC curve above another detector curve, given the same length of the observation vector, SNR, and threshold (or PFA). This means that the detector has a higher PD for a specific PFA or a lower PFA for a specific PD. The limiting curve $PD = PFA$ corresponds to a random detector. As an example, Figure 2.1 shows the theoretical ROC curves of an ED for several signal-to-noise ratios. The interval of SNR shown extends from -8 dB to 4 dB. As seen, as the SNR increases, a higher PD is obtained for the same PFA, indicating a higher detector performance.

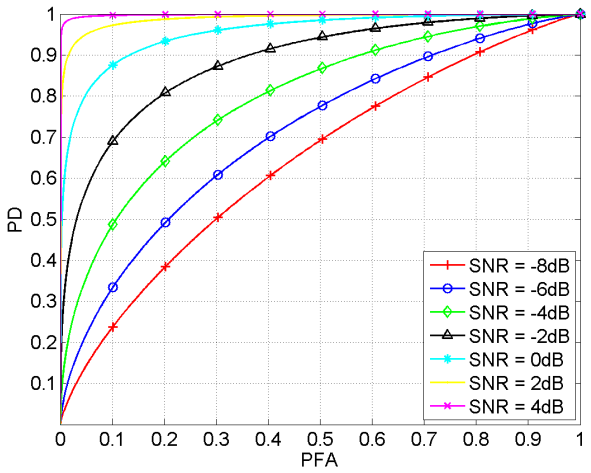


Figure 2.1: Theoretical ROC curves of the ED for different $SNRs$.

However, sometimes relying only on a cursory observation of the ROC curves can be misleading, and incorrect conclusions may be drawn due to the fact that the curves may behave differently, depending on the area of the curve under study. For this reason, a new parameter is needed to measure and compare the performance of different detectors through a consideration of the entire ROC curve, or only those parts in which the designated detector will work.

Let us define the two vectors of PFAs and PDs, which define the ROC curve of a detector:

$$\begin{aligned} \mathbf{a} &= [PFA_1, PFA_2, PFA_3, \ldots, PFA_L] \\ \mathbf{d} &= [PD_1, PD_2, PD_3, \ldots, PD_L] \end{aligned} \tag{2.28}$$

where L is the number of total points used to characterize the ROC curve. Therefore, to measure the performance of two detectors (D_1 and D_2), their ROC curves are compared by means of a new factor, denoted by γ. This new parameter is computed using a subset of PFAs and PDs extracted only from the interval of interest of the ROC curve. This interval is determined by the maximum PFA ($maxPFA$) considered, as seen in Figure 2.2.

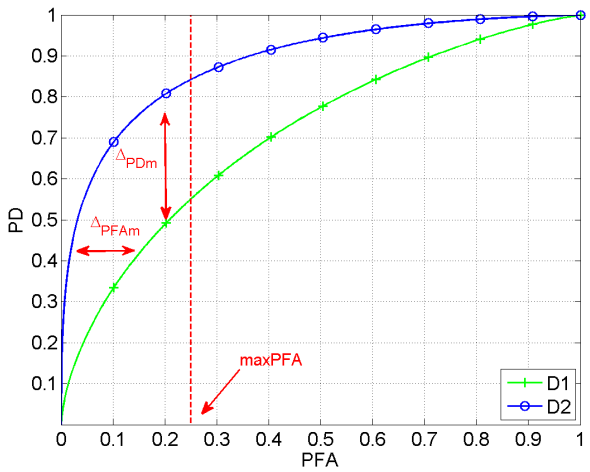


Figure 2.2: Calculation of the ROC factor (γ) for detectors D1 and D2 ($maxPFA = 0.25$).

Let us denote $\mathbf{a}_1$ as the vector of PFAs of D_1, and the corresponding vector of PDs as $\mathbf{d}_1$. Similarly, for D_2 vectors $\mathbf{a}_2$ and $\mathbf{d}_2$ are used. Thus, the ROC factor is obtained first by substracting the PDs values of both detectors for the same PFAs (between 0 and $maxPFA$). After this, the same procedure is repeated, but in this case substracting the PFAs of both detectors, given the same PDs. A vector of differences is obtained as a result, and then their mean can be computed as shown below:

$$\bar{\Delta}_{PD} = \frac{1}{M} \sum_{m=1}^{M} (\mathbf{d}_2(m) - \mathbf{d}_1(m)) \tag{2.29}$$

and

$$\bar{\Delta}_{PFA} = \frac{1}{M} \sum_{m=1}^{M} (\mathbf{a}_2(m) - \mathbf{a}_1(m)). \tag{2.30}$$

M is the number of PFAs (or PDs) used to compute the ROC factor, and its value will depend on the sections of the ROC under analysis. It must be noted that these factors represent an absolute mean value of improvement or deterioration in the PD and PFA of D_2 with respect to D_1 (depending on if

they are > 0 or < 0, respectively). Furthermore, considering that increases in $\bar{\Delta}_{PD}$ and decreases in $\bar{\Delta}_{PFA}$ indicate that D_2 is better than D_1, the γ factor is defined as shown below:

$$\gamma = \frac{\bar{\Delta}_{PD} - \bar{\Delta}_{PFA}}{2}. \tag{2.31}$$

Note that $\bar{\Delta}_{PD}$ and $\bar{\Delta}_{PFA} \in [0, 1]$, and therefore $\gamma \in [-1, 1]$. The performance of this factor can be summarized as follows:

- $\gamma > 0$: the ROC curve of D_2 is above the ROC curve of D_1, and therefore, this detector performs better. This improvement will be more significant as γ increases.
- $\gamma < 0$: the ROC curve of D_2 is below the ROC curve of D_1, and therefore, this detector performs poorly. This deterioration will be more significant as γ decreases.
- $\gamma = 0$: the ROC curve of D_2 is equal to the ROC curve of D_1, and therefore, both detectors have the same performance.

In what follows, we will use this parameter to evaluate and compare the performance of two detectors in similar conditions. This is a more concise parameter, and an alternative to simply observing the ROC curves.

2.3.2 Comparison of the ED and the PED in Gaussian noise

In the previous sections, the ED was presented as a practical solution to the detection of unknown signals in presence of Gaussian and independent noise. Furthermore, the particular case of dependence between the noise samples was analyzed and it was studied how the PED can address these situations when the noise is Gaussian.

Let us now evaluate and compare the performance of both detectors in the presence of Gaussian noise and observe how they behave in two different situations, independent and non-independent samples. This fact can considerably affect the outcomes obtained by the detectors, often leading to unacceptable PFA or poor PD.

Figure 2.3 reveals the experimental ROC curves of the ED and the PED, obtained from (2.14) and (2.22), respectively, when detecting a random Gaussian signal in the presence of non-independent Gaussian noise samples. Both results are compared for the same SNR, with the theoretical ROC curve of the ED (ED_{theo}) obtained from (2.17) and (2.18), where independent Gaussian noise is assumed. As seen in the graph, the performance of the ED significantly decreases in comparison to the ED_{theo} curve, since the ED is not prepared to work correctly with non-independent noise samples; consequently, a negative ROC factor of $\gamma = -0.1459$ is obtained (dashed arrow).

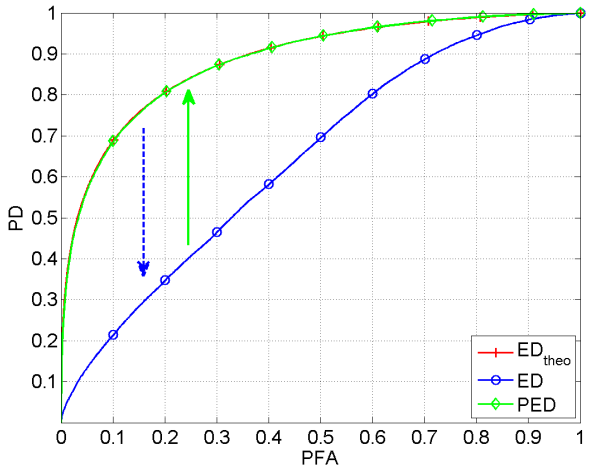


Figure 2.3: Simulated ROC curves of the ED and the PED (correlated Gaussian noise) in comparison to the ED_{theo} (uncorrelated Gaussian noise) for $SNR = -2$ dB.

However, when using the PED in the same situation, a pre-whitening transformation is applied prior to calculating the energy, it can be observed how a significant improvement in detection performance is achieved (solid arrow). In this case, the ROC curve is overlapped with the theoretical curve, and computing the ROC factor of the PED curve in comparison to the ED_{theo}, a value near 0 ($\gamma = 9.1e^{-4}$) is obatined. Therefore, we can conclude that the PED compensates for the deterioration experimented by the ED with non-independent Gaussian noise.

2.4 Conclusions

In this chapter, 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. The Bayes and Neyman Pearson criterion were studied for the detector design and 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 has been 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. However, the particular case of the ED having correlated noise samples was examined leading to the PED. A new ROC factor was presented to evaluate the detector per-

formance in terms of its ROC curve. Finally, the behavior of the ED and the PED was compared in the case of correlated Gaussian noise, showing the significant improvement achieved with the PED in comparison to the ED for the same noise conditions and SNR.

Part II

Theoretical framework

Chapter 3

Extensions of the energy detector in non-Gaussian and non-independent noise

It is the unsolved problems that keep the mind alert, not the ones that have been solved already.

Erwin Guido

In this chapter, the Rao test is presented as a possible solution to address the problem of detecting signals in the presence of non-Gaussian noise. Some limitations of this test are studied when it is used in the particular case of energy detection. Therefore, some alternative extensions based on the use of two possible non-linear functions are presented. The principal objective is to reach a noise distribution in which samples are as Gaussian distributed as possible. Furthermore, when the background noise is not only non-Gaussian distributed, but also non-independent, the detection problem becomes more complicated. Thus, the use of a linear transformation based on independent component analysis (ICA) is also presented, leading to further extensions of the energy detector (ED) and the matched subspace filter (MSF).

3.1 Detection in presence of non-Gaussian noise

This section offers a brief analysis of different types of noises that do not follow a Gaussian distribution, as well as two approaches to characterizing them. The detection of deterministic signals in non-Gaussian noise is

reviewed. Then the problem is generalized to unknown signal detection, leading to a study of the Rao test.

3.1.1 Study of Gaussianity

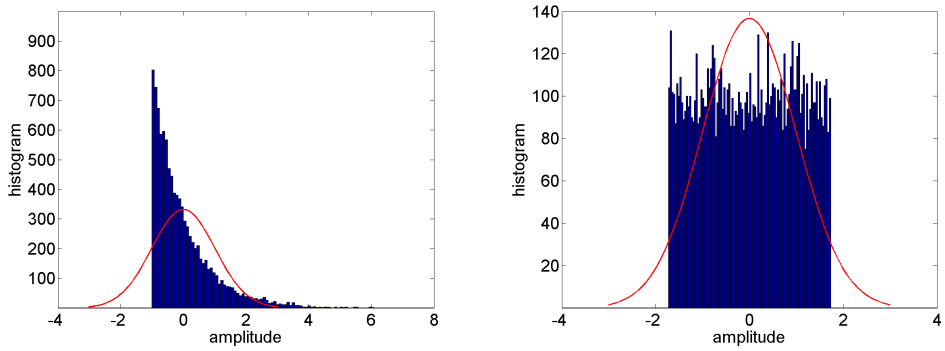
To properly evaluate the behavior of a detector, it is very important to recognize or to characterize the background noise in which the signal to be detected is immersed. It is therefore necessary to have some parameter or method to perform even a generic classification of the noise.

A random variable which follows a Gaussian distribution is defined with the terms described in Appendix A. However, the bibliography lists a widely studied range of distributions [80], such as Laplacian, Gamma, Rayleigh and Uniform. The utilization of the ED to detect unknown signals in the presence of these noise distributions leads to an increase of unacceptable false alarms. Therefore, it is of special interest to study all these distributions.

There are several methods to test the non-Gaussianity in a particular distribution of given data, but in our case two main techniques are used:

- **Histogram**: This method is simple and widely employed because it allows one to approximate the PDF of the data. This method also allows one to identify, graphically, the shape of the distribution. If it is plotted in combination with the theoretical Gaussian PDF, it is possible to differentiate between Gaussian and non-Gaussian by evaluating, subjectively, the similarity between the two distributions. Figure 3.1 shows the histograms of two different non-Gaussian distributions, Gamma and Uniform, each with zero mean and unit variance. Both are shown alongside a Gaussian distribution of the same standard deviation and mean for purposes of comparison. As can be observed, the PDFs differ from the Gaussianity assumption.

- **Kurtosis**: The degree of non-Gaussianity can be more specifically calculated by estimating the *kurtosis*. Defined as the fourth cumulant divided by the square of the second cumulant, *kurtosis* measures the degree of peakedness or flatness of a distribution, in comparison with a Gaussian PDF [4, 37]. For the Gaussian case, the *kurtosis* value is 3; and, depending on whether the *kurtosis* of the data is > 3 or < 3, platikurtic or leptokurtic distributions can be respectively differentiated. Thus, a *kurtosis* higher than 3 defines a more peaked distribution while, on the other hand, a *kurtosis* lower than 3 defines a flatter distribution than a Gaussian distribution. Randomly generated examples of non-Gaussian data distributions are listed here, along with their respective estimated *kurtosis* value in parentheses: Laplacian (6), Rayleigh (3.26), Gamma (8.45), and Uniform (1.8).



(a) (b)

Figure 3.1: PDF of two different non-Guassian distributions in comparison to a Gaussian PDF (solid line) of the same standard deviation and mean. (a) Gamma data distribution. (b) Uniform data distribution.

3.1.2 Detection of random signals: Rao test

As seen in the previous chapter (Section 2.2), 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 for both Bayes and Neyman-Pearson criteria when the signal and noise follow a Gaussian distribution with zero-mean, and the samples are uncorrelated.

However, 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). For example, taking into account the approach used in Neyman-Pearson, the search for the optimal detector is based on maximizing the PD for a given PFA. In this case, the ED will not reach the maximum PD if the assumption of the Gaussian noise model is not achieved. Moreover, considering the most general case in which the characteristics of the signal are completely unknown, the statistic $\mathbf{y}^T\mathbf{y}/\sigma^2_{\mathbf{w}}$ could have an unpredictable distribution, and therefore the calculation of the appropriate threshold would be a challenge.

Unfortunately, there is no general optimum solution for the ED when the noise is non-Gaussian. However, for the subspace signal case $\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where the matrix $\mathbf{H}$ with dimension $N \times p$ represents the subspace signal, $\boldsymbol{\theta}$ denotes the parameters of the unknown signal, and $\mathbf{w}$ is the noise vector whose components are random and independent and identically distributed (i.i.d.), it is possible to attain the following, a particular implementation of

the general Rao test [43]:

$$\frac{g(\mathbf{y})^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T g(\mathbf{y})}{P_{g(w)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{3.1}$$

Prior to calculating the energy of the vector $\mathbf{y}$, function $g(\cdot)$ is applied element-wise to the vector $\mathbf{y}$ in the following manner:

$$g(\mathbf{y}) = [g(y_0)g(y_1)\dots g(y_{N-1})]. \tag{3.2}$$

If $p(w)$ represents the noise PDF, it can be defined as:

$$g(w) = -\frac{\frac{dp(w)}{dw}}{p(w)} \tag{3.3}$$

and

$$P_{g(w)} = \int_{-\infty}^{\infty} \frac{\left[\frac{dp(w)}{dw}\right]^2}{p(w)} dw. \tag{3.4}$$

Note that the preprocessed noise mean-power $P_{g(w)}$ can be expressed as follows:

$$P_{g(w)} = \int_{-\infty}^{\infty} g^2(w)p(w)dw = E[g^2(w)] \tag{3.5}$$

where $E[\cdot]$ calculates the statistical average.

Therefore, particularizing the Rao test for the energy detector and assuming that $p = N$, $\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T = \mathbf{I}$ is obtained. From this, it is possible to arrive at the following expression:

$$\frac{g(\mathbf{y})^T g(\mathbf{y})}{P_{g(w)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{3.6}$$

Hence, (3.6) is a generalization of (2.14) in the sense that it is necessary to apply some pre-processing function $g(\cdot)$ to our observation data vector prior to the calculation of the normalized energy. Note that for the Gaussian case, $g(w) = w$, so $g(\mathbf{y}) = \mathbf{y}$. The parameter $P_{g(w)} = P_w$ will take the value $\sigma^2_{\mathbf{w}}$, thus (2.14) and (3.6) will be equivalent.

Finally, it should be noted that the statistic $g(\mathbf{y})^T g(\mathbf{y})/P_{g(w)}$, as in the Gaussian case, also follows a chi-square distribution (χ^2_N), and therefore, it is possible to estimate the threshold value λ for the PFA required by the system.

Rao test example and limitations

An example of typical non-Gaussian noise is one that follows a Laplacian distribution. The PDF with a zero-mean is given in this case by the expression:

$$p(w) = \frac{1}{\sqrt{2}\sigma} exp\left(-\frac{\sqrt{2}}{\sigma}|w|\right) \qquad -\infty < w < \infty \tag{3.7}$$

where σ is the noise standard deviation. Comparing this distribution with a Gaussian having the same variance and zero-mean (see Appendix A.1), it can be observed from Figure 3.2 that the principal difference is the appearance of noise samples of greater magnitude in the Laplacian distribution versus the Gaussian example. This leads to the appearance of spikes and outliers in the time domain, clearly seen in Figure 3.3, where a realization of each type of noise is represented.

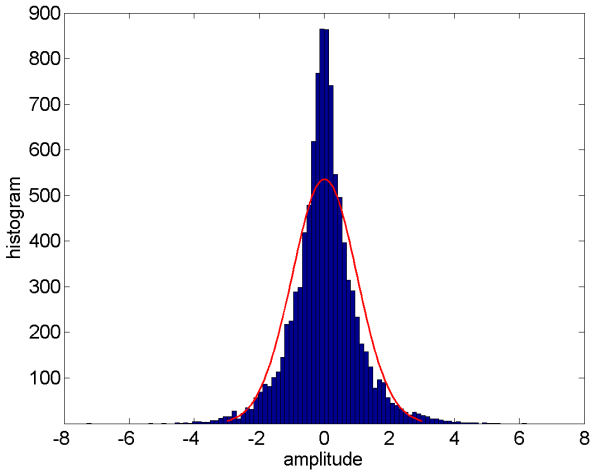
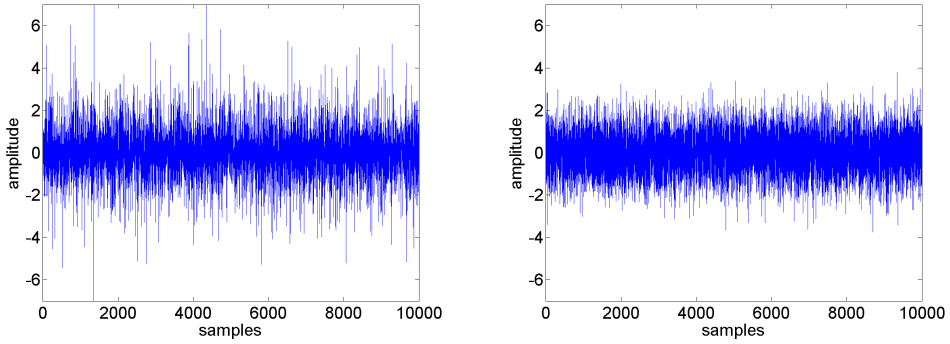


Figure 3.2: PDF of Laplacian distribution in comparison to a Gaussian PDF (solid line) of the same standard deviation.

In the presence of non-Gaussian noise, one would expect the detector to be able to take into account these outliers from the background noise. Thus, this fact would not affect nor excessively damage the detector PFA. As seen earlier, it is possible to use detectors that are able to manage such situations. As a solution to this problem, the particularization of the Rao test was presented in (3.1) and (3.6) for the case of subspace deterministic signals with unknown parameters and for completely unknown signals, respectively. In both cases, the use of a non-linear function $g(\cdot)$ was employed with the aim of reducing the spikes and outliers of the non-Gaussian noise. However, it will be demonstrated how it is not always possible to use the expression



(a) (b)

Figure 3.3: Time realization of two different noise distributions. (a) Laplacian noise distribution. (b) Gaussian noise distribution.

$g(\cdot)$ given in (3.3), particularly in the generalization of the Rao test for the energy detector given in (3.6).

In (3.6), the function $g(\cdot)$ must be known in order to apply the appropriate pre-processing to the data when the noise is non-Gaussian. As seen in (3.3), the function $g(\cdot)$ depends on the probability density function of the noise, $p(w)$. For example, to detect the presence of an unknown signal in a Laplacian background noise with independent samples, a non-linear function, which takes into account both (3.3) and (3.7), is defined as:

$$\begin{aligned} g(w) &= -\frac{\frac{dp(w)}{dw}}{p(w)} = -\frac{d\ln p(w)}{dw} \\ &= \sqrt{\frac{2}{\sigma^2}}\frac{d|w|}{dw} = \sqrt{\frac{2}{\sigma^2}}sgn(w) \end{aligned} \tag{3.8}$$

where

$$sgn(w) = \begin{cases} 1 & \text{if } w \geq 0 \\ -1 & \text{if } w \leq 0 \end{cases} \tag{3.9}$$

and the pre-processed noise mean-power is given by $P_{g(w)} = 2/\sigma^2$ [42].

Therefore, the Rao test can be expressed as:

$$\begin{aligned} T(\mathbf{y}) &= \frac{g(\mathbf{y})^T g(\mathbf{y})}{P_{g(w)}} \\ &= \frac{\left(\sqrt{\frac{2}{\sigma^2}} sgn(\mathbf{y})\right)^T \left(\sqrt{\frac{2}{\sigma^2}} sgn(\mathbf{y})\right)}{2/\sigma^2} \\ &= sgn(\mathbf{y})^T sgn(\mathbf{y}) = N \end{aligned} \tag{3.10}$$

where N is the dimension of the observation vector $\mathbf{y}$. The energy is always constant, equal to N, and independent of the observation vector. Hence, it is necessary to find alternative non-linear functions for the Rao test when it is particularized for the energy detector case. This is discussed in the following section, where two different transformations are presented.

3.2 Extended energy detector (EED)

3.2.1 Alternative estimates of the non-linear function

Let us first consider the simplest case in which the components of the noise vector $\mathbf{w} = [w_0, w_1, \ldots, w_{N-1}]^T$ are 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 $g(\cdot)$ is proposed to convert a random variable w 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 $g(\cdot)$, to every component of the observation vector $\mathbf{y}$ before computing the energy, and then by subsequently implementing the following test [64]:

$$\frac{g(\mathbf{y})^T g(\mathbf{y})}{P_{g(w)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{3.11}$$

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 detection threshold fixed by the PFA selected. The non-linear transformation $g(\cdot)$ in (3.11) can be implemented using different techniques. Of these, the well-known parametric Box-Cox transformation will be examined first; followed by the presentation of a new non-linear transformation, itself 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.

3.2.2 The parametric power transformation

One of the most common transformations used to achieve Gaussianity was proposed by Box and Cox in [10], where a parametric power transformation was presented to reduce departures from normality. These authors demonstrated that a distribution can be approximated to Gaussian only by applying a transformation based on raising to a power of the non-Gaussian variable w as follows:

$$g_\beta(w) = \begin{cases} \dfrac{w^\beta - 1}{\beta} & \text{if } \beta \neq 0 \\ \log w & \text{if } \beta = 0 \end{cases} \tag{3.12}$$

where the Gaussianization process can be controlled by the transformation parameter β. The most recommended studies and theoretical analyses use the expression given by (3.12), instead of merely w^β, since this is continuous at $\beta = 0$. For vectors $\mathbf{w}$, the transformation is applied to each element separately with the same β, so independent components of the vectors are assumed.

The transformation in equation (3.12) is only valid for $w > 0$, and therefore another version of this transformation must be considered. Accordingly, Box and Cox proposed a shifted power version where w can take negative values:

$$g_\beta(w) = \begin{cases} \dfrac{(w+\beta_2)^{\beta_1} - 1}{\beta_1} & \text{if } \beta_1 \neq 0 \\ \log(w+\beta_2) & \text{if } \beta_1 = 0 \end{cases} . \tag{3.13}$$

In this case, $\beta = (\beta_1, \beta_2)$; in practice, the β_2 must be selected so that $w+\beta_2 > 0$ for any value of w. Therefore, β_1 can be considered the transformation parameter as β_2 is only necessary to ensure positive values.

This technique has been extensively studied, although there is a family of possible variants that partially modify this transformation. Some of them are reviewed in [74]. For example, in [54] an exponential transformation is proposed; in [38] the so-called modulus transformation was introduced; and in [6] another modification was suggested in order to support the normal distribution.

Nevertheless, our efforts here are focused on the original transformation proposed in (3.13) and on the estimation of the parameter β_1, denoted in the following by β. In [74], a review of the most common techniques used in the estimation of the transformation parameter β is presented. One is the maximum likelihood method, commonly used since it is conceptually easy to implement and the profile likelihood function is easy to compute in this case. Another approach is based on a Bayesian method, as presented in [71]. Some others are focused on the estimation of the transformation parameter on the basis of enforcing a particular assumption [84]. However, we will not attempt to estimate this parameter using these aforementioned methods, but rather,

we shall experimentally study the effect of β; that is, observe the resulting ROC curve of the EED when it implements this transformation, as will be shown in Section 3.5.

It is important to note that, clearly, not all non-Gaussian data may be power-transformed to normal. This problem was studied in [25], where it was concluded that even in cases where no power transformation could bring the distribution to exact normal, the usual estimates of β will lead to a distribution which tends toward this requirement. Therefore, and in general, we can use the central assumption that, for some β, the transformed observations can be treated as more normally distributed than the original data.

3.2.3 Transformation based on the data PDF

The Box-Cox transformation shown in the previous section can be easily implemented, but it presents the disadvantage of not taking into account the distribution of the original data when the non-linear transformation is applied. As a result, it is possible that some distributions are closer to Gaussianity, and as such, need a softer transformation than others that are not so Gaussian. It is therefore necessary to find another transformation that best fits the input data.

Considering the solutions proposed in [70], it is not possible to *directly* transform a non-Gaussian random variable into a Gaussian one. However, a random variable with Uniform distribution can be transformed into another with a specific distribution (Gaussian in this case). The problem can thus be addressed by introducing an intermediate step in which the data distribution becomes Uniform and is then transformed into the desired Gaussian distribution. Hence, the transformation process described consists of transforming a random variable w with a known PDF $F_w(w)$ into another called v with a Gaussian distribution function defined by $F_v(v)$. For this purpose, the function $g(w)$ must be found so that the distribution of the random variable $v = g(w)$ follows the specified data distribution. Therefore, the process will be divided into two steps, as illustrated in Figure 3.4:

- **Known PDF to Uniform distribution ($F_w(w) \rightarrow F_z(z)$)**: given a random variable w with a known distribution $F_w(w)$, we find the function $g_1(w)$ to transform it into another called z, which is uniformly distributed in the interval (0,1). In this case, the transformation is $g_1(w) = F_w(w)$ and, therefore:

$$z = F_w(w) \quad \text{where} \quad F_z(z) = z \quad \text{for} \quad 0 \leq z \leq 1. \tag{3.14}$$

  The random variable z can be considered as the output of a non-linear system without memory, whose input is w, and whose transfer function is defined by $F_w(w)$.

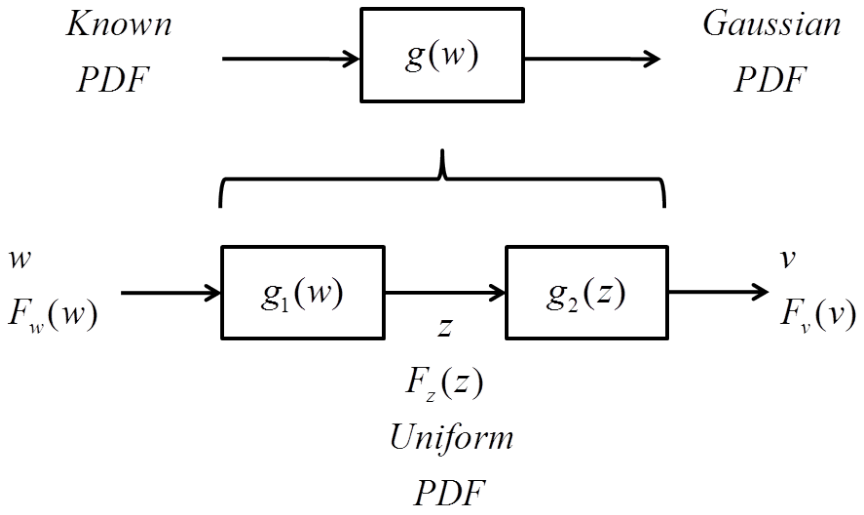


Figure 3.4: Transformation of a known PDF into a Gaussian distribution.

- **Uniform to Gaussian distribution** ($F_z(z) \rightarrow F_v(v)$): in this case, given the random variable z uniformly distributed in the interval (0,1), we wish to find the function $g_2(z)$ that yields a new random variable v with the distribution defined by the function $F_v(v)$. In this case, the required transformation is given by $g_2(z) = F_v^{-1}(z)$ and, hence:

$$v = F_v^{-1}(z). \tag{3.15}$$

Finally, by applying the above two steps, a random variable v, whose data follow a Gaussian distribution $F_v(v)$, with zero-mean and variance σ_g^2 is obtained. The resulting transformation can be expressed as follows:

$$v = g(w) = F_v^{-1}(F_w(w)) \tag{3.16}$$

where

$$F_v(z) = \frac{1}{\sigma_g\sqrt{2\pi}} \int_{-\infty}^{z} e^{-(1/2\sigma_g^2)t^2} dt. \tag{3.17}$$

3.2.3.1 Known noise PDF

Taking into account (3.16), and considering the case in which the random variable to be transformed corresponds to the noise w, it becomes necessary to know the distribution function $F_w(w)$ to obtain $g(w)$. If this information were known in advance, the transformation could be immediately calculated.

For example, a Laplacian noise distribution can be transformed into a Gaussian distribution following the steps above. So, from (3.16), it is necessary to calculate $F_w(w)$ (of the Laplacian noise), and F_v^{-1} (of the Gaussian

noise). As the Laplacian data distribution (with zero-mean and standard deviation σ_l) is defined by (3.7), the expression $F_w(w)$ is yielded:

$$F_w(w) = \frac{1}{2}\left[1 + sgn(w)\left(1 - e^{\frac{-\sqrt{2}|w|}{\sigma_l}}\right)\right] \tag{3.18}$$

where sgn is a function defined in (3.9). Then, using (3.17), the inverse of the Gaussian distribution function is obtained as follows:

$$F_v^{-1}(v) = \sigma_g\sqrt{2}\, erf^{-1}(2v-1) \tag{3.19}$$

where σ_g represents the standard deviation of the Gaussian noise. The error function can be expressed as:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt. \tag{3.20}$$

Hence, assuming (3.18) and (3.19), the non-linear transformation that should be applied to a Laplacian noise variable w to convert it into Gaussian noise can be defined by the follow expression:

$$g(w) = \sigma_l\sqrt{2}\, erf^{-1}\left(sgn(w)\left(1 - e^{\frac{-\sqrt{2}|w|}{\sigma_l}}\right)\right). \tag{3.21}$$

3.2.3.2 Unknown noise PDF

In most applications, the information about the noise distribution is unavailable. Therefore, to maintain the generalization of the transformation and to ensure its applicability in several scenarios of background noises, a non-parametric estimation of the noise PDF ($F_w(w)$) must be attempted. For this purpose, the non-parametric estimator described in [78] is employed to estimate the PDF of the random variable w using a set of samples $w_l, l = 1, \ldots, L$ corresponding to various realizations of w. As a result, the following expression is attained:

$$\hat{p}(w) = a\sum_{l=1}^{L} exp\left(-\frac{1}{2}\left(\frac{w - w_l}{h}\right)^2\right) \tag{3.22}$$

where $a = \dfrac{1}{Lh\sqrt{2\pi}}$ is a normalization constant and h is a parameter which controls the degree of smoothing of the estimated PDF.

With (3.22), the estimation of $F_w(w)$ can therefore be derived:

$$\begin{aligned}\hat{F}_w(w) &= \int_{-\infty}^{w} a\sum_{l=1}^{L} exp\left(-\frac{1}{2}\left(\frac{x - w_l}{h}\right)^2\right)dx \\ &= \frac{1}{L}\sum_{l=1}^{L}\frac{1}{2}\left[1 + erf\left(\frac{w - w_l}{h\sqrt{2}}\right)\right].\end{aligned} \tag{3.23}$$

In (3.23), samples w are the components of the noise vectors to be transformed into Gaussian, and samples $w_l, l = 1, \ldots, L$ correspond to the components of the original training noise, where L is the total number of training samples.

Hence, taking into account (3.16); (3.19) with $\sigma_g = 1$; and (3.23); it becomes possible to express the non-linear and non-parametric transformation as follows:

$$\begin{aligned} \hat{g}(w) &= \Phi^{-1}(\hat{F}_w(w)) \\ &= \sqrt{2}\, erf^{-1}\left(\frac{2}{L}\sum_{l=1}^{L}\frac{1}{2}\left[1 + erf\left(\frac{w - w_l}{h\sqrt{2}}\right)\right] - 1\right) \end{aligned} \tag{3.24}$$

where $\Phi(x)$ is defined in Appendix (A.4).

It is possible to illustrate the behavior of this function $g(\cdot)$ for different non-Gaussian data distributions, as displayed in Figure 3.5 for the following distributions: Laplacian, Rayleigh, Gamma and Uniform. It becomes apparent that one of the effects of the function $g(\cdot)$ consists of cutting the peaks and outliers of the signal, as is expected.

3.2.4 Optimality study of the EED

An extension of the ED has been presented to deal with the most general case of non-Gaussianity of the background noise. This detector applies a scalar non-linear function with the aim of converting the non-Gaussian noise in Gaussian. As a rule, the extended detector will not be optimum, but it will be shown that, under the weak signal assumption, it is GLRT.

The EED test is obtained by applying one of the two non-linear transformations studied in the previous section to every component of the observation vector $\mathbf{y}$. However, the present focus is on the EED when implementing the non-parametric transformation, since this type of processing is more appropriate when different types of noise are present. Therefore, for the study of optimality, the following definition of the new transformed variable u is used:

$$u = g(w) = \Phi^{-1}(F_w(w)). \tag{3.25}$$

Assuming that the samples are i.i.d., the energy is thereby calculated:

$$g(\mathbf{y})^T g(\mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{3.26}$$

Notice that as the transformed random variable u in (3.25) is zero mean and unit-variance, the noise mean-power $P_w = E[g^2(w)] = 1$.

Regarding the control of the PFA, the proposed EED statistic in (3.26) follows a chi-squared distribution (χ^2_N) under H_0, as in the ED. This allows

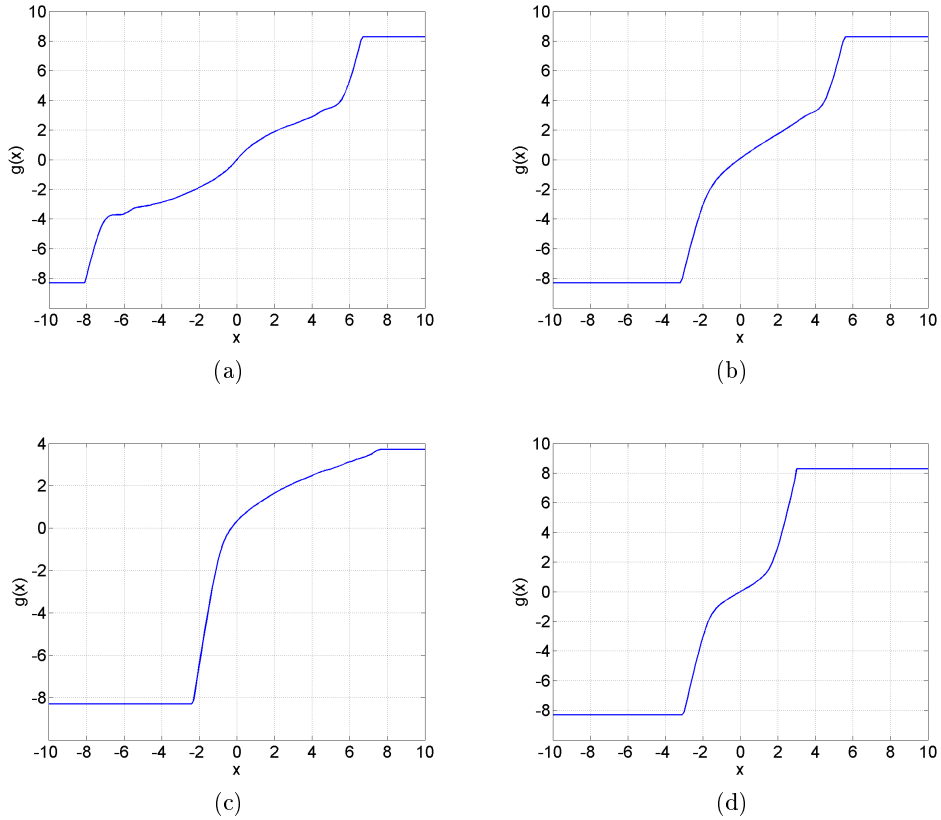


Figure 3.5: Non-linear function $g(\cdot)$ for different unit-variance and zero-mean non-Gaussian noise distributions. (a) Laplacian noise distribution. (b) Rayleigh noise distribution. (c) Gamma noise distribution. (d) Uniform noise distribution.

one to calculate the required threshold λ for a specific PFA. But, what can we say about the optimality of the new detector? Note that the actual performance of the EED depends on how the non-linear transformation affects the signal under H_1. Therefore, it is difficult to establish the optimality of the EED in a general form. However, let us assume the more interesting case of a low signal-to-noise ratio (SNR), since every reasonable detector must work well with a high SNR. Thus, the next detection model will be considered:

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{w} \qquad \mathbf{w} : p_{\mathbf{w}}(\mathbf{w}) = p_w(w_0) \cdot p_w(w_1) \dots p_w(w_{N-1}) \\ H_1 &: \mathbf{y} = \mathbf{s} + \mathbf{w} \end{aligned} \tag{3.27}$$

where $p_{\mathbf{w}}(\mathbf{w})$ is an arbitrary noise probability density function. By applying the non-linear transformation to these hypotheses, it is possible to obtain a

new model:

$$\begin{aligned} &H_0 : g(\mathbf{y}) = g(\mathbf{w}) && g(\mathbf{w}) : N(0, \mathbf{I}) \\ &H_1 : g(\mathbf{y}) = g(\mathbf{s}+\mathbf{w}) = g(\mathbf{w}) + diag[g'(\mathbf{w})]\mathbf{s} = g(\mathbf{w}) + \mathbf{z} \end{aligned} \tag{3.28}$$

where the assumption of low SNR is considered by means of a linear approximation of $g(\mathbf{s}+\mathbf{w})$ and following a Taylor series expansion. The matrix $diag[g'(\mathbf{w})]$ is defined as a diagonal matrix whose main diagonal is formed by vector $g'(\mathbf{w})$. Notice that $\mathbf{z}$ cannot be considered, generally, as a Gaussian distribution with uncorrelated variables, even in the case that $\mathbf{s}$ could follow a Gaussian distribution with uncorrelated samples. Hence, all these considerations lead us to conclude that EED will not be optimum in most cases. However, considering $\mathbf{z}$ as a completely unknown vector, the achieved Gaussianity of the noise after transformation guarantees that ED is GLRT. This is an interesting property, but it remains necessary to consider how the SNR is modified after the non-linear transformation.

To perform this analysis, an enhancement factor, denoted α, is defined. It is calculated as the quotient of the SNR after the non-linear transformation (SNR_g), and before it (SNR_o). Using the results and definitions of (3.27) and (3.28), α may be expressed as follows:

$$\begin{aligned} \alpha &= \frac{SNR_g}{SNR_0} = \frac{E[\mathbf{z}^T\mathbf{z}]/E[g(\mathbf{w})^T g(\mathbf{w})]}{E[\mathbf{s}^T\mathbf{s}]/E[\mathbf{w}^T\mathbf{w}]} = \frac{E[\mathbf{s}^T diag(g'^2(\mathbf{w}))\mathbf{s}]}{E[\mathbf{s}^T\mathbf{s}]/E[w^2]} \\ &= \frac{E[g'^2(w)]\cdot E[\mathbf{s}^T\mathbf{s}]}{E[\mathbf{s}^T\mathbf{s}]/E[w^2]} = E[g'^2(w)]\cdot E[w^2]. \end{aligned} \tag{3.29}$$

The following expressions were considered concurrently:

$$\begin{aligned} E[\mathbf{w}^T\mathbf{w}] &= \sum_{n=1}^{N} E[w_n^2] = N\cdot E[w^2] \\ E[g(\mathbf{w})^T g(\mathbf{w})] &= \sum_{n=1}^{N} E[g^2(w_n)] = N\cdot E[g^2(w)] = N \\ E[\mathbf{s}^T diag(g'^2(\mathbf{w}))\mathbf{s}] &= E\left[\sum_{n=1}^{N-1} g'^2(w_n)s_n^2\right] \\ &= \sum_{n=1}^{N-1} E[g'^2(w_n)]\cdot E[s_n^2] = E[g'^2(w)]\cdot E[\mathbf{s}^T\mathbf{s}]. \end{aligned} \tag{3.30}$$

Possible changes in the SNR (resulting from the effects of the non-linear transformation) are indicated by α. To calculate this factor, it is necessary

to obtain $g'(w)$ by using the expression from (3.25):

$$\begin{aligned} g'(w) &= \frac{du}{dF_w(w)}\frac{dF_w(w)}{dw} = \frac{1}{dF_w(w)/du}\, p_w(w) = \frac{1}{\Phi'(u)}\, p_w(w) \\ &= \frac{1}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}}\, p_w(w) = \sqrt{2\pi}\, e^{\frac{1}{2}g^2(w)}\, p_w(w). \end{aligned} \tag{3.31}$$

And finally, using (3.29) and (3.31), the factor α can be defined as:

$$\begin{aligned} \alpha &= E[g'^2(w)] \cdot E[w^2] \\ &= 2\pi \int_{-\infty}^{\infty} e^{g^2(w)} p_w^3(w) dw \cdot \int_{-\infty}^{\infty} w^2 p_w(w) dw. \end{aligned} \tag{3.32}$$

As detailed in (3.32), the factor α can be computed to evaluate the change in SNR due to the non-linear transformation for a specific noise distribution. Note that, as expected, for zero-mean Gaussian noise distribution, and utilizing (3.25), $g(w) = w/\sigma_w$ is a linear transformation; hence, $g'(w) = 1/\sigma_w$ (obtained from (3.31)), and thus $\alpha = 1$. It is important to emphasize that for $\alpha > 1$, an EED operating in non-Gaussian background noise performs better than an ED operating in Gaussian background noise, given the same SNR. This fact is not guaranteed when $\alpha < 1$, but improvements in the PD of the EED with respect to the ED, and with both operating on the same non-Gaussian noise, are still possible. This performance can only be verified experimentally for each specific noise distribution, as we will see in Section 3.5.

3.3 Detection with statistical dependence between noise samples

3.3.1 Introduction

Both the ED (studied in Section 2.2) and the EED (3.26) 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 seen in Section 2.2.3, independence and uncorrelation are equivalent for the Gaussian case; hence simple pre-whitening is sufficient and the pre-processed energy detector (PED) can be utilized.

However, the statistical dependence problem for the non-Gaussian case is not so easily solved. One of the techniques used to reach vectors with independent components is to apply an independent component analysis (ICA), explained more fully in Section 3.3.2. 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 $\mathbf{y}$:

$$\mathbf{y}_p = \mathbf{U}\mathbf{y}. \tag{3.33}$$

Equation (3.33) is a generalization of (2.21), 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 $\mathbf{y}$, as performed in equation (2.21), and the second yields the desired independence by means of a unitary transformation (equivalent to a rotation).

Although there are several options to implement the decorrelation process, the pre-whitening step shown in (2.21) is used here to enhance the generalization from Gaussian to non-Gaussian. Thus, matrix $\mathbf{U}$ can be decomposed as:

$$\mathbf{U} = \mathbf{Q}\mathbf{R}_{\mathbf{w}}^{-\frac{1}{2}} \quad \text{where} \quad \mathbf{Q}^T\mathbf{Q} = \mathbf{I}. \tag{3.34}$$

As a result, the ED for non-Gaussian, non-independent noise is defined as follows:

$$g(\mathbf{Q}\mathbf{R}_{\mathbf{w}}^{-\frac{1}{2}}\mathbf{y})^T g(\mathbf{Q}\mathbf{R}_{\mathbf{w}}^{-\frac{1}{2}}\mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{3.35}$$

It will be referred to hereafter as the pre-processed extended energy detector (PEED). Notice that, as in (3.26), normalization by the noise mean-power is not required, as the non-linear transformation generates random zero-mean unit-variance Gaussian variables. Section 3.3.3 will address the problem of estimating the transformation $\mathbf{U}$ in (3.35) from a training set of noise data samples.

All the considerations given in Section 3.2.4 are now valid and applicable to the linearly transformed observation vector $\mathbf{y}_p = \mathbf{Q}\mathbf{R}_{\mathbf{w}}^{-\frac{1}{2}}\mathbf{y}$. Therefore, (3.35) implements a GLRT test for the detection of unknown signals, whose observation vector is represented as $\mathbf{z}_p = diag[g'(\mathbf{w}_p)]\mathbf{s}_p$, in a Gaussian and uncorrelated background noise. In this case, notice that the transformation $\mathbf{Q}$ is unitary (a rotation), so the energy of the pre-whitened signal is conserved. Hence, the parameter α can still be considered an indicator of how PEED enhances the performance of the PED when the noise is non-Gaussian and non-independent.

3.3.2 Review of ICA model

ICA is a generalization of Principal Component Analysis (PCA) [50], whose goal is to find a linear transformation from an original data set with a mixture of components, so that the transformed data samples are statistically independent, or at least as independent as possible. As a result, ICA algorithms have proven successful in separating linear mixtures of independent source

signals in many applications, for example, biomedical data analysis (EEG, ERP, fMRI, optical imaging) and computational modeling, as reviewed in [83].

To define the ICA problem, it is convenient to use the vector-matrix notation. Thus, let $\mathbf{w}$ denote the random vector whose elements are the mixture samples $w_1, \ldots, w_N$, and $\mathbf{u}$ the random vector with elements $u_1, \ldots, u_N$. In addition, $\mathbf{A}$ shall denote the matrix with elements a_{ij}. Using this vector-matrix notation, the above mixing model is written as:

$$\mathbf{w} = \mathbf{A}\mathbf{u} \tag{3.36}$$

which is known as the ICA model [36]. The independent components cannot be directly observed and the mixing matrix is assumed to be unknown. The only observed data is the random vector $\mathbf{w}$, and both $\mathbf{A}$ and $\mathbf{u}$ must be estimated by using it. The starting point of the ICA is the very simple assumption that the components u_i are statistically independent, while the independent components must have non-Gaussian distributions. Therefore, after estimating $\mathbf{A}$, we can compute the inverse, denoted as $\mathbf{U}$, and thereby attain not only uncorrelated but also the independent components simply by equating:

$$\mathbf{u} = \mathbf{U}\mathbf{w}. \tag{3.37}$$

There are several approaches that can be implemented for the ICA estimation. Some are based on the minimization of the mutual information [5], but we shall focus on the maximum likelihood estimation. In this case, a statistical model of the data is often assumed; but when this assumption is inaccurate, the algorithms perform suboptimally (or even fail to produce the source separation), as commented in [9]. For this reason, an alternative method is turned to. It consists of employing a more flexible model for the PDF of the source signals, based on a non-parametric kernel density estimation technique of the PDF [78], as will be described in greater detail in the following section.

3.3.3 Estimating the transformation matrix U

First of all, let us offer some important comments about ICA identifiability. Problems with identifiability appear when ICA is applied to blind source separation (BSS): sources can be recovered up to some scaling factor and permutation, with the only constraint being a maximum of one Gaussian source [18]. In our case, ICA enters as a consideration in the context of detection not to separate sources, but rather to obtain transformed vectors with components as i.i.d. as possible. This transformation is achieved using a training set of dependent vectors; consequently, the constraints appearing in BSS become of no concern in this case. There are many algorithms in block or iterative versions which are able to obtain estimates of matrix $\mathbf{U}$.

Let us consider a maximum likelihood approach for this estimate. Thus, the set of training observation noise vectors $\mathbf{w}_k$ with $k = 1, \ldots, K$, are grouped into matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$. From now on, it will be assumed that $\mathbf{w}_k$ are independent observation vectors of non-independent noise. In practice, this means that vectors $\mathbf{w}_k$ must correspond to non-overlapped (and rather well separated) segments of the noise record, or, preferably, that different noise records be used for every $\mathbf{w}_k$. A very popular approach for estimating the ICA model is to find the maximum likelihood estimation of matrix $\mathbf{U}$, given matrix $\mathbf{W}$. Therefore, after employing logarithms, the log-likelihood function is defined as:

$$L(\mathbf{W}/\mathbf{U}) = \log p(\mathbf{W}/\mathbf{U}) = \sum_{k=1}^{K} \log p(\mathbf{w}_k/\mathbf{U}). \tag{3.38}$$

But, from (3.33) and using well-known properties of functions for random variables [77], (3.38) can be expressed as:

$$L(\mathbf{W}/\mathbf{U}) = \sum_{k=1}^{K} [\log |det\mathbf{U}| + \log p(\mathbf{U}\mathbf{w}_k)] \tag{3.39}$$

where maximization leads to the following expression

$$\begin{aligned} \frac{dL(\mathbf{W}/\mathbf{U})}{d\mathbf{U}} &= \sum_{k=1}^{K} \left[\frac{d \log |det\mathbf{U}|}{d\mathbf{U}} + \frac{(dp(\mathbf{U}\mathbf{w}_k)/(d\mathbf{U}))}{p(\mathbf{U}\mathbf{w}_k)} \right] \\ &= \sum_{k=1}^{K} \left[(\mathbf{U}^T)^{-1} - f(\mathbf{U}\mathbf{w}_k)\mathbf{w}_k^T \right]. \end{aligned} \tag{3.40}$$

In (3.40), assuming that the components of vectors $\mathbf{w}_{pk} = \mathbf{U}\mathbf{w}_k$ are i.i.d., the same non-linear scalar function $f(\cdot)$ can be separately applied to every component of the linearly transformed vector $\mathbf{w}_{pk}$ as follows:

$$f(\mathbf{U}\mathbf{w}_k) = [f(w_{pk1}), \ldots, f(w_{pkN})]^T. \tag{3.41}$$

Notice that assuming the same non-linear function for all the components guarantees that all the elements of the transformed vector have identical distribution. Similarly, separate application of the scalar function entails independence. Therefor, given the training set of dependent vectors, a transformation is obtained that leads to new vectors with components as i.i.d. as possible.

Equating (3.40) to zero, a set of non-linear equations with unknown matrix $\hat{\mathbf{U}}$ (estimated $\mathbf{U}$) can be obtained:

$$\hat{\mathbf{U}}^T = \left(\frac{1}{K} \sum_{k=1}^{K} f(\hat{\mathbf{U}}\mathbf{w}_k)\mathbf{w}_k^T \right)^{-1}. \tag{3.42}$$

For the Gaussian case, $f(\hat{\mathbf{U}}\mathbf{w}_k) = \hat{\mathbf{U}}\mathbf{w}_k$ and $\hat{\mathbf{U}}$ must satisfy the following expression:

$$\hat{\mathbf{U}}^T = \left(\frac{1}{K}\sum_{k=1}^{K}\hat{\mathbf{U}}\mathbf{w}_k\mathbf{w}_k^T\right)^{-1} = \left(\frac{1}{K}\sum_{k=1}^{K}\mathbf{w}_k\mathbf{w}_k^T\right)^{-1}\hat{\mathbf{U}}^{-1} \tag{3.43}$$

where

$$\hat{\mathbf{U}}^T\hat{\mathbf{U}} = \left(\frac{1}{K}\sum_{k=1}^{K}\mathbf{w}_k\mathbf{w}_k^T\right)^{-1}. \tag{3.44}$$

Note that for the Gaussian case, the solution for matrix $\hat{\mathbf{U}}$ can be expressed as:

$$\hat{\mathbf{U}} = \hat{\mathbf{U}}^T = \hat{\mathbf{R}}_{\mathbf{w}}^{-\frac{1}{2}} = \left(\frac{1}{K}\sum_{k=1}^{K}\mathbf{w}_k\mathbf{w}_k^T\right)^{-\frac{1}{2}}. \tag{3.45}$$

In general, however, iterative procedures [23, 28] are necessary to solve (3.42). For example, one possible solution is to use conventional gradient algorithms to obtain the following expression for matrix $\hat{\mathbf{U}}$:

$$\begin{aligned}\hat{\mathbf{U}}_{i+1} &= \hat{\mathbf{U}}_i + \beta\frac{\hat{d}L(\mathbf{W}/\mathbf{U})}{\hat{d}\mathbf{U}} \\ &= \hat{\mathbf{U}}_i + \beta\sum_{k=1}^{K}\left[(\hat{\mathbf{U}}_i^T)^{-1} - \hat{f}_i(\hat{\mathbf{U}}_i\mathbf{w}_k)\mathbf{w}_k^T\right].\end{aligned} \tag{3.46}$$

Considering this algorithm and the decomposition of $\mathbf{U}$ indicated in (3.34), convergence can be accelerated by first estimating $\mathbf{R}_{\mathbf{w}}$ from a set of training noise vectors, and subsequently employing a classical estimator as follows:

$$\hat{\mathbf{R}}_{\mathbf{w}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{w}_m\mathbf{w}_m^T. \tag{3.47}$$

Then, an iterative gradient algorithm is applied to the pre-whitened training vectors in order to estimate the rotation matrix $\mathbf{Q}$ in the iteration $i+1$, as follows:

$$\hat{\mathbf{Q}}'_{i+1} = \hat{\mathbf{Q}}_i + \beta\sum_{k=1}^{K}\left[\left(\hat{\mathbf{Q}}_i^T\right)^{-1} - \hat{f}\left(\hat{\mathbf{Q}}_i\mathbf{w}_{pk}\right)\mathbf{w}_{pk}^T\right] \tag{3.48}$$

where in this case $\mathbf{w}_{pk} = \hat{\mathbf{R}}_{\mathbf{w}}^{-\frac{1}{2}}\mathbf{w}_k$ are the pre-whitened training noise vectors. To estimate $\mathbf{Q}$, a unitary matrix must be implemented in each iteration, yielding the following normalization:

$$\hat{\mathbf{Q}}_{i+1} = \hat{\mathbf{Q}}'_{i+1}\left(\hat{\mathbf{Q}}'^T_{i+1}\hat{\mathbf{Q}}'_{i+1}\right)^{-\frac{1}{2}}. \tag{3.49}$$

The implementation of the iterative algorithm in (3.48) and (3.49) requires the estimation of function $\hat{f}_i(x) = \frac{\hat{p}'_i(x)}{\hat{p}_i(x)}$ at every iteration i. To perform such an estimate, it is necessary to first know the data distribution $\hat{p}(x)$. Hence, to reach a general applicability valid for any distribution, a non-parametric estimation of $p(x)$, as described in (3.22), is used. Let us consider the set of samples x_l, with $l = 1, \ldots, L$ corresponding to realizations of the random variable x (whose PDF must be estimated). In this case, the samples correspond to components of $\mathbf{Q}_i \mathbf{w}_{pk}$, $k = 1, \ldots, K$ obtained at every iteration; thus, a total number of $L = K \cdot N$ samples will be calculated. Therefore, the following expression is derived:

$$\begin{aligned} \hat{p}'(x) &= a \sum_{l=1}^{L} exp\left(-\frac{1}{2}\left(\frac{x - x_l}{h}\right)^2\right)\left(-\frac{x - x_l}{h}\right)\frac{1}{h} \\ &= \frac{a}{h^2}\left[\sum_{l=1}^{L} x_l \, exp\left(-\frac{1}{2}\left(\frac{x - x_l}{h}\right)^2\right)\right. \\ &\quad \left. - \sum_{l=1}^{L} x \, exp\left(-\frac{1}{2}\left(\frac{x - x_l}{h}\right)^2\right)\right] \end{aligned} \tag{3.50}$$

where x_l (with $l = 1, \ldots, L$) corresponds to a set of training samples of the data distribution to be estimated. Hence, the estimate of function $f(x)$ will be defined as:

$$\hat{f}(x) = \frac{1}{h^2}\left[x - \frac{\sum_{l=1}^{L} x_l \, exp\left(-\frac{1}{2}\left(\frac{x - x_l}{h}\right)^2\right)}{\sum_{l=1}^{L} exp\left(-\frac{1}{2}\left(\frac{x - x_l}{h}\right)^2\right)}\right]. \tag{3.51}$$

Using this expression it is possible to calculate $\hat{\mathbf{R}}_{\mathbf{w}}$ and $\hat{\mathbf{Q}}$ in (3.47) and (3.49), respectively, and, in turn, to obtain the final transformation matrix $\hat{\mathbf{U}}$.

3.4 Generalized matched subspace filter: GMSF

3.4.1 Introduction

The matched subspace filter (MSF) is known to be an uniformly most powerful (UMP) detector for the detection of a subspace signal in a background of uncorrelated Gaussian noise. Optimality of the MSF is kept even in the presence of subspace interferences [76]. Also in [22], similar detector solutions are

proposed in the presence of interference for specific types of independent non-Gaussian noise (called generalized Gaussian distributions). Unfortunately, there is no general UMP solution for the subspace signal detection problem when the noise is non-Gaussian and non-independent. Hence, suboptimal detectors will be devised.

A GLRT could be implemented, but it requires a maximum likelihood estimate (MLE) of the unknown parameters (in our case, μ and $\boldsymbol{\theta}$), which is not a very practical option. Other well-known suboptimal alternatives exist, such as the Wald and Rao tests [43]. The Wald test, although simpler to implement than the GLRT, also requires estimates of the involved parameters under H_1. The Rao test, in contrast, does not have such a requirement; in its most general form, only an MLE of the nuisance parameters under H_0 is required. Notice that in the problem considered here, the vector of parameters is given by $\boldsymbol{\theta}_r = \mathbf{0}$ (under H_0) and $\boldsymbol{\theta}_r = [\mu, \boldsymbol{\theta}^T]^T$ (under H_1); i.e., no nuisance parameters appear. Thus, the Rao test becomes an attractive alternative to deal with the non-Gaussian and independent noise case. It is worth mentioning that GLRT, Wald, and Rao tests are asymptotically equivalent in any case. Moreover, it has been recently shown that, for a finite number of observations, the three tests either coincide or are statistically equivalent in a number of typical detection problems [20, 21]. In particular, it is demonstrated in [21] that coincidence exists in detection problems without nuisance parameters when the observation PDF belongs to the exponential family, thus covering a broad range of practical cases.

3.4.2 Rao test generalization

Here, our attention turns to the Rao test. 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. This has been already shown in (3.1), and it can also be expressed in the form:

$$\frac{g(\mathbf{y})^T \mathbf{P} g(\mathbf{y})}{P_{g(w)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{3.52}$$

where $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the projection matrix onto the subspace defined by the columns of matrix $\mathbf{H}$.

Therefore, (3.52) is an extension of the MSF seen in Section 2.2.4, in the sense that a non-linear transformation $g(\cdot)$ is applied to the original observation vector prior to computing the normalized subspace energy. Note that for the Gaussian case $g(w) = w$; therefore, $g(\mathbf{y}) = \mathbf{y}$ and $P_{g(w)} = P_w = \sigma_w^2$. So, (2.14) and (3.52) are equivalent. In addition to the aforementioned equivalences with the GLRT and the Wald test, the Rao test has another interesting property: the statistic $g(\mathbf{y})^T \mathbf{P} g(\mathbf{y})/P_{g(w)}$ is χ_p^2 as in (2.14), hence the value λ can be easily computed for a required PFA. In the following,

the expression defined in (3.52) and found suitable for independent and non-Gaussian noise will be denoted by MSF.

3.4.3 Extension of the MSF

In this section, a further generalization of the MSF given in (3.52) is proposed in order to consider the most general case of non-Gaussian and non-independent noise. It is based on the use of ICA, which implements a linear matrix transformation to make the components of the transformed observation vector as independent as possible. The same idea was exploited in Section 3.3 to derive energy detectors in the presence of non-Gaussian and non-independent noise samples. Actually, ICA could be an option to implement a linear pre-processing step in any detection problem involving non-Gaussianity and statistical dependence among the observed vector components. However, each detector requires particular attention due to the specific implementation of the non-linear transformation $g(\cdot)$. An "ad hoc" non-linear function was proposed in Section 3.2.3 to make the linearly transformed observation vector as Gaussian as possible, but the GLRT condition of the extended energy detector used was not entirely demonstrated. Nevertheless, this is not the case in the detection problem considered in this section since the non-linear transformation required in (3.52) is inherent to the Rao test (defined in (3.3)), thereby preserving its properties.

Test (3.52) assumes that the components of $\mathbf{w}$ are i.i.d. When this is not the case, the non-independent observation vector could be transformed into a new one having independent components. This can be done by means of ICA and an appropriate linear transformation leading to a new generalization of the MSF that will be termed hereafter as generalized matched subspace filter (GMSF) [63]. 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:

$$\frac{g(\mathbf{U}\mathbf{y})^T\mathbf{P}_U g(\mathbf{U}\mathbf{y})}{P_{g(u)}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \tag{3.53}$$

where $P_{g(u)}$=$E[g^2(u)]$, and the pre-processed subspace matrix can be expressed as:

$$\mathbf{P}_U = \mathbf{H}_U(\mathbf{H}_U^T\mathbf{H}_U)^{-1}\mathbf{H}_U^T \quad \text{with} \quad \mathbf{H}_U = \mathbf{U}\mathbf{H}. \tag{3.54}$$

Notice that the non-linear transformation appearing in (3.53) must be defined from the PDF of the linearly transformed noise samples u, which generally have a different (non-Gaussian) PDF from the original noise w. Hence, using (3.3), and changing w to u, we obtain:

$$g(u) = -\frac{\dfrac{dp(u)}{du}}{p(u)}. \tag{3.55}$$

Thus, it is guaranteed that equation (3.53) implements a Rao test in the linearly transformed observation vector $\mathbf{y}_p$=$\mathbf{U}\mathbf{y}$, while maintaining all the mentioned properties of the Rao test such as the χ^2_p distribution of the statistic $g(\mathbf{U}\mathbf{y})^T\mathbf{P}_U g(\mathbf{U}\mathbf{y})/P_{g(u)}$, and the asymptotic equivalence with the GLRT and the Wald test.

Different criteria have been proposed to estimate the required ICA transformation, but in essence, all of them are trying to minimize some appropriate measure of the dependence between the vector components. In particular, the same approach as the one presented in Section 3.3.1 (where a matrix $\mathbf{U} = \mathbf{Q}\mathbf{R}_{\mathbf{w}}^{-\frac{1}{2}}$ was applied to the dependent data) is used.

Equation (3.48) and test (3.53) require knowledge of function $\hat{f}(.)$ and the pre-processing function $g(\cdot)$, respectively. In this case, both functions are equivalent and depend on the noise PDF $p(u)$ as shown in (3.55). If there is *a priori* knowledge of $p(u)$, the non-linear function $g(\cdot)$ can be directly computed. But, in general, there will not be any available knowledge about the PDF of u; hence, a non-parametric approach becomes necessary. Let us consider the set of samples $\{u_l\}, l = 1 \dots L$, corresponding to realizations of the random variable u (whose PDF is to be estimated). The classical non-parametric estimator takes the form $\hat{p}(u)$ of (3.22). Assuming that the quality of the estimate $\hat{p}(u)$ is appropriate to consider $p(u) \cong \hat{p}(u)$, the non-linear pre-processing function $g(\cdot)$ can be computed as in (3.55), thereby leading to:

$$g(u) = \frac{1}{h^2}\left[u - \frac{\sum_{l=1}^{L} u_l \cdot \exp\left(-\frac{1}{2}\left(\frac{u-u_l}{h}\right)^2\right)}{\sum_{l=1}^{L} \exp\left(-\frac{1}{2}\left(\frac{u-u_l}{h}\right)^2\right)}\right]. \tag{3.56}$$

One may identify in (3.56) a linear and a non-linear term. The latter accounts for the possible non-Gaussianity of the random variable u.

3.5 Evaluation results of the EED in non-Gaussian independent noise

In order to illustrate the performance of the two possible transformations described in Section 3.2 (parametric Box-Cox and non-parametric) and which can be implemented when using the EED, a number of experiments using a variety of simulated signals and noises of different characteristics were performed. First, the procedure followed in estimating the power transformation parameter (β), required for the Box-Cox transformation, is detailed. Then, both methods were compared, analyzing the performance of the two non-linear functions, and finally, a conclusion was drawn about which was most

suitable for the EED, when confronted with the general task of detecting unknown signals in the presence of non-Gaussian and independent noise. In both methods, the new ROC factor explained in Section 2.3.1 was used to compare the performance of different detectors.

3.5.1 Estimation of the power transformation parameter

In situations where an assumption of noise normality is not guaranteed, one of the possible solutions is to transform the data such that the distributions are nearer to the Gaussian assumption. Box and Cox proposed a parametric power transformation to reduce departures from normality, as shown in Section 3.2.2. Although there is a family of possible variants that can be implemented, our study is focused on analyzing the EED when implementing the original transformation described in (3.13). Since this transformation is parametric dependent, the usual practice when using some of these techniques is to first estimate the parameter β.

As mentioned in Section 3.2.2, the transformation parameter β is found experimentally by observing the detection results of the EED. Following an evaluation of the resultant ROC curves (indicative of a successful transformation of the non-Gaussian noise), the optimum parameter is then selected from the best EED performance. The following simulations are focused on estimating the most suitable β parameter, required to detect a Gaussian signal among different independent and non-Gaussian noises (i.e. Laplacian, Rayleigh, Gamma and Uniform distributions). A total of $40 \cdot 10^3$ vectors and $2 \cdot 10^3$ different PFAs from 0 to 10^{-10} were considered in order to calculate the experimental ROC curves of the EED. The range of PFA used to compute the ROC factor was set between 0 and 0.1, since this is a reasonable operation area for a detector.

Figure 3.6 illustrates the ROC factor of the EED (using the original Box-Cox variant) for different β values in comparison with an ED, but under the same conditions. Four different non-Guassian noises were used and, due to the variety of the examples selected, a wide interval of β values were tested, extending from -3 to 3. Several SNR were simulated (from 6 dB to -6 dB) but since the same performance was observed in all cases, only the results for $SNR = -2$ dB are presented. These results reveal that the β parameter is fairly related to the characteristics of the non-Gaussian noise in which the signal is immersed. For this particular Box-Cox variant, when $\beta > 0$, positive ROC factors were achieved; this means that in the remaining cases the EED performs poorer than the ED. Upon observing these curves, it is possible to estimate the best β parameter for a particular non-Gaussian noise by selecting that which yields the maximum ROC factor.

Table 3.1 gives the optimum β values for each non-Gaussian noise. As can be observed, there is no common value for all of them, and the precision with which this parameter is estimated depends greatly on the noise

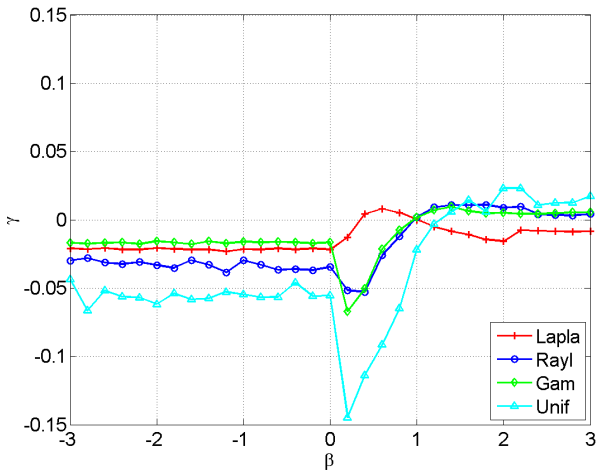


Figure 3.6: ROC factor (γ) for different β values when using the original Box-Cox variant for detecting a Gaussian signal in presence of several non-Gaussian noises with a $SNR = -2$ dB and $N = 25$.

characteristics. We must also notice that, in principle, the EED is expected to outperform the ED since the former can address the detection problem in the presence of non-Gaussian noise. However, it is possible that even when the optimum β is used, the EED curve will not follow this behavior. The explanation for this effect is related to the fact that when applying the Box-Cox transformation, we cannot generally ensure that the transformed observations are independent and normally distributed. To some extent the reasonableness of this assumption can be tested with a comparative analysis of the ROC curves generated by both the EED and the ED, as shown in the next section. It follows that the corresponding ROC factor between both detectors can be considered an indicator of the effectiveness of the Box-Cox transformation.

	Laplacian	**Rayleigh**	**Gamma**	**Uniform**
β	0.6	1.8	1.4	2

Table 3.1: Optimum β from the original Box-Cox transformation when detecting a Gaussian signal in the presence of different non-Gaussian and independent noise distributions.

3.5.2 Comparison of the EED using the non-parametric and parametric Box-Cox transformation

In the previous section, simulation results led to the derivation of the optimum β to be used with the parametric Box-Cox transformation for each particular non-Gaussian noise tested. Now, the non-parametric approach for the EED proposed in 3.2.3 must also be analyzed in order to compare both methods. Thus, several simulations were performed to study the unknown signal detection problem in the presence of non-Gaussian and independent noise, when using the EED.

First, let us describe in greater detail the implementation employed for the non-parametric EED. In this case, the non-linear function described in (3.24) was utilized and, based on a set of training samples, the noise PDF was estimated as proposed in (3.22). In this particular experiment, a total number of $10 \cdot 10^3$ noise samples were arbitrarily selected and used in a previous training step. Thereafter, several experiments with different SNR, spanning from -6 dB to 4 dB, were performed (although only a portion of the results are shown here). In addition, four different noise distributions (Laplacian, Rayleigh, Gamma, and Uniform) were tested.

Figure 3.7 illustrates the resultant ROC curves of the EED after applying both the non-parametric (EED*nopar*) and the Box-Cox transformation (EED*boxcox*). A Gaussian signal is detected in the presence of the four different non-Gaussian noises with independent samples. In addition, the ROC curves of the ED in the same noise conditions are likewise shown to facilitate a useful comparison of relative performance. As expected, the EED utilizing both non-linear functions outperformed the ED when the noise did not follow a normal distribution. Furthermore, one can observe how the improvement reached by the non-parametric EED is more significant than that reached with the parametric EED over the ED. This improvement is related to the α parameter, obtained for the non-parametric case using (3.29) and given in Table 3.2 for each non-Gaussian noise. Not surprisingly, high values of α are associated with large improvements in the EED ROC curve over the ED curve.

	Laplacian	**Rayleigh**	**Gamma**	**Uniform**
α	1.3055	1.3732	3.9912	2.0355

Table 3.2: Enhancement factor α obtained for the non-parametric EED when detecting a Gaussian signal in the presence of different non-Gaussian and independent noise distributions.

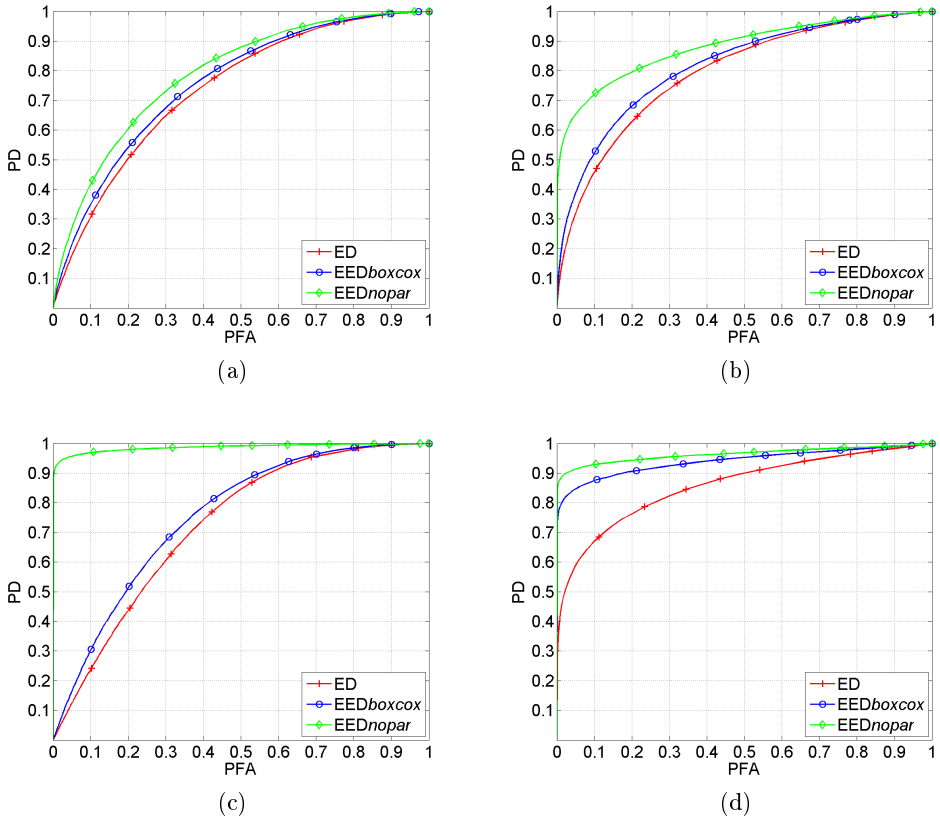


Figure 3.7: ROC curves of the EED using the Box-Cox (EED*boxcox*) and the non-parametric approach (EED*nopar*). Generated when detecting a Gaussian signal in the presence of different noise distributions with $SNR = -4$ dB and $N = 25$. (a) Laplacian noise. (b) Rayleigh noise. (c) Gamma noise. (d) Uniform noise.

In Table 3.3, a more precise evaluation of both the EED and the ED is presented. The objective of this table is to summarize the behavior of the EED versus the ED, and to analyze in greater detail the ROC curves shown in Figure 3.7. It must first be pointed out how the ROC factor obtained is positive when comparing both the EED*nopar* as well as the EED*boxcox* with the ED. This means that the corresponding ROC curves of the EED lie above those of the ED. Then, it is also important to remark that the maximum values are reached for the Gamma noise in both cases. In light of the ROC curves shown in Figure 3.7, this result falls in line with our expectations. Furthermore, greater values of the factor are achieved with the EED*nopar* than with the EED*boxcox*, which is indicative of better performance under the same conditions. Finally, when comparing the ROC curves of the two

EED transformations, the ROC factor is again above 0; hence, it can be concluded that the EED*nopar* performs better in all noise conditions.

Comparison	***Laplacian***	***Rayleigh***	***Gamma***	***Uniform***
EED*nopar* Vs ED	0.0142	0.2101	0.4392	0.3025
EED*boxcox* Vs ED	0.0088	0.0237	0.0117	0.2411
EED*nopar* Vs EED*boxcox*	0.0079	0.1731	0.4373	0.0651

Table 3.3: ROC factor comparing the ED with the EED non-parametric and the EED using the Box-Cox transformation with $SNR = -4$ dB and $N = 25$.

It has been demonstrated that the non-parametric transformation used in the EED performs better based on the improved ROC curves. However, other considerations must be taken into account when deciding which of the two approaches to implement. Although the implementation of the Box-Cox transformation is much less time consuming, it is worthwhile in this case to use the non-parametric approach as it can adapt to changes in noise characteristics. Therefore, when referring to the EED, only the non-parametric transformation will be used.

3.6 Performance evaluation of the PEED

In Section 3.5, the problem of detecting unknown signals in the presence of non-Gaussian noises was evaluated. In that case, the samples of the non-Gaussian noise were assumed to be i.i.d.. When this is not the case, simple pre-whitening is insufficient; a more complex detector must be implemented, as detailed in Section 3.3. This detector was termed a PEED, and this section presents a precise analysis of its performance when the noise is not only non-Gaussian, but also non-independent. In addition, it will be compared with other detectors, such as the ED and the PED (described in Chapter 2), in order to examine the differences between them in non-independent scenarios, as well as with independent noise samples.

3.6.1 Non-independent noise samples

In the first simulation setup, the detector evaluation required the generation of random variables corresponding to non-Gaussian and non-independent noise, denoted w_d. They were obtained by first generating non-Gaussian independent random variables w_i, leading to noise vectors $\mathbf{w}_i$. They were subsequently transformed as follows: $\mathbf{w}_d = \mathbf{U}^{-1}\mathbf{w}_i$. The elements of the mixture matrix $\mathbf{U}^{-1}$ (dimension $N \times N$) were calculated from a random variable with a standard uniform distribution across the open interval (0,1).

A total of $3{\cdot}10^4$ noise vectors were generated for each simulation. In addition, a Gaussian signal was generated and different SNR conditions covering the range of -8 dB to 4 dB were simulated.

In order to evaluate the PEED, it was assumed that both the estimate of the linear transformation $\mathbf{U}$ and the pre-processing function $g(\cdot)$ required in (3.35) were obtained in a previous training step where only noise was present. Thus, selecting the number of noise training vectors K used to estimate of $\mathbf{U}$ becomes an important issue of practical interest. Obviously, good estimates of $\mathbf{U}$ require a relatively high value of K. However, it cannot be made arbitrarily large as there are some limiting factors, namely, the computational burden and the time interval duration allowed for training. In general, this decision may be considered part of the overall calibration of the detector for each application, but unfortunately, it is not easy to find a closed analytic equation providing the best values. Instead, an experimental fitting was used, taking the value that resulted from the best performance (in terms of the ROC curves). For this task, a total of $2 \cdot 10^3$ training vectors were used.

Figure 3.8 summarizes the performance of the PEED when detecting a Gaussian signal in the presence of four different non-Gaussian and non-independent background noises, with $SNR = -4$ dB. In addition, the ROC curves of the PED and the EED are also shown to better compare and understand the performance of the PEED. First, it is important to notice how the EED offers a truly poor performance: a random detector for all noise conditions resulted, as it could not address the noise sample dependence. Thus, let us examine the performance of the PED. As described in (2.22), this detector applies a simple pre-whitening to the observation vector and, as can be observed, the ROC curve experiments a considerable improvement in performance, particularly when viewed alongside the EED. However, the results of this solution remain insufficient: when the non-Gaussian noise samples are dependent, not only pre-whitening but also ICA must be applied. This requirement makes using the PEED more attractive, as its ROC curve presents the best results of all non-Gaussian examples. The improvement reached is related to the enhancement factor α, given in brackets for each case: Laplacian ($\alpha = 1.2998$), Rayleigh ($\alpha = 1.3196$), Gamma ($\alpha = 3.9956$), and Uniform ($\alpha = 2.0546$).

Table 3.4 summarizes the results obtained in Figure 3.8 after evaluating the performance of the PEED versus the EED and the PED, on the basis of the ROC factor. As observed, and concurring with the results illustrated in Figure 3.8, the calculated ROC factor revealed that the PEED outperforms the other two detectors, in particular for the Gamma and Uniform noises.

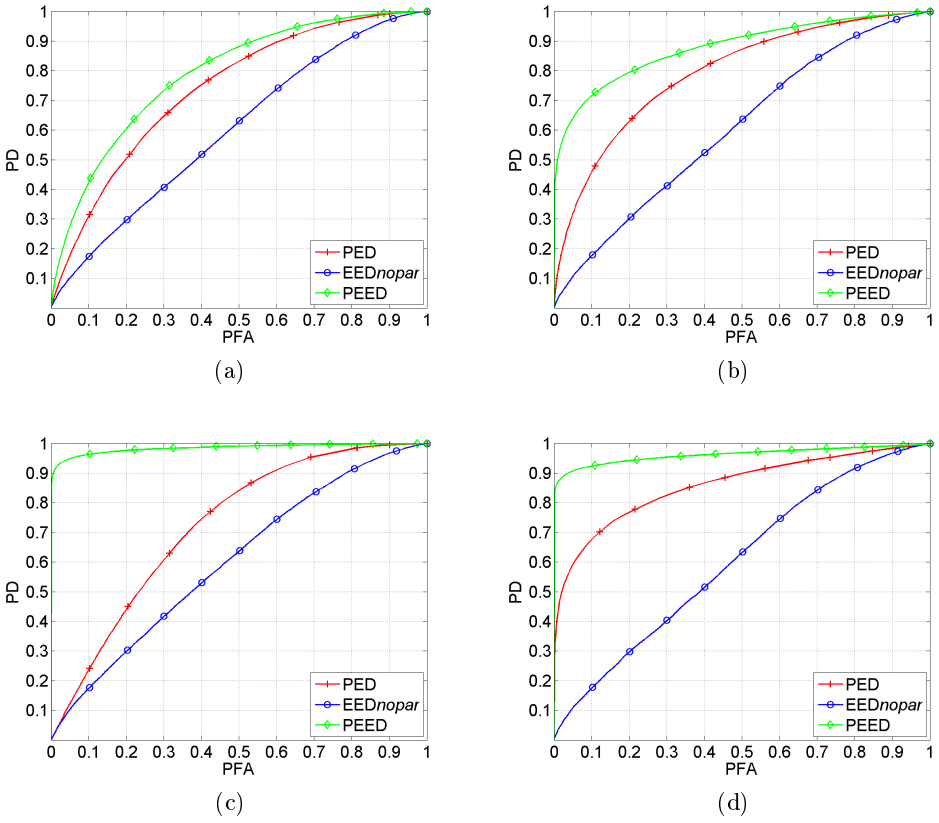


Figure 3.8: ROC curves comparing the PEED with the PED and the EED. The non-parametric approach was used to detect a Gaussian signal in the presence of different noise distributions with dependent samples, $SNR = -4$ dB, and $N = 25$. (a) Laplacian noise. (b) Rayleigh noise. (c) Gamma noise. (d) Uniform noise.

Comparison	***Laplacian***	***Rayleigh***	***Gamma***	***Uniform***
PEED vs PED	0.0177	0.1813	0.4337	0.3048
PEED vs EED	0.0398	0.2251	0.4355	0.4218

Table 3.4: ROC factor comparing the PEED with the PED and the EED, using the non-parametric transformation with $SNR = -4$ dB and $N = 25$.

3.6.2 Independent noise samples

To conclude with our evaluation of the PEED, let us study its performance when the noise is non-Gaussian, yet independent. In this case, no pre-whitening nor linear transformation are required and thus the estimation

of matrix $\mathbf{Q}$ should tend toward the identity $\mathbf{I}$, while $\hat{\mathbf{R}}_{\mathbf{w}}$ shall only represent normalization.

Figure 3.9 represents the ROC curves of the EED and the PEED are represented when detecting a Gaussian signal in the presence of the four previously mentioned non-Gaussian noises. It is possible to observe here how the PEED performance is similar to that of EED, and as expected, consistently better than the ED. Therefore, under these particular noise conditions, we can conclude that both detectors are equivalent; thus, the general applicability of the PEED in the detection of unknown signals in the presence of any kind of noise is demonstrated.

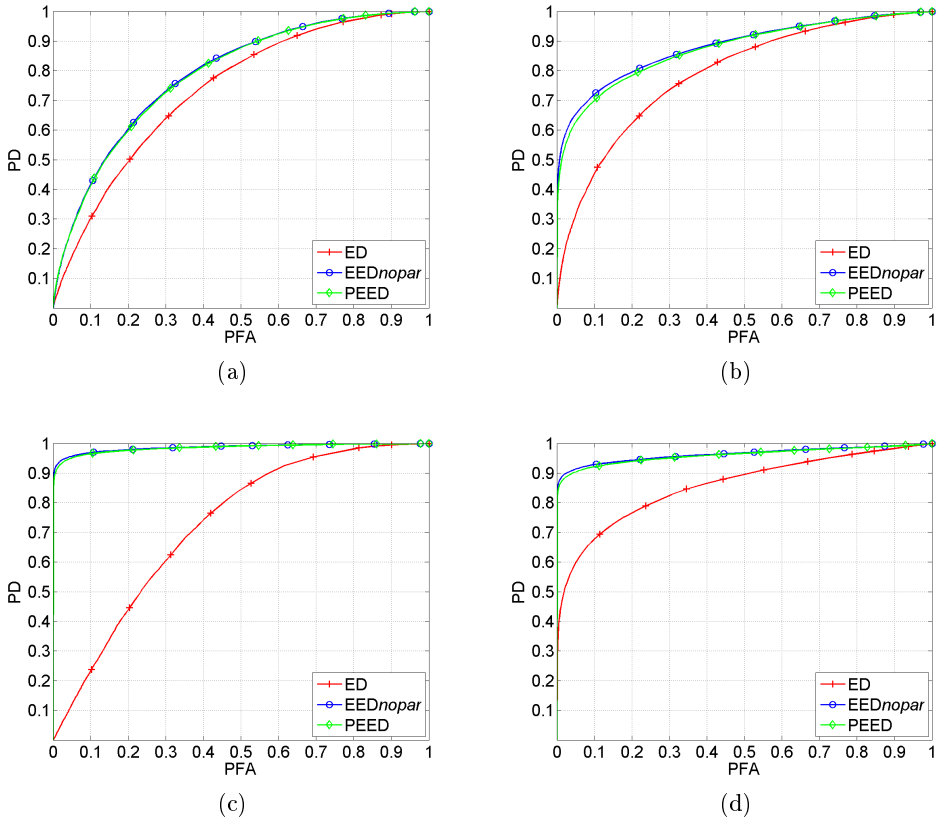


Figure 3.9: ROC curves comparing the PEED with the ED and the EED. The non-parametric approach was used to detect a Gaussian signal in the presence of different noises with independent samples, $SNR = -4$ dB and $N = 25$. (a) Laplacian noise distribution. (b) Rayleigh noise distribution. (c) Gamma noise distribution. (d) Uniform noise distribution.

3.7 Experimental results of the GMSF

Finally, several experiments were performed to verify the theoretical derivations and performance of the proposed GMSF. In this case, the signal was assumed to be in a subspace, but in the presence of non-Gaussian and non-independent noise. In order to evaluate the improvement of the GMSF versus the MSF, different experiments were conducted using (3.53) and (3.52), respectively.

3.7.1 Signal model

Subspace signals will be considered band-limited, i.e., formed by the sum of one or more sinusoids as defined in (3.57). This is a specific subspace signal, and it is of particular interest to examine the expected improvements when using the GMSF, as will be seen in the experiments performed. Hence, the detection problem is expressed as follows:

$$\begin{aligned} H_0 &: y[n] = w[n] \\ H_1 &: y[n] = \sum_{m=1}^{M} A_m \cos(2\pi m f_0 n + \phi_m) + w[n] \end{aligned} \tag{3.57}$$

where $n = 0, 1, \ldots, N-1$, and $m = 1, \ldots, M$ represents the number of sinusoids used. The amplitude A_m and phase ϕ_m are assumed to be unknown, f_0 is assumed to be known, and $w[n]$ are the non-independent and non-Gaussian noise samples. It is possible to rewrite the data model in the linear model form as $\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, with the subspace matrix given by:

$$\mathbf{H} = \left[\begin{pmatrix} 1 & 0 \\ \cos[\omega_1] & \sin[\omega_1] \\ \vdots & \vdots \\ \cos[\omega_1(N-1)] & \sin[\omega_1(N-1)] \end{pmatrix} \Bigg\| \begin{matrix} \cdots \\ \cdots \\ \\ \cdots \end{matrix} \right.$$
$$\left. \begin{matrix} \cdots \\ \cdots \\ \\ \cdots \end{matrix} \Bigg\| \begin{pmatrix} 1 & 0 \\ \cos[\omega_M] & \sin[\omega_M] \\ \vdots & \vdots \\ \cos[\omega_M(N-1)] & \sin[\omega_M(N-1)] \end{pmatrix} \right]. \tag{3.58}$$

Here each sinusoid ω_m=$2\pi m f_0$ is represented by two column vectors, and the operator $\|$ simply signifies column augmentation; that is, the abutting parentheses are removed to offer a wider matrix from the operands of the operator $\|$. Using the same notation, the unknown parameter vector $\boldsymbol{\theta}$ can be expressed as $\boldsymbol{\theta} = [(\alpha_1\beta_1)\| \ldots \|(\alpha_M\beta_M)]^T$, where $\alpha_m = A_m \cos\phi_m$ and $\beta_m = -A_m \sin\phi_m$.

3.7.2 Experimental setup

On the one hand, the detector evaluation required the generation of random variables corresponding to non-Gaussian and non-independent noise. They were obtained in a manner similar to that previously described in Section 3.6.1 for the PEED. A total of $3 \cdot 10^4$ noise vectors were generated for each simulation. On the other hand, generation of the subspace signals was also required. In this case, the frequency f_0 was set to 0.1, ϕ_m were samples of a uniform random variable between $[-\pi, \pi]$, and A_m were selected for every required SNR, defined as:

$$SNR = \sum_{m=1}^{M} \frac{|A_m|^2}{\sigma^2_{\mathbf{w}_i}}. \tag{3.59}$$

To evaluate the GMSF, it was assumed that the estimate of the linear transformation $\mathbf{U}$ and the pre-processing function $g(\cdot)$ required in (3.53) were obtained as part of a previous training step where only noise was present. As mentioned in the previous section, selecting the corresponding number of noise training vectors K is a challenging task since many factors can influence this decision. However, it is also possible to use an experimental fitting in order to select the number of noise training vectors K, used in the estimation of $\mathbf{U}$. In this case, we took the value of the best ROC curve performance. For example, Figure 3.10 shows the results achieved for non-Gaussian and non-independent noise $\mathbf{w}_d$ (generated from Laplacian PDF), with $SNR = -2$ dB, $N = 25$, $M = 1$, and different K setups. It can be observed how for $K > 2 \cdot 10^3$, there is no significant improvement in the ROC curve. Similar results were achieved by varying SNR, N and M; therefore, in our experiments K was set to $2 \cdot 10^3$.

Once $\hat{\mathbf{U}}$ was obtained, it was then necessary to estimate the non-parametric function $g(\cdot)$. To do so, the estimated linear transformation $\hat{\mathbf{U}}$ was applied to the K available noise training vectors ($\mathbf{w}_d$) as follows: $\hat{\mathbf{u}} = \hat{\mathbf{U}}\mathbf{w}_d$. This yielded K linearly pre-processed vectors $\hat{\mathbf{u}}$, and to a total set of $K \cdot N$ independent samples of the random variable u, with PDFs similar to the original w_i. The parameter L was set based upon experiments similar to those conducted to determine the most suitable value of K. To reduce computational requirements, only a subset of $L = NK/4$ randomly selected samples were used to estimate $g(\cdot)$, using (3.55).

3.7.3 Comparing GMSF and MSF

Several experiments were performed in order to assess the improvements of the GMSF with respect to the MSF by varying different parameters involved in the detection problem.

First, it is of particular interest to observe how the GMSF behaved in the presence of independent non-Gaussian noise ($\mathbf{w}_i$). Figure 3.11 presents the

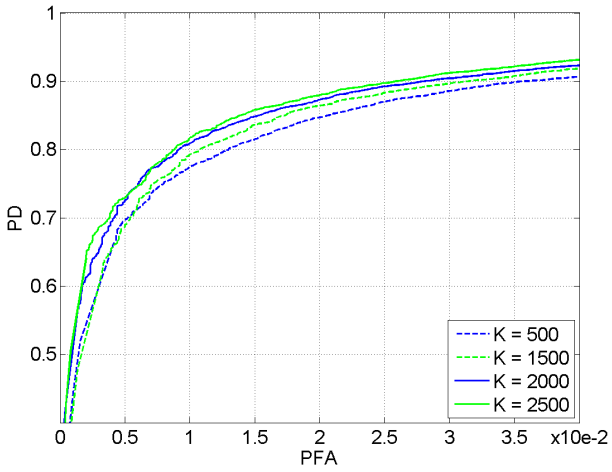


Figure 3.10: ROC curves of GMSF for non-Gaussian and non-independent noise, generated from Laplacian noise PDF with $SNR = -2$ dB, $N = 25$, $M = 1$, and different K setups.

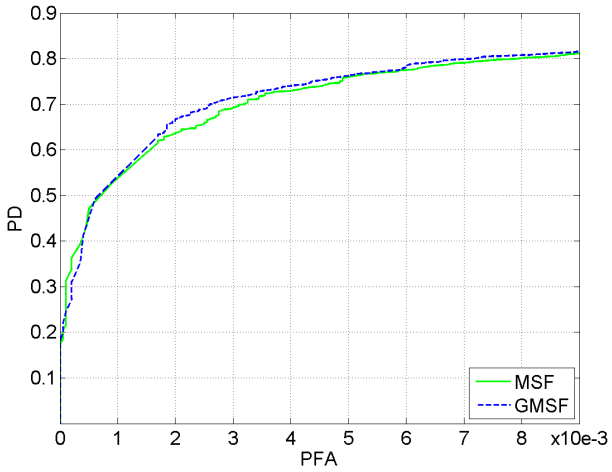


Figure 3.11: ROC curves of MSF and GMSF for independent Laplacian noise ($\mathbf{w}_i$), with $SNR = -2$ dB, $N = 25$, and $M = 1$.

ROC curves of both detectors when using independent Laplacian noise for $SNR = -2$ dB, $N = 25$, and $M = 1$. It can be observed that, as expected, both detectors behaved similarly. Secondly, Figure 3.12 shows the ROC curves obtained with the MSF and the GMSF when utilizing different types of non-independent noises ($\mathbf{w}_d$) for $SNR = -6$ dB, $N = 25$, and $M = 1$.

As described in the previous sections, these dependent noise samples corresponded to linearly transformed independent noises having non-Gaussian PDFs: Rayleigh, Laplacian, and Gamma. Furthermore, the ROC curve of the MSF is also represented for the same parameters and non-Gaussian noise PDFs, but with independent samples ($\mathbf{w}_i$). A comparison of the three curves

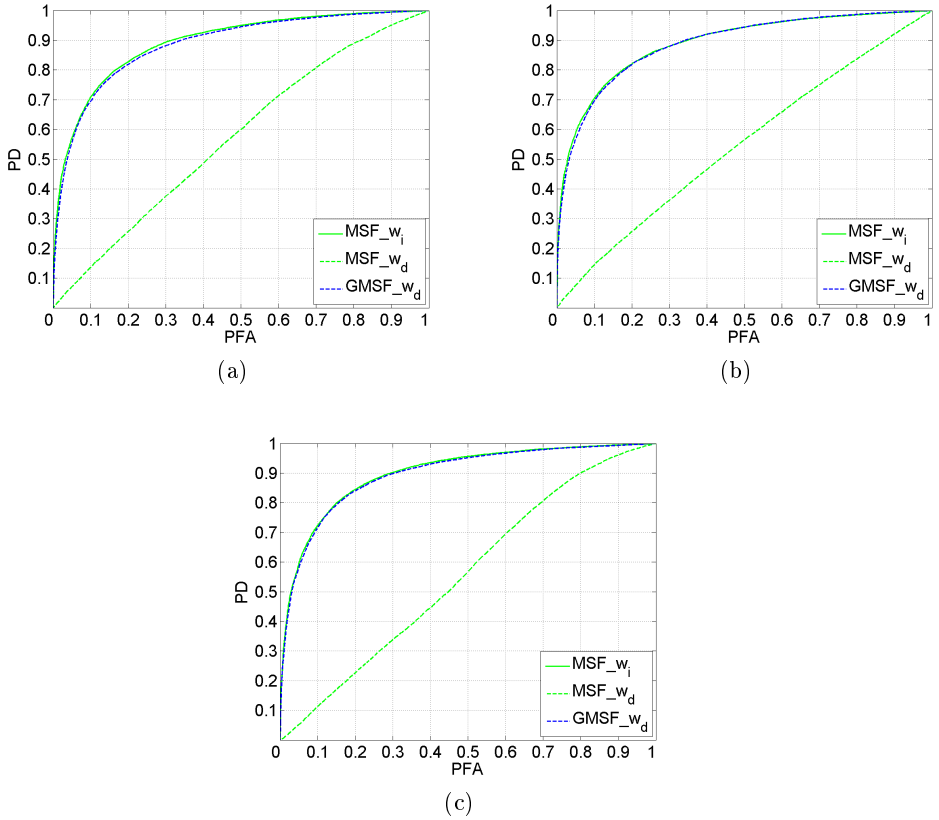


Figure 3.12: ROC curves of MSF and GMSF for different types of non-independent ($\mathbf{w}_d$) and independent ($\mathbf{w}_i$) non-Gaussian noise distributions, with $SNR = -6$ dB, $N = 25$, and $M = 1$. (a) Laplacian noise distribution. (b) Rayleigh noise distribution. (c) Gamma noise distribution.

across all noise distributions finds that the MSF detector experiments a considerable deterioration with non-independent noise, while the GMSF curves practically coincide with those corresponding to the MSF for independent noise. Therefore, the results show the capability of the GMSF to compensate for the degradation of the MSF, which itself becomes a random detector in the presence of dependent noise. The above results are of particular importance, since it is possible to demonstrate the generalization property of the GMSF: it significantly improves the MSF performance in the presence

of non-independent noise (Figure 3.12) and it behaves like the MSF in the presence of independent noise (Figure 3.11).

The previous results are expanded in Figure 3.13 by gauging the respective influence of different experimental parameters. Figures 3.13a, 3.13b, and 3.13c show the ROC curves of the GMSF and the MSF with non-Gaussian and non-independent noise (generated from Laplacian PDF) for different values of SNR, N, and M, respectively. The improvements of the GMSF with respect to the MSF become evident across all the examples, and the influence of the parameter value follows this expected behavior. First, Figure 3.13a illustrates how the PD increases with the SNR. It must be noticed that this increase is not very significant for the MSF, thus indicating that the SNR should be much higher to compensate for the degradation of the MSF in

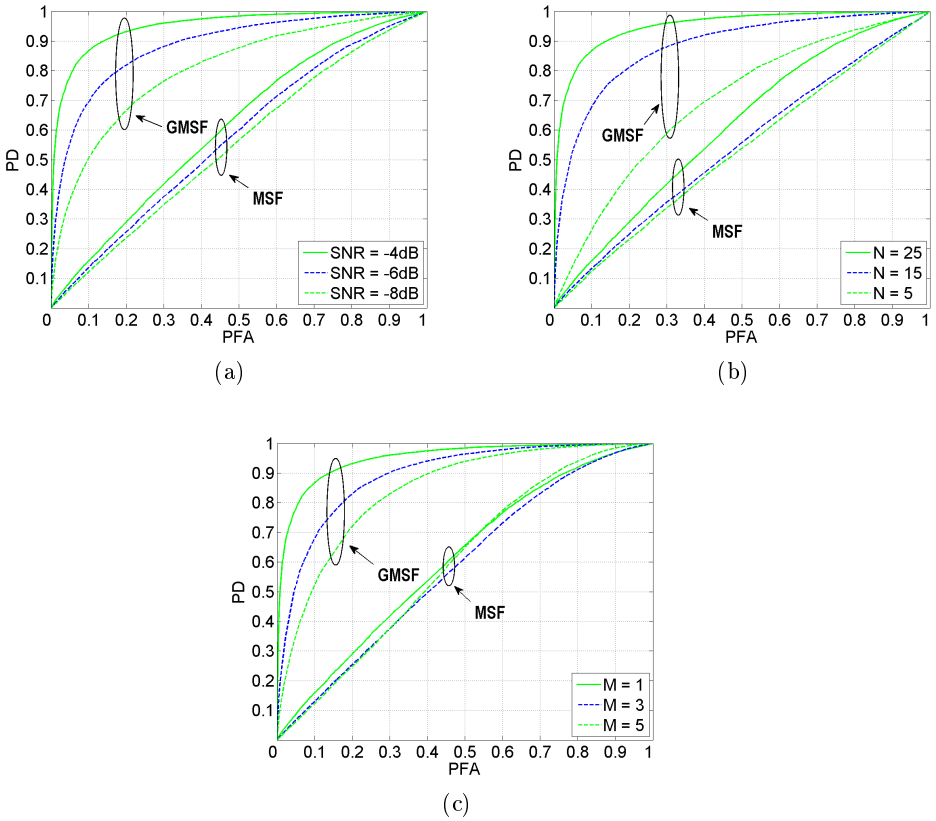


Figure 3.13: ROC curves of MSF and GMSF with non-Gaussian and non-independent noise distributions generated from Laplacian noise PDF. (a) Different SNR, $N = 25$ and $M = 1$. (b) Different observation vector length (N), $SNR = -4$ dB and $M = 1$. (c) Different subspace matrix dimension (M), $SNR = -4$ dB and $N = 25$.

the presence of dependent noise. Similarly, Figure 3.13b reveals that, as expected, the PD also improves with the observation size N, since the normal behavior of any detector implies that the test statistic increases the signal to noise ratio gain with N. Again, it can be observed how the enhancement in PD (when using the MSF) is not as significant as that obtained with the GMSF. Finally, Figure 3.13c shows the influence of varying M, the number of involved sinusoids (i.e., varying the subspace dimensions). As expected, the PD increases in the case of the GMSF when the subspace becomes more restrictive (reduced dimensions), which means that the signal bandwidth becomes narrower, although this effect is insignificant in the MSF when it is compared with the GMSF.

As a general conclusion, the degradation of the MSF brought on by the presence of non-independent noise cannot be easily compensated for by raising the SNR, by increasing the observation size N, nor by reducing the signal subspace dimension M. Therefore, use of the GMSF is perceived as a practical solution to this problem.

3.8 Conclusions

This chapter has addressed the detection of totally unknown signals and signals in a subspace (in both cases, with non-Gaussian and non-independent noise).

When the noise is non-Gaussian, the Rao test was examined as a suboptimum solution to the detection problem. However, it was demonstrated that in some particular situations it fails to offer a satisfactory solution. Therefore, alternative non-linear functions were proposed, leading to an extended version of the ED, termed EED. Two possible transformations were studied: one that implements the parametric Box-Cox transformation, and the other based on the non-parametric estimation of the data PDF to be transformed. An analysis of the best Box-Cox parameter was conducted for different non-Gaussian distributions by means of an ROC factor evaluation. The two transformations were compared for the same conditions, and it was observed how the non-parametric solution offered better detection results in terms of PD. Thus, it was decided to use it when implementing the EED. In this latter case, a new enhancement factor α was defined, indicating possible changes in SNR induced by the non-linear transformation.

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 $\mathbf{U}$ that makes as much i.i.d. as possible the components of the data vectors.

Thus, a detailed evaluation was carried out in order to compare the performance of all the generalizations and extensions of the ED. First, the

EED and the ED were evaluated across four different non-Gaussian noise distributions with independent samples. By calculating the enhancement factor α for each case, the expected improvement attained by the EED was demonstrated to be greater than the ED. Similar simulations were then carried out with non-independent noise samples. In this case, the PEED revealed the best results in comparison with the EED and the PED. A final simulation showed the performance results of the GMSF versus the MSF for non-independent non-Gaussian noise and for different parameter setups. All cases showed that the MSF suffers a performance degradation in comparison with the GMSF in the presence of non-independent noise.

Chapter 4

Structure of multiple energy detectors (MED)

Courage and perseverance have a magical talisman, before which difficulties dissapear and obstacles vanish into air.

John Quincy Adams

In this chapter, a novel approach is proposed to overcome the detection problem introduced by signal duration uncertainty. Instead of using only one detector for every possible signal duration, the implementation of multiple energy detectors (MED), matched to different possible signal durations is introduced. Among the possible segmentation strategies that could be studied, we specifically concentrate on the study of a particular case of successive segmentations of the original observation vector. One of the main goals of this chapter is to obtain the receiver operating characteristic (ROC) curves of this structure in time and frequency domains, and to study the possible improvements offered by this new method when compared with the use of a single energy detector (ED).

4.1 Detection of signals with unknown duration

In the context of novelty or event detection, where the characteristics of the signals to be detected are not known, there is also a lack of information about the signal duration since the environment under study is susceptible to any type of event. The selection of the temporal duration of the observation vector is a very challenging task that may significantly affect detection.

Therefore, an approach based on the use of multiple energy detectors, each of which is matched to different observation duration and bandwidth is considered. Thus, the simplicity of the energy detectors is maintained and the complicated approach employed by the non-linear functions used in [45] is likewise avoided.

4.1.1 Revision of the ED

As described in Chapter 2, the ED is a good solution for detecting unknown signals in the presence of a background noise, and it is also generalized likelihood ratio test (GLRT) for the case of white Gaussian noise. In these conditions, under hypothesis H_0, the resulting statistic $\mathbf{y}^T\mathbf{y}/\sigma^2_{\mathbf{w}}$ is chi-squared distributed with N degrees of freedom (χ^2_N), where N corresponds to the observation vector dimension. Under hypothesis H_1, the resulting statistic is non-central chi-squared distributed with parameters of non-centrality equal to the signal-to-noise ratio defined as $SNR = \mathbf{s}^T\mathbf{s}/\sigma^2_{\mathbf{w}}$, where $\mathbf{s}$ is the signal vector and $\mathbf{w}$ the noise vector.

At this point, 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, 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 $SNR + N$ and a variance $4SNR + 2N$. In consequence, the expression of the ROC for the ED when N is too large can be easily obtained, as in [43]:

$$PD \approx Q(Q^{-1}(PFA) - SNRN) \tag{4.1}$$

where Q is the error function defined in Appendix (A.5).

The term $SNRN = SNR/\sqrt{2N}$ is defined as a normalized SNR, and, taking into account (4.1), 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 is proposed and will be analyzed in detail in the sections that follow.

4.1.2 Description of the multiple energy detector structure

This section describes the new structure formed by multiple energy detectors (MED). First, the distribution of the multiple EDs is detailed, followed by a description of the decision rule employed.

The MED consists of using several EDs, each of them using a different length of the observation vector. One could devise many strategies for partitioning the initial observation vector of dimension N, but in absence of any *a priori* information, L layers of partitions are considered to correspond to different segmentation degrees of the original observation length [89].

Figure 4.1 shows the proposed layered structure of the detectors, composed of successive halved divisions of the original observation vector. Layer $l = 0$ (top level), is formed by the original interval of N samples; while at layer $l = 1$, the original interval is divided in two, thereby obtaining two non-overlapped segments of $N/2$ samples each. Similarly, layer $l = 2$ is composed of four non-overlapped segments of $N/4$ samples each. This patterns continues successively until layer $l = L - 1$ (bottom level), where 2^{L-1} segments of $N/2^{L-1}$ samples are present. Hence, L represents the total number of layers in the structure.

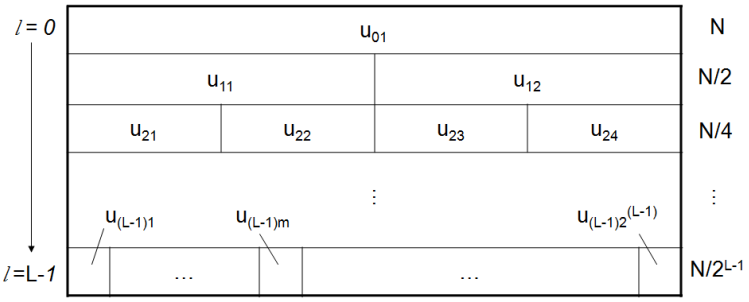


Figure 4.1: Layered MED structure of L levels and a total number of T detectors.

With the previously described structure, let us now introduce the new decision rule by defining $\mathbf{y}$ as the observation vector of the top level, whose number of samples is equal to N. Thus:

$$U = \sum_{i=1}^{T} u_i \begin{cases} U > 0 & \text{decides} \quad H_1 \\ U = 0 & \text{decides} \quad H_0 \end{cases} \tag{4.2}$$

where T is the total number of detectors (defined as $T = 2^L - 1$), and u_i is the decision of an ED applied to an observation vector $\mathbf{y}_i$ of dimension $N_i \leq N$, which itself is a segment of the original observation vector $\mathbf{y}$. In this case, the presence of a signal is determined if at least one of the energy detectors decides it. Each ED will be characterized by its PD_i and PFA_i values; depending on the signal energy in each segment, different situations may arise. If the signal energy in a particular segment is zero, then $PD_i = PFA_i$; if it is greater than zero, $PD_i > PFA_i$. In this latter scenario, the corresponding PD_i will depend on $SNRN_i$, independent of the

signal energy distribution across the segment. Therefore, it is independent whether the signal is distributed throughout the observation vector or, on the contrary, it is concentrated in a given area.

However, the analysis of this multiple energy detector structure is complicated by the fact that the individual decisions of each ED used in the structure are statistically dependent, due to the fact that segments are obtained by subdivision of other segments. This is a problem which is quite similar to the fusion of correlated decisions extensively considered in the framework of multi-sensor processing [88, 90]. 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.

4.2 Theoretical performance of the MED

The analysis of the multiple energy detector structure, presented in the previous section, is complicated by the fact that the individual decisions are statistically dependent as the segments are obtained from the successive subdivision of others. To evaluate the behavior of the MED and to be able to compare it with the ED, it is therefore necessary to calculate the probabilities of false alarm and detection in the new structure. For this purpose, let PFA_{MED} and PD_{MED} represent the respective probabilities of false alarms and detection in the MED.

4.2.1 Derivation of the PFA_{MED}

The first step consists of calculating the PFA_{MED}. Since there are no special reasons to do otherwise, the probability of false alarm of each detector ED is assumed to be fixed to the same value for all EDs present in the structure, $PFA_i = PFA$, $\forall i$. This implies that different thresholds are required for each different segment size N_i to maintain the same PFA. So, the threshold λ_i corresponding to the $i-$th energy detector is given by:

$$PFA_i = PFA = Q\left(\frac{\lambda_i - N_i}{\sqrt{2N_i}}\right) \tag{4.3}$$

where Q stands for the error function defined in Appendix (A.5).

Based on the structure of MED described in Figure 4.1, let us now introduce a double index notation for each energy detector. Thus, u_{lm} will refer to the decision corresponding to the m-th energy detector at layer l; where $l = 0, \ldots, L-1$; and $m = 1, \ldots, 2^l$. Taking into account (4.2), the following is yielded:

$$\begin{aligned} PFA_{MED} &= P(U > 0/H_0) = 1 - P(U = 0/H_0) \\ &= 1 - P(u_{lm} = 0, \forall l, m/H_0). \end{aligned} \tag{4.4}$$

Now, let us define the probability that all detectors at layer l decide H_0 (conditioned to H_0), and that all the detectors at lower levels have decided H_0:

$$P_l = P(u_{l1} = 0, u_{l2} = 0, \ldots, u_{l2^l} = 0/H_0; \\ u_{l'm'} = 0, l' > l, m' = 1, \ldots, 2^{l'}). \tag{4.5}$$

Then,

$$P(U = 0/H_0) = \prod_{l=L-1}^{0} P_l. \tag{4.6}$$

Considering the WGN model $\mathbf{w} : N(0, \sigma^2_{\mathbf{w}}\mathbf{I})$, and that all the segments belonging to the same layer are non-overlapped, the corresponding energies will be independent random variables. Therefore, the energy detectors of the same layer will be statistically independent and hence,

$$P_l = \prod_{m=1}^{2^l} P_{lm} \tag{4.7}$$

where P_{lm} is the probability that the m-th detector at layer l decides H_0, (conditioned to H_0 and to all the detectors at lower levels having decided H_0). Thus, this probability can be expressed as:

$$P_{lm} = P(u_{lm} = 0/H_0; u_{l'm'} = 0, l' > l, m' = 1, \ldots, 2^{l'}). \tag{4.8}$$

Now, we concentrate on evaluating P_{lm}. At the lowest level $(L-1)$ the following expression is reached:

$$P_{(L-1)m} = 1 - PFA, \quad \forall m. \tag{4.9}$$

And, for $l < L-1$, (explained in Appendix B):

$$P_{lm} = \frac{1 - PFA}{1 - Q\left(\sqrt{2} \cdot Q^{-1}(PFA)\right)}. \tag{4.10}$$

Therefore, going back to (4.7), the probability that all the detectors at layer l-th decide H_0, and that all detectors at lower levels have decided H_0, will be given by:

$$P_l = \begin{cases} \displaystyle\prod_{m=1}^{2^l} P_{lm} = \left(\frac{1 - PFA}{1 - Q\left(\sqrt{2} \cdot Q^{-1}(PFA)\right)}\right)^{2^l} & \text{if } \; l < L-1 \\ \displaystyle\prod_{m=1}^{2^{L-1}} P_{(L-1)m} = (1 - PFA)^{2^{L-1}} & \text{if } \; l = L-1 \end{cases} \tag{4.11}$$

So, combining (4.4) and (4.6) we may finally write the PFA_{MED} expression as follows:

$$\begin{aligned} PFA_{MED} &= 1-(1-PFA)^{2^{l-1}} \prod_{l=L-2}^{0} \left(\frac{1-PFA}{1-Q\left(\sqrt{2}\cdot Q^{-1}(PFA)\right)} \right)^{2^l} \\ &= 1-\frac{(1-PFA)^{2^L-1}}{(1-Q(\sqrt{2}\cdot Q^{-1}(PFA)))^{2^{L-1}-1}}. \end{aligned} \tag{4.12}$$

Analyzing in detail the expression obtained in (4.12), a number of interesting conclusions can be derived. On the one hand, the effect of the statistical dependence among the different detectors on PFA_{MED} can be observed. If the detectors were independent, the expression would be reduced to:

$$PFA_{MED} = 1-(1-PFA)^{2^L-1} \tag{4.13}$$

and hence, the denominator in (4.12) is due to the existence of dependence among some of the individual detectors. On the other hand, notice that the energy detector is a particular case of MED when only one layer in the structure ($L=1$) is considered. In this case, the expression is $PFA_{MED} = PFA$. It is also noteworthy that PFA_{MED} depends only on the PFA and the number of levels L, but not on N.

Finally, it is crucial to describe the procedure required to calculate the thresholds of the detectors in the different layers of the MED structure. First, we begin by fixing the desired PFA_{MED}, and then, taking into account (4.12) and the determined number L of layers, the PFA is calculated. After that, from the PFA and using (4.3) the thresholds λ_l is calculated for each layer $l = 0, \ldots, L-1$, taking into account that the detectors at the same level will have the same threshold and that the number of samples of the detector in each level is $N_l = N/2^l$, where N is the vector length at the top layer.

4.2.2 Derivation of the PD_{MED}

Following a similar procedure to that just described, we derive the PFA_{MED}. It is then possible to obtain the expression of the PD_{MED} for the entire set of detectors in the new structure. In this case, hypothesis H_1 is considered instead of H_0, and the signal presence is included where appropriate (described below).

In order to simplify the calculation of the general PD_{MED} expression for any signal duration, the particular case in which the signal length is equal to the segment duration of the lowest MED layer is first studied, and then, this result is generalized.

4.2.2.1 Particular case

Let us assume that the signal is completely included in the first segment in the structure's bottom layer ($l = L - 1$). In this situation, the signal is also entirely included in the first half of the first segment of the upper level ($L - 2$). This pattern continues on this way until layer $L = 0$, as seen in Figure 4.2 for the case of $L = 5$.

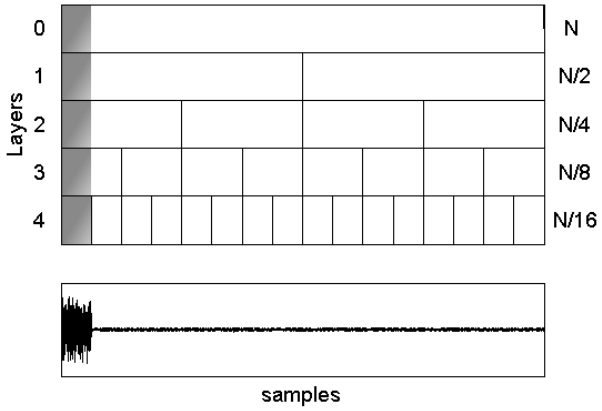


Figure 4.2: MED structure representation of $L = 4$ when the signal extends its energy uniformly throughout the first segment of the bottom layer (shaded).

Concentrating exclusively on the signal in the first segment does not imply any loss of generality in the calculation: the order in which multiple detector decisions are added in (4.2) is irrelevant.

Taking into account (4.4)-(4.6), it is possible to achieve similar expressions to calculate the PD_{MED}, as follows:

$$\begin{aligned} PD_{MED} &= P(U > 0/H_1) = 1 - P(U = 0/H_1) \\ &= 1 - P(u_{lm} = 0, \forall l, m/H_1), \\ P_l^1 &= P(u_{l1} = 0, u_{l2} = 0, \ldots, u_{l2^L} = 0/H_1; \\ &\quad u_{l'm'} = 0, l' > l, m' = 1, \ldots, 2^{l'}), \\ P(U = 0/H_1) &= \prod_{l=L-1}^{0} P_l^1, \\ P_l^1 &= \prod_{m=1}^{2^l} P_{lm}^1, \\ P_{lm}^1 &= P(u_{lm} = 0/H_1; u_{l'm'} = 0, l' > l, m' = 1, \ldots, 2^{l'}). \end{aligned} \tag{4.14}$$

Moreover, a derivation similar to that described in Appendix B (until the first equality of (B.11)) can be followed, substituting H_0 with H_1:

$$P_{lm}^1 = \frac{P(E_{lm} < \lambda_1/H_1)}{P(E_{lm} < 2\lambda_{l+1}/H_1)}. \tag{4.15}$$

Note that in those segments unaffected by the signal presence, the conditioning on H_0 is the same as the conditioning on H_1. Due to the fact that only the first segment of each layer contains the signal, the following expression can be derived:

$$P_{lm}^1 = P_{lm} = \frac{1 - PFA}{1 - Q\left(\sqrt{2} \cdot Q^{-1}(PFA)\right)}; \quad m = 2, \ldots, 2^l. \tag{4.16}$$

To calculate the P_{lm}^1 in the first segment, let us consider the approximate expression of (4.1) and the derivation found at the end of Appendix B between (B.12) and (B.16). It must be remembered that we are under hypothesis H_1; therefore, the following expression is valid under hypothesis H_0 (the energy follows a chi-square distribution):

$$\Delta_l = Q\left(\frac{2\lambda_{l+1} - N_l}{\sqrt{2N_l}}\right) \tag{4.17}$$

and it must be substituted by

$$\Delta_l^1 = Q\left(\frac{2\lambda_{l+1} - N_l - SNR}{\sqrt{4SNR + 2N_l}}\right) \tag{4.18}$$

where the energy follows a non-central chi-squared distribution with mean $N_l + SNR$ and variance $4SNR + 2N_l$. Then, for $l > L - 1$:

$$P_{l1}^1 = \frac{P(E_{l1} < \lambda_1/H_1)}{P(E_{l1} < 2\lambda_{l+1}/H_1)} = \frac{1 - Q(Q^{-1}(PFA) - SNRN_l)}{1 - Q(\sqrt{2} \cdot Q^{-1}(PFA) - SNRN_l)} \tag{4.19}$$

and, for $l = L - 1$:

$$P_{L-1,1}^1 = 1 - Q(Q^{-1}(PFA) - SNRN_{L-1}) \tag{4.20}$$

where $SNRN_l = SNR/\sqrt{2N_l}$, as defined in (4.1). Using (4.16) and (4.20), the PD_{MED} can be expressed as a simple modification of (4.12). So, changing the probabilities corresponding to the first segment at every level and keeping the remaining probabilities unchanged, it is possible to reach

$$\begin{aligned} PD_{MED} \;=\; & 1 - (1 - PFA)^{2^{L-1}-1}(1 - Q(Q^{-1}(PFA) - SNRN_{L-1})) \\ & \cdot \prod_{l=L-2}^{0} \left[\left(\frac{1 - PFA}{1 - Q(\sqrt{(}2) \cdot Q^{-1}(PFA))} \right)^{2^l - 1} \right. \\ & \left. \cdot \frac{1 - Q(Q^{-1}(PFA) - SNRN_l)}{1 - Q(\sqrt{2} \cdot Q^{-1}(PFA) - SNRN_l)} \right] \end{aligned} \tag{4.21}$$

where, adding terms, the following expression can be derived for $L > 1$:

$$PD_{MED} = 1 - \frac{(1-PFA)^{2^L-1-L}}{(1-Q(\sqrt{2}\cdot Q^{-1}(PFA)))^{2^{L-1}-L}} \cdot \frac{\prod_{l=L-1}^{0}[1-Q(Q^{-1}(PFA)-SNRN_l)]}{\prod_{l=L-2}^{0}[1-Q(\sqrt{2}\cdot Q^{-1}(PFA)-SNRN_l)]}. \quad (4.22)$$

By analyzing in greater detail (4.22), the expression of PD_{MED} can be obtained for a single level by replacing $L = 1$. Taking into account that the denominator of the second term is 1 (since it is defined only for $L > 1$), the following expression can be reached:

$$PD_{MED} = Q(Q^{-1}(PFA) - SNRN_0). \quad (4.23)$$

Notice that this expression concurs with (4.1), where the PD for a single detector has been defined. In the same way, it is also essential to notice that PD_{MED} depends on not only the PFA, but also on L and $SNRN_0$, the latter defined as the normalized signal-to-noise ratio at the top level ($L = 0$). The normalized signal-to-noise ratio for a particular level of the structure is thus expressed as follows:

$$SNRN_l = SNRN_0 \cdot \sqrt{2^l} \quad (4.24)$$

considering that $N_l = N/2^l$, where N is the number of observation samples at the top level.

4.2.2.2 Generalization of the PD_{MED}

To derive the generalization of the PD_{MED} expression viewed in (4.22) for any signal length, it is necessary to modify the expression of the PFA_{MED} viewed in (4.12) in a manner similar to the preceding case. Therefore, the probabilities corresponding to the segments affected by the presence of the signal have to be changed at every layer and the remaining probabilities

should stay unchanged. This brings us to the general equation:

$$\begin{aligned} PD_{MED} &= 1-(1-PFA)^{2^{L-1}-k(L-1)} \\ &\cdot \prod_{m=1}^{k(L-1)} (1-Q(Q^{-1}(PFA)-SNRN_{L-1m})) \\ &\cdot \prod_{l=L-2}^{0} \left[\left(\frac{1-PFA}{1-Q(\sqrt{(2)}\cdot Q^{-1}(PFA))} \right)^{2^l-k(l)} \right. \\ &\cdot \left. \prod_{m=1}^{k(l)} \left(\frac{1-Q(Q^{-1}(PFA)-SNRN_{lm})}{1-Q(\sqrt{2}\cdot Q^{-1}(PFA)-SNRN_{lm})} \right) \right]. \end{aligned} \tag{4.25}$$

And, by grouping terms:

$$\begin{aligned} PD_{MED} &= 1-\frac{(1-PFA)^{2^L-1-K}}{(1-Q(\sqrt{2}\cdot Q^{-1}(PFA)))^{2^{L-1}-1-K+k(L-1)}} \\ &\cdot \frac{\prod_{l=L-1}^{0}\prod_{m=1}^{k(l)}[1-Q(Q^{-1}(PFA)-SNRN_{lm})]}{\prod_{l=L-2}^{0}\prod_{m=1}^{k(l)}[1-Q(\sqrt{2}\cdot Q^{-1}(PFA)-SNRN_{lm})]} \end{aligned} \tag{4.26}$$

where the following terms have been introduced:

- $k(l)$: number of affected segments at layer l where a signal is present.
- $K_M = \sum_{l=0}^{L-1} k(l)$: total number of affected segments with a signal presence in the MED.
- $SNRN_{lm}$: normalized signal-to-noise ratio of the m-th affected segment at layer l.

In order to better understand the simplification of these variables, let us consider an example where a MED structure of 5 layers is used to detect a signal concentrated in the first segment of the third layer, (Figure 4.3). In this case, it is possible to obtain the aforementioned variables observing Figure 4.4, where the detectors affected by the presence of signal at each layer have been colored. Therefore, the variable $k(l)$ is a vector containing the following values $[1, 1, 1, 2, 4]$, and K_M is equal to 9.

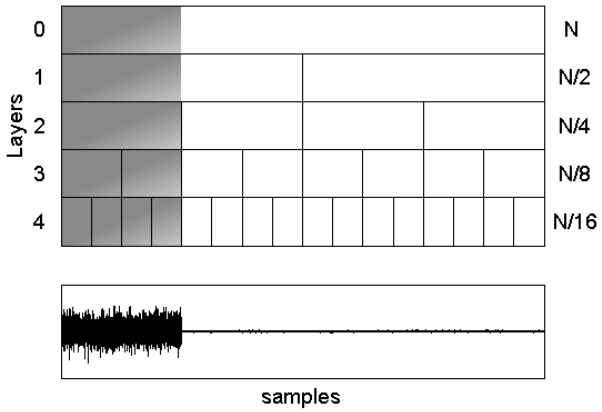


Figure 4.3: MED structure of $L = 4$ when the signal extends its energy uniformly throughout the first segment of the third layer (shaded).

Considering the process as a whole until (4.26), the importance that (4.12) plays throughout must be highlighted. This is not surprising, however, considering the derivation of the PD_{MED} is closely related to the PFA_{MED}. This can be attributed to the fact that the model of energy propagation from lower to upper layers, described in (B.2), is independent of the conditioning on H_0 or H_1. Following this model, if a segment is not affected by the

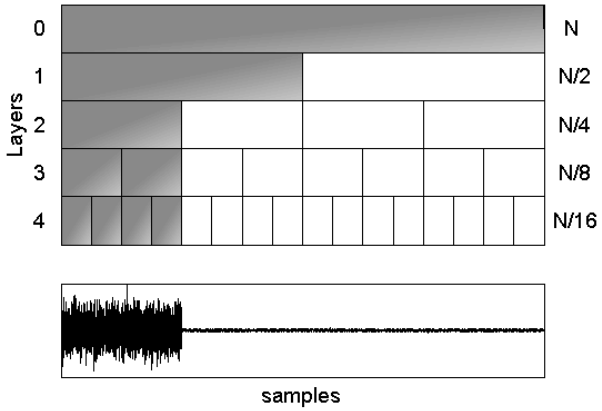


Figure 4.4: Detectors affected at each layer by the signal presence in a MED structure of $L = 4$. The signal is uniformly distributed throughout the first segment of the third layer.

presence of the signal, all the segments from lower layers, which are subsets of it, will not be affected either. Hence, the derivation of the conditioned probabilities for non-affected segments is independent of the signal presence and coincides with that obtained in the derivation (4.12). On the other hand, in segments with signal, the derivation of the probabilities is the same as in the previous case. But consideration must be made during the final steps for the effect introduced by the $SNRN_{lm}$ in the calculation of such conditional probabilities.

4.3 MED theoretical evaluation

As the behavior of the MED and its underlying theoretical principles have been described, it is essential to validate its interest (as opposed to the use of classic ED) when the signal has different lengths (compared to the duration of the observation segments) in different layers of the MED. Assuming that the length of the upper segment of the MED is fixed by a number of computational limitations, it will be demonstrated that the detector which gives the best results is the one that is adapted to the signal duration. Thus, it is necessary to verify the improvements achieved when the signal being detected has a short duration, and when multiple detectors (MED) with an observation segment length comparable to the signal duration are used instead of a single ED with a fixed number of observations samples. In addition, it is also necessary to evaluate the possible degradation that is introduced by using subdivisions of the original detector which are not required when the signal duration is similar to the duration of the original observation interval. To this end, three examples are considered below to bring forth the advantages and disadvantages of the MED.

4.3.1 The signal duration is comparable to the observation vector length at the top layer

First of all, the least favorable case for the MED (in comparison with ED) is studied. This case occurs when the signal energy extends uniformly across the whole initial observation vector, as seen in the Figure 4.5. This implies that all detectors with shorter duration segments will corrupt the detection. In this situation, all segments in the structure are affected by the signal. With this in mind, and taking into account (4.26), the following expressions can be derived:

- $k(l) = 2^l$;
- $K_M = 2^L - 1$;
- $SNRN_{lm} = \dfrac{SNRN_l}{2^l} = \dfrac{SNRN_0 \cdot \sqrt{2^l}}{2^l} = \dfrac{SNRN_0}{\sqrt{2^l}}$.

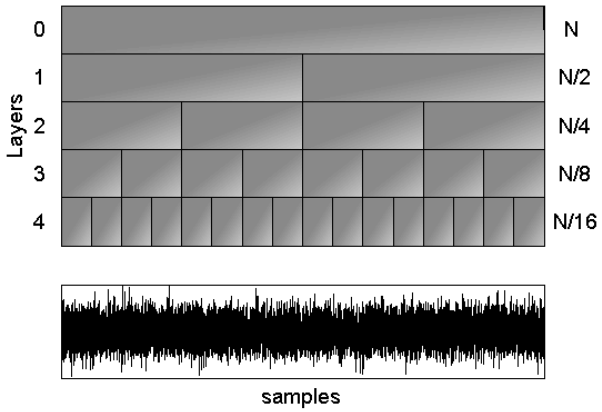


Figure 4.5: Detectors affected at each layer by the signal presence in a MED structure of $L = 4$. The signal is uniformly distributed throughout the top layer.

By substituting these values in the expression of PD_{MED} in (4.26), we meet the next case:

$$PD_{MED} = 1 - \frac{\prod_{l=L-1}^{0} [1 - Q(Q^{-1}(PFA) - SNRN_0/\sqrt{2^l})]^{2^l}}{\prod_{l=L-2}^{0} [1 - Q(\sqrt{2} \cdot Q^{-1}(PFA) - SNRN_0/\sqrt{2^l})]^{2^l}}. \quad (4.27)$$

The expression of the PFA_{MED} is given by (4.12), as it is not affected by the particular signal properties and it only depends on the noise characteristics.

In Figure 4.6, the ROC curves represent the theoretical expressions of PFA_{MED} and PD_{MED} shown in (4.12) and (4.27), respectively. In the graph, the behavior of the MED using several layers L and with $SNRN_o = 2$ become noticeable. This value was selected to ensure good detectability for $L = 1$, which corresponds to the ED. As expected, some degradation in the behavior of the MED curves is seen when using several layers of the structure ($L > 1$) compared with the ED ($L = 1$). This is due to using unnecessary subdivisions of the original observation segment, whose length is similar to the signal. However, this deterioration in the MED performance is not very significant considering that, for example, a value of $L = 7$ entails subdivisions up to a factor of 64 of the original segment and a total number of 127 energy detectors in the MED layered structure.

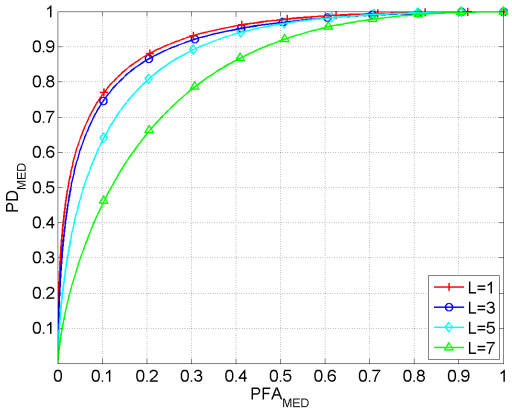


Figure 4.6: Theoretical ROC curves of the MED for $SNRN_o = 2$ and for different number of layers used (L). The signal energy is concentrated in the first segment of the top layer.

In this example, the "worst-case" scenario in which the use of the MED can damage the detection results has been studied. However, in the following examples it will be shown how it is possible to justify this behavior because of the improvements achieved when the signal concentrates its energy in segments of lower levels.

4.3.2 The signal duration is comparable to the observation vector length at the bottom layer

The second example corresponds to the opposite case of that found in 4.3.1. Here, the signal concentrates its energy in the first segment of the bottom layer of the MED, as shown in Figure 4.2. Therefore, we are searching for the most favorable case wherein the MED would clearly be a better choice in terms of performance than the classical ED.

In this situation, the expressions of PD_{MED} and PFA_{MED} shown in (4.22) and (4.12), respectively, can be used directly to obtain the corresponding ROC curves. In Figure 4.7, the ROC curves of the MED have been represented for different values of L and for $SNRN_o = 0.5$. This particular value was selected because of the interest in studying the ED ($L = 1$) in the worst case (i.e. when it behaves nearly identically to random detector $PD_{MED} \cong PFA_{MED}$), thereby allowing one to observe the performance of the MED in this situation. As expected, the degree of improvement achieved with the MED increases with the number of layers required to match the actual signal duration.

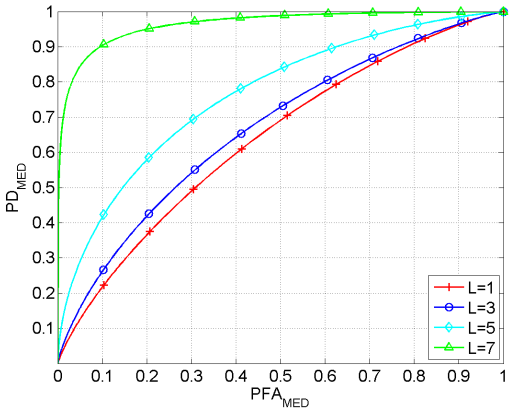


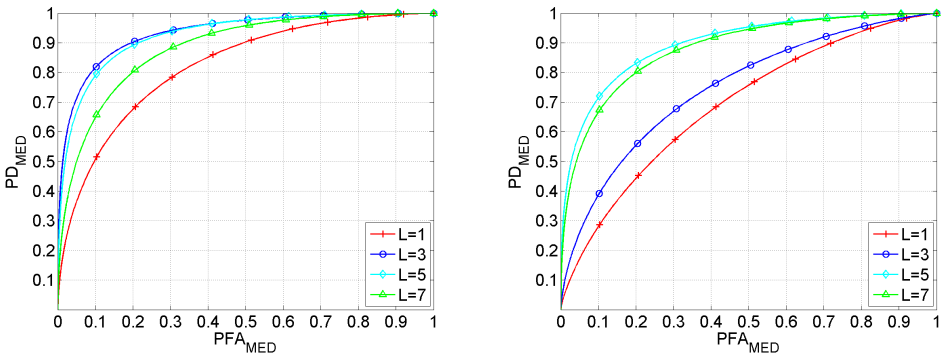
Figure 4.7: Theoretical ROC curves of the MED for $SNRN_o = 0.5$ and for different number of layers used (L). The signal energy is concentrated in the first segment of the bottom layer.

4.3.3 The signal duration is comparable to the observation vector length at intermediate layers

Finally, two intermediate cases are compared in this scenario with the preceding examples: in particular, those cases where the signal energy is uniformly distributed across the first segment of the third and fifth level, representing the 25% and the 6.25% of the whole length of the original observation vector, respectively.

Figure 4.8a shows the theoretical ROC curves of the MED using different values of L when the signal has a duration similar to the segment length of layer 3. The normalized signal-to-noise ratio was set to $SNRN_o = 1.3$ because this particular value offered a good detectability for the case of $L = 3$. As expected, the best result was reached when only the first three levels of MED were used. However, notice that when using more levels the resultant curves do not outperform those of the optimum; nevertheless, it is possible to reach a performance enhancement in comparison with the classic ED ($L = 1$). For example, for $L = 5$ (segments with duration 1/4 of the actual signal duration were used at the bottom layer), a nearly identical performance to that reached by the best case of $L = 3$ was achieved.

In Figure 4.8b, the signal had a similar duration to the observation segment of the fifth level and the $SNRN_0$ was set to 0.7. The best ROC curve was obtained for $L = 5$, as expected. As in the previous example, the improvement attained (relative to the ED) by increasing the number of levels was quite significant.



(a) (b)

Figure 4.8: Theoretical ROC curves of the MED for different numbers of layers used (L). (a) Signal duration is equal to the segment length of layer 3 and $SNRN_o = 1.3$. (b) Signal duration is equal to the segment length of layer 5 and $SNRN_o = 0.7$.

4.4 General comments on the MED

The previous section analyzed in detail the performance of the MED in different situations and examples. It is now possible to extract a general conclusion about its behavior. In view of the results, using the MED (versus the classic ED) reports a significant improvement in specific situations. In those cases where the MED uses an overestimated number of required layers, a deterioration is produced. However, the improvement achieved is always better than the possible degradation suffered. This fact can be observed in Table 4.1, where two opposing examples are examined for $SNRN_o = 0.7$, using a MED structure with a total number of 7 layers:

- *Example A*: The signal is mainly concentrated in the segment of the top layer.
- *Example B*: The signal duration is comparable to the segment duration of the bottom layer.

This table shows the increase in PD_{MED} obtained for both situations and three different PFA_{MED} values. The results are derived by subtracting the PD_{MED} achieved using all layers ($L = 7$), along with that obtained for only one layer ($L = 1$). Thus, a negative value indicates MED degradation when using all layers and, on the contrary, a positive value corresponds to an enhancement in performance. As expected, one can clearly appreciate how improvement is always greater than the resultant deterioration. Or more informally, we can affirm that using the MED offers "*a lot to gain and little to lose.*"

signal duration	$\boldsymbol{PFA_{MED}}$		
	10^{-1}	10^{-2}	10^{-3}
Example A	-0.12	-0.034	-0.0069
Example B	0.71	0.92	0.93

Table 4.1: PD_{MED} increases in the MED structure for different PFA_{MED} values comparing the case of using $L = 7$ with respect to $L = 1$. Two examples of different signal lengths are considered for the same $SNRN_0 = 0.7$.

More specific conclusions about the behavior of the MED in other situations are hard to obtain unless prior information about the signal duration and distribution energy across the observation vector is available. However, due to the scope of application of such structures where the signals to be detected are not completely known, it is difficult to have any prior information about the signal. Therefore, any assumption in this direction would be unrealistic and it would reduce the generic character of the MED. Thus, the aspect of greatest interest is that the theoretical expressions of PD_{MED} and PFA_{MED}, described in (4.12) and (4.26), can be applied to the analysis of every particular scenario. Hence, it is necessary to consider the general validity of these two equations, taking into account different aspects, as discussed below.

4.4.1 Segmentation techniques

First, it is necessary to emphasize the relationship between the PFA_{MED} and PD_{MED} expressions with the technique used to segment the original observation vector. In this case, the expressions (4.12) and (4.26) are linked to the segmentation approach described in Figure 4.1, where the observation segment in the top layer is successively divided by two non-overlapping segments. This detector structure has been used since, in principle, the information signal is completely unknown. However, in those cases where some *a priori* information about the signal to be detected was available, it would be possible to use other partition strategies that are more optimal than and different from the aforementioned. In this case, it would be necessary to conduct a new analysis of the resulting detector structure to derive the new expressions of the PFA_{MED} and the PD_{MED}. Hence, it can be concluded that the optimal structure and their distribution are closely related to the characteristics of the signal to be detected.

Thus, for a given application, a more appropriate structure of detectors could be devised by knowing *a priori* the maximum (N_{max}) and the minimum (N_{min}) durations of the expected signals. Then, it seems reasonable to choose the shortest segment duration N_{min} at the bottom layer of the structure of

detectors, while at the top layer an observation segment will be selected whose duration will be given by the following expression:

$$R(N_{max}) = N_{min} \cdot 2^{L-1} \tag{4.28}$$

where $R(x)$ indicates upper round-off to the closest power of two. So, it is possible to determine the optimum number of layers that the MED will have from the following condition:

$$L = 1 + \frac{ln\left(\frac{R(N_{max})}{N_{min}}\right)}{ln2}. \tag{4.29}$$

In any case, as discussed in the beginning of this section, it is possible to find other situations and particular applications in which some other information about the signal detector could be available *a priori*. These would be cases that need to be addressed specifically to find the optimal structure of the MED.

4.4.2 Signal and noise models

Next, let us consider the validity of (4.12) and (4.26) with respect to the different noise and signal models. First of all, the expression of PFA_{MED}, shown in (4.12), is valid with the only constrain that the threshold at every individual detector is computed by using (4.3). This means that the background noise should be white Gaussian, although extension to non-independent non-Gaussian noise could be attempted by using the extended energy detector proposed in Chapter 3. Similarly, it would also be possible to expand the use of the multiple detector structure to include the scenario when the background noise is non-stationary. In this case, it is necessary to employ a method to estimate the variance of the noise and normalize the energy of the individual detector (ED) of the MED.

Secondly, the signal model used does not influence the behavior of the PFA_{MED}; nevertheless, it could, in principle, affect the validity of the PD_{MED} expression viewed in (4.26). However, any distribution of the signal energy across the observation interval may be considered. Therefore, signal energy is not required to be uniformly distributed across one or more segments at every layer. This is because the parameters including the presence of the signal are the individual normalized signal-to-noise ratios $SNRN_{lm}$. Thus, very irregular distributions of the signal energy can be tested using (4.26). In the selected examples of the preceding sections, some particular cases were looked at in which the propagation of $SNRN_{lm}$ through the different layers adapted some simple forms. A myriad of possibilities exists for the distribution of the signal energy, but in the absence of prior information the selected examples are enough to draw a reliable conclusion when evaluating the MED in comparison to the ED.

4.4.3 Real-time applications

Finally, the proposed detector can be applied in a real-time framework, as any other detector, by considering streams of successive observation intervals. In every incoming observation interval, the partition of Figure 4.1 would be applied. If the successive observation intervals are non-overlapped, on average one false alarm is obtained for every $1/PFA_{MED}$ observation interval; a correct decision is made on average every $1/PD_{MED}$ observation interval where a signal is present. However, if the successive observation intervals are overlapped (for example, in the extreme case of only one sample changing from one interval to the next), a more complex derivation is required to determine the average number of false alarms or correct detections due to the dependence among successive decisions. Regardless, the PFA_{MED} and PD_{MED} are always the key parameters in determining the performance of the detector.

4.5 Performance of the MED in the time and frequency domains

The purpose of this section is to validate with simulations the theoretical expressions obtained for the MED, and then to compare the results with the classical ED. In particular, the performance of the MED is evaluated, not only in the time but also in the frequency domain, since detection of signals with unknown bandwidth is also of particular interest.

4.5.1 Frequency MED structure

The signals to be detected are completely unpredictable and the spectrum can also provide additional information. Some recent studies are devoted to the evaluation of the detection performance in the frequency domain. In [13], an optimal detector is derived in the discrete frequency domain and the paper discussed the procedure for selecting the segment length as part of the spectral estimation method. In this case, the knowledge of the signal bandwidth is required. In some other applications (e.g, echo detection in radar, sonar or acoustics) an algorithm is applied to the automatic target detection of acoustic signals [16]. Again, an approximate knowledge of the signal bandwidth is required. However, the detection problem in the frequency domain can be attempted by using the MED structure in order to detect signals of different bandwidth. In this latter case, it is possible to apply the same methodology previously described to derive a similar MED structure. In consequence, both problems can be considered as conceptually equivalent, and the frequency detection problem of the MED is addressed by transforming the original observation vector $\mathbf{y}$ of dimension N (at layer 0),

into the frequency domain $\mathbf{Y}$ by using the discrete Fourier transform (DFT) as follows:

$$\mathbf{Y}[k] = \sum_{n=0}^{N-1} \mathbf{y}[n] e^{-j2\pi kn/N}. \tag{4.30}$$

The DFT length was set to the same dimensions as the original observation vector at the top layer (N), and the normalized angular frequency was given by $\Omega = 2\pi k/N$. The same subdivision strategy as the one used in the time domain was then applied, and the detection problem at each individual segment was addressed using the classical ED and by subsequently computing the test statistic in a similar manner. Considering $\mathbf{Y}_{ml}$ as the observation vector corresponding to the *m-th* frequency segment at layer l, it is possible to rewrite the hypothesis test as follows:

$$\begin{array}{ll} H_0 : \mathbf{Y}_{lm} = \mathbf{W}_{lm} & \mathbf{W}_{lm} : N(0, \sigma^2_{\mathbf{W}}\mathbf{I}) \\ H_1 : \mathbf{Y}_{lm} = \mathbf{S}_{lm} + \mathbf{W}_{lm} & \end{array} \tag{4.31}$$

where $\mathbf{W}_{lm}$ is the noise vector with zero-mean white Gaussian distribution and $\mathbf{S}_{lm}$ is the signal vector. Both are set to the *m-th* frequency slot, with $m = 2^l$ equivalent to the number of subdivisions (or EDs) at layer l ($l = 1, \ldots, L$). In hypothesis H_0, the received signal in the *m-th* frequency slot is assumed to contain only noise; if H_1 is true, noise and signal are present.

In the following section, the performance of the MED structure in the time and in the frequency domain will be studied, and results from both will be compared with the classical ED. To do so, several experiments were performed by considering the simple detection problem in which a Gaussian signal ($\mathbf{s}$) of different time durations had to be detected in the presence of white Gaussian noise ($\mathbf{w}$) with zero-mean. The variance of the noise was fixed to unity, while different normalized signal-to-noise ratios ($SNRNs$) were generated by properly scaling the Gaussian amplitude of the signal. A MED structure of 7 layers ($L = 7$) was implemented and compared with the classical ED ($L = 1$). The original observation vector size N was set to 16384, leading to an observation segment on the bottom layer of 256 samples. The ROC curves, determined by the PFA_{MED} and the PD_{MED}, were utilized to summarize the robustness of the MED detector. The results obtained were averaged over $40 \cdot 10^3$ independent trials.

4.5.2 Time ROC curves

This section presents the results reached for the MED in the time domain and in different situations. Figure 4.9 shows the experimental (dashed) and theoretical (solid) ROC curves when using the MED to detect a Gaussian signal with several time durations in the presence of additive WGN. Different numbers of layers (L) were employed, but only the ROC curves of the most representative ones are depicted. Notice that when $L = 1$, the

MED structure corresponds to the classical ED since only one detector of the original observation length N was being used to detect the signal.

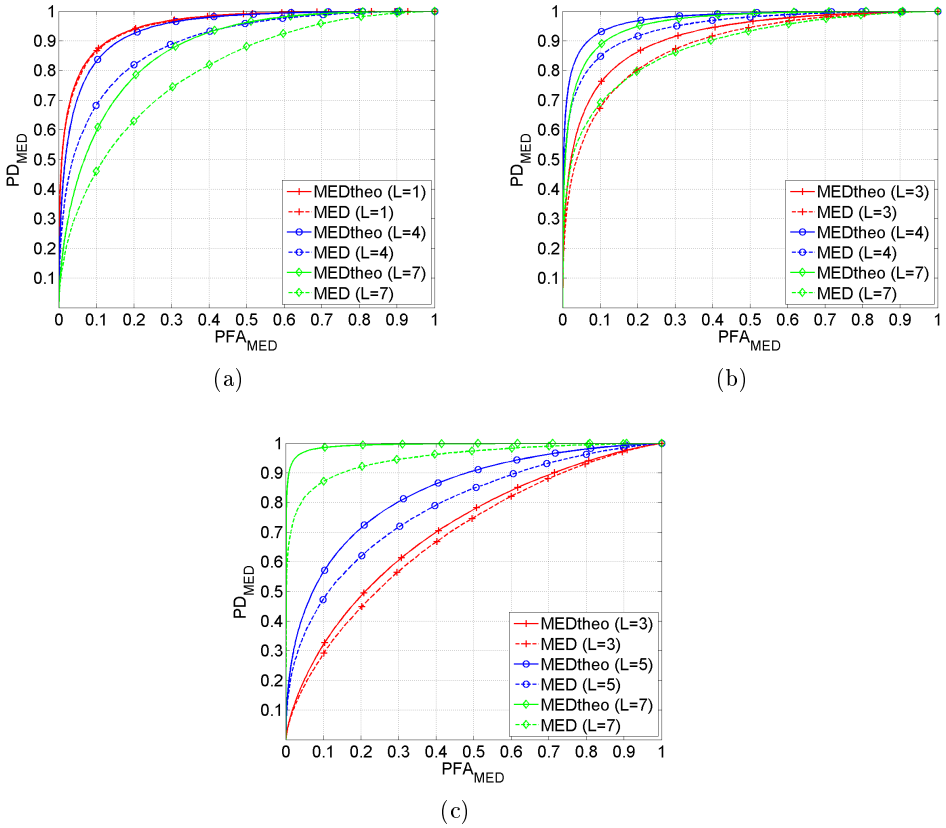


Figure 4.9: Theoretical and simulated ROC of time MED when detecting a Gaussian signal uniformly distributed throughout the first segment of different layers and in the presence of white Gaussian noise. (a) Signal in segment of $L = 1$ and $SNRN_o = 2.4$. (b) Signal in segment of $L = 4$ and $SNRN_o = 1.2$. (c) Signal in segment of $L = 7$ and $SNRN_o = 0.6$.

In Figure 4.9a, the signal had a time duration similar to that of the original observation vector (N), and in Figures 4.9b and 4.9c, the signal energy was uniformly distributed through the first segment of layers 4 and 7, respectively. Several $SNRN_0$ were simulated, but only the most suitable ones were selected from each case to show the MED performance results. In all cases, the simulation results agreed quite closely with the theoretical derivations. The reason that all experimental curves are below the theoretical curves lies in that the chi-square distribution in (4.1) was approximated to a Gaussian distribution for large N. This fact can be better appreciated for large values of L, which results in a smaller number of samples in the

observation vector. Apart from this point, the three examples reveal that the best results are obtained when using a number of layers that includes an ED with an observation length comparable to the signal duration. Therefore, the optimum number of layers is $L = 1$ in the first case, and $L = 4$ and $L = 7$ in the other two cases.

Let us now compare the two opposite examples shown in Figures 4.9a and 4.9c. It is important to highlight the fact that in the first case the degradation suffered when the MED used unnecessary layers ($L > 1$), in comparison to the optimum result ($L = 1$), is not as significant as that experimented in the second case, where the layers are necessary and are not used. Therefore, this fact shows the superiority of the proposed structure and demonstrates the general applicability to a wide variety of areas, as will be seen in the following chapters.

4.5.3 Frequency ROC curves

The following analysis is based on the evaluation of the MED frequency structure when detecting a Gaussian signal immersed in additive WGN. Figure 4.10 depicts the same three time examples studied in the previous section, but in this case, the time signal of the original observation vector at the top layer was transformed beforehand in the frequency domain (using the DFT algorithm) and then partitioned into different segments. It can be observed how the best performance in detection is achieved for $L = 1$ (classical ED) in all cases. Furthermore, increasing the number of layers used in the MED leads to poorer behavior in the resultant ROC curves. This is due to the flat spectrum of the simulated Gaussian signal used: its frequency transformation has a uniform energy distribution throughout the original observation vector, independent of the signal's time duration. In this case, the MED frequency structure does not provide any valuable information.

Experiments similar to the preceding examples were performed; but this time three Gaussian signals with different bandwidth were used. The signal time durations were the same as those used in Section 4.5.2. In order to simulate these situations, a white Gaussian signal with a fixed time duration (comparable to the segment length of layer 1, with the dimension $N = 16384$) was filtered by applying a low pass filter. By using different frequency stop parameters, it was possible to generate a signal spectrum with different bandwidths. Considering a sample frequency (f_s) of 24 kHz and using a DFT length equal to the observation vector (N), the following cut-off frequencies were set: $f_s/2$, $f_s/4$ and $f_s/7$. This led to a filtered signal with a bandwidth whose energy was principally concentrated in the first segments of layers 1, 3 and 6, respectively. The resulting ROC curves for each case are presented in Figure 4.11. The simulation results show how the detector performs better in each case with a different value of L. As expected, the optimum number of layers used corresponded to the layer containing a de-

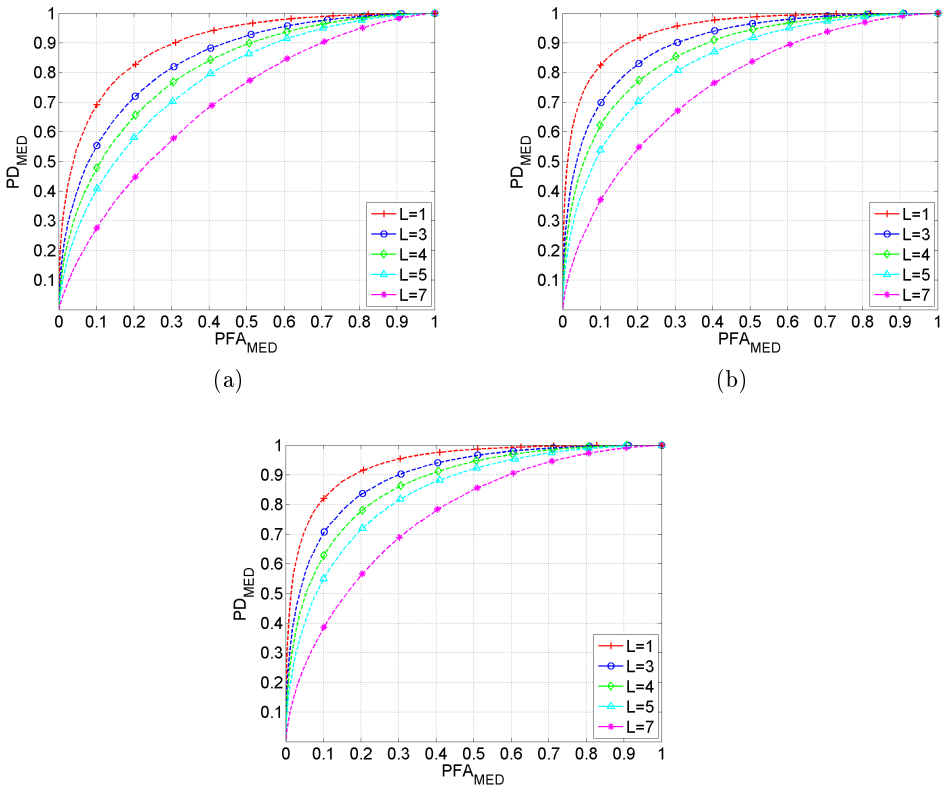


Figure 4.10: Simulated ROC of frequency MED when detecting a Gaussian signal uniformly distributed throughout the first segment of different layers and in the presence of white Gaussian noise. (a) Signal in segment of $L = 1$ and $SNRN_o = 2.4$. (b) Signal in segment of $L = 4$ and $SNRN_o = 3$. (c) Signal in segment of $L = 7$ and $SNRN_o = 3$.

tection segment with a length similar to that of the signal bandwidth. Once more, again it is better to use as many layers as possible: the improvement witnessed is always worthwhile in comparison to the possible deterioration experimented when the number of layers is overestimated. These results also demonstrate the necessity of using the frequency information for detection. As observed, by using the frequency MED, valuable information about the frequency characteristics of the signal to be detected is extracted. Furthermore, this information can be properly combined with the time MED to better exploit the characteristics of the signals. The combined results of time and frequency can offer more reliable detection decisions. This aspect remains a challenging task which should be considered as a future line of research. Moreover, it is also possible to work with the information provided

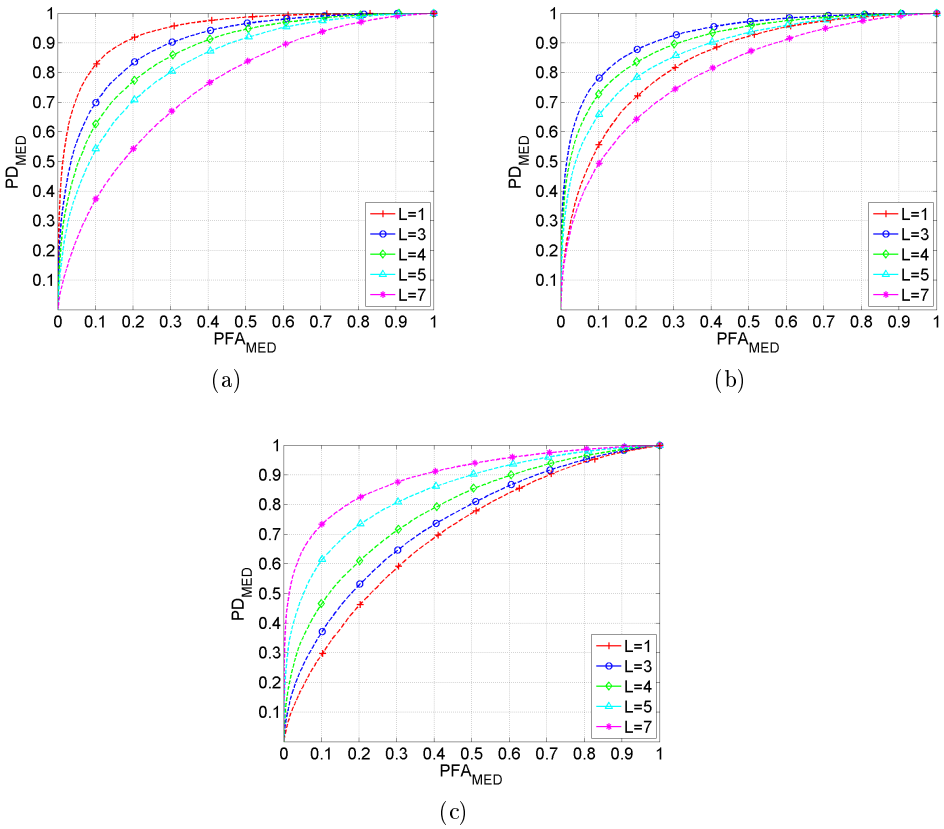


Figure 4.11: Simulated ROC of frequency MED when detecting a Gaussian signal which bandwidth is mainly distributed throughout the first segment of different layers, and in the presence of white Gaussian noise. (a) Signal bandwidth in segment of $L = 1$ and $SNRN_o = 3$. (b) Signal bandwidth in segment of $L = 3$ and $SNRN_o = 1.8$. (c) Signal bandwidth in segment of $L = 6$ and $SNRN_o = 0.8$.

by each MED structure, then to implement an enhanced recognition process, as shown in Chapter 6.

4.6 Conclusions

This chapter has addressed the problem of detecting unknown signals with undetermined duration. It has been demonstrated that the signal duration incorporates an important uncertainty that influences the detection problem. Therefore, a novel approach based on the implementation of a MED structure was proposed. The MED was formed by multiple EDs matched to different signal durations, and the segmentation strategy followed in this case

was based on several layers. Each layer was formed by successive segmentations (of a factor of 2) of the original observation vector. The individual decisions of the EDs were statistically dependent and therefore, a detailed study and derivation of the new PFA and PD was conducted. After that, the theoretical expressions were evaluated for different signal time durations. The results showed the improvement realized with the MED in comparison to the single ED when the signal duration was smaller than the original observation vector. It was also demonstrated that when the MED uses an overestimated number of required layers, deterioration arose. However, this deterioration was acceptable since it was always smaller than that experimented by the single ED in the previous case. Extensions of the MED to the frequency domain were also considered. Both time and frequency structures were evaluated by performing several simulations with signals of different durations and in the presence of independent Gaussian noise. Again, the results showed that the best performance was always achieved by the MED when the number of layers used corresponded to the layer containing the detection segment that was comparable to the signal length.

Part III

Acoustic applications

Chapter 5

Acoustic event detection

If you are not prepared to be wrong, you will never come up with anything original.

Ken Robinson

The focus of this chapter is to present a detailed evaluation of the different detectors, presented in Chapters 3 and 4, with real acoustic signals. Several acoustic sources and a variety of real and simulated noise scenarios were tested. In particular, the ED is compared with a new generalization that rejects false alarms induced by the difficulty of estimating the noise statistics in non-stationary conditions. Furthermore, the EED and the PEED are evaluated in real-world scenarios to deal with non-Gaussian and non-independent noise. In addition, as no information about the duration of the acoustic event is available, a MED structure matched to different time durations is used in the time and frequency domains. This evaluation is conducted by measuring the performance of the different energy detectors in terms of the ROC.

5.1 Detection in real acoustic applications

There are a great number of areas in which the detection of unknown events is required. One one of the most interesting fields of research is acoustic scene analysis, devoted to surveillance applications in which the signals recorded by a set of microphones are processed to extract as much information as possible about the environment [93]. Moreover, under some particularly adverse situations, such as hidden objects or low lighting scenarios, it is possible to use acoustic sensors to capture information not perceived by video sensors [91].

In real acoustic applications, sound sources are not entirely known and thus, the design of an appropriate detector is more difficult. In this case, energy detection is of interest since an ED is optimal for the automatic detection of unknown signals in the presence of independent and Gaussian noise. However, non-Gaussian, non-independent and non-stationary acoustic background noises are expected to be present in real-world scenarios. Therefore, alternative EDs are needed which are capable of continuously adapting to the noise statistics and of calculating the appropriate threshold for a required PFA.

Much of the recent work in this research area does not take into account the background noise and its particular characteristics, nor of the time or frequency duration of the event. As shown in Chapters 3 and 4, these facts normally decrease the performance of the detection; the noise samples can be highly correlated and they do not necessarily follow a Gaussian distribution. Consequently, it is necessary to apply 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. In previous chapters, the different extensions of the ED were evaluated with simulated signals and noises; however, now it is necessary to extend this study to the case of real-world signals and noises.

5.1.1 Surveillance applications

Part of the present work was developed within the framework of a collaborative research project between various companies and Spanish universities called HESPERIA (in English: Homeland Security: Technologies for the Security in Public Spaces and Infrastructures) [79]. 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)" of the "Universitat Politècnica de València (UPV)" 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. Some results related to this line of research can be found in [30, 62].

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 chapters 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 5.1. Furthermore, the audio techniques employed will be enhanced to achieve the necessary level of autonomy and functionality that the general system requires by generating the required presentation contents.

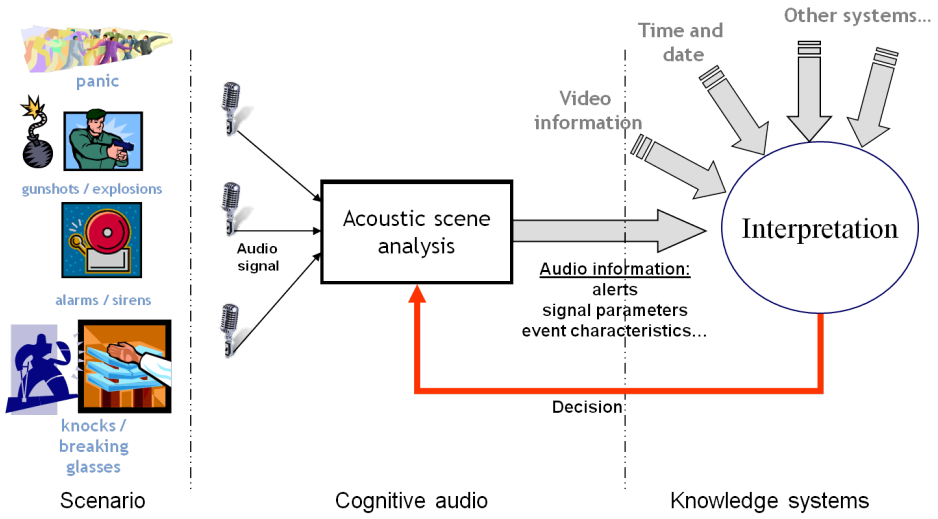


Figure 5.1: General description of the surveillance system.

5.1.2 Equipment and scenarios

One of the goals of this work was to carry out a measurement campaign of acoustic signals in various areas where experimental recordings of possible background noises and events were available. This section describes the different types of recordings (acoustic database) and the acoustic signals used to evaluate the potential performance of the different energy detectors and extensions described in previous chapters. On the one hand, different types of background noises that can be found in the scenarios considered were recorded and, on the other hand, a subset of possible acoustic signals present in these environments was also generated. For that purpose, a measurement campaign of acoustic sounds was carried out in different scenarios.

Prior to commencing, the preliminary selection of the equipment was of

particular importance. Some common limiting factors of this type of acoustic systems needed to be taken into account:

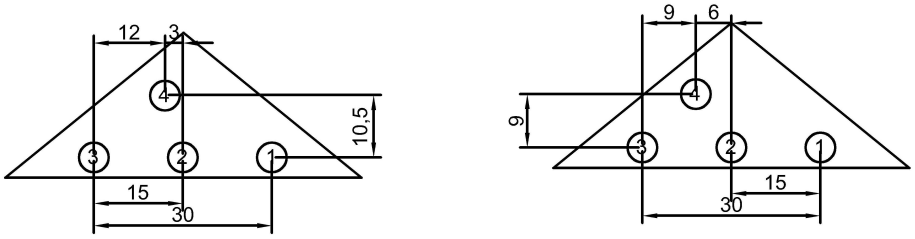
- The decrease in signal-to-noise ratio when increasing the distance between the acoustic event and the microphone.
- The background noise changes depending on the time of the day, location of the sensors and other factors. These changes can affect both the level and the statistics of the signal, and therefore it is necessary that the system continuously adapts to the background noise.
- The waveform of the detected sound is not known, nor its duration, frequency, intensity, etc. These parameters vary depending on many factors that must necessarily be considered: the repetition of the sound, the distance, and the impulse response between source and sensor.

(a) (b)

Figure 5.2: Recording equipment. (a) Microphone. (b) Data acquisition card.

Accounting for the aforementioned limitations, the acoustic events and background noises were recorded using a set of 8 omni-directional electret condenser microphones similar to that one shown in Figure 5.2a. The frequency response of the microphones was between 25 Hz and 20 kHz, which was a sufficient range to cover all the frequency requirements of sounds and noises recorded. In addition, a multichannel audio data acquisition unit (8 channels) with a sampling frequency of 48 kHz was used and is shown in Figure 5.2b. The distribution of the microphones used consisted of two arrays of four microphones each, with 109 cm of separation between each array. Each microphone array was roughly an inverse t-shape geometry with a total width of 30 cm, as can be observed in Figure 5.3.

Typical audio events and background noises that can arise in surveillance applications are not easily available, mainly because of the confidential nature of the data and also because such abnormal situations or events are rarely recorded. To be as close as possible to real-world surveillance application conditions, acoustic data from two different scenarios was collected, each



(a) (b)

Figure 5.3: Michrophones array geometry (cm). (a) Array 1. (b) Array 2.

comprised of multiple recordings. In the first, acoustic events were recorded, while in the second, noise recordings were recorded. Therefore, two different recording campaigns were conducted in the following scenarios:

- **Indoor**: For acoustic event recordings, it is necessary to have a very low noise level. Thus, an office room was selected. Figure 5.4 shows the layout along with the microphone placement. The three different measurement locations are also marked; they were placed according to a practical criterion which took into account the microphones' positions and the SNR of the recordings.

- **Outdoor**: For the background noise recordings, two different outdoor scenarios were selected. In the first case, the noise present in a public area of a shopping center was recorded (Figure 5.5a). In the second case, an open-field space of around $250\,m^2$, shown in Figure 5.5b, was selected to record the noise of a power generator.

5.1.3 Detection system

A robust detection system suitable for the automatic detection of dangerous acoustic events in public places and infrastructures was evaluated. The proposed system processes and interprets the audio signals acquired by a distributed microphone array in order to discriminate between noise and novel events. It must be pointed out that the system is capable of integrating any of the energy detectors studied in Chapters 3 and 4. In this section, the overall architecture and main functionalities of the detection surveillance system are presented. The system is based on the processing of the acoustic signals and is made up of the following subsystems:

- Initialization module.

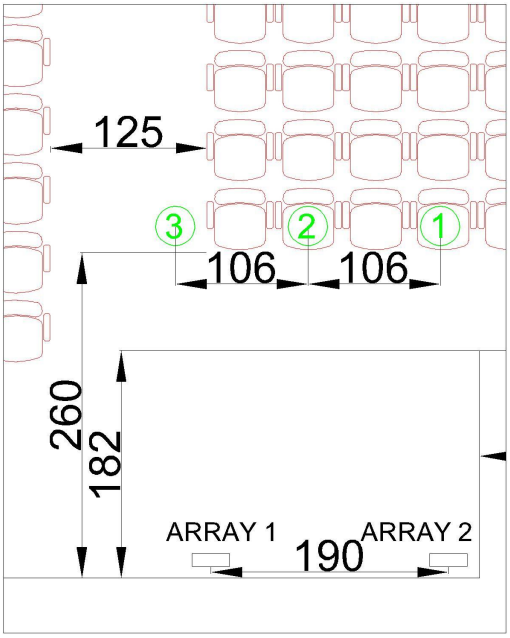


Figure 5.4: Room layout and microphone placement (measurements in cm).

(a) (b)

Figure 5.5: Acoustic scenarios used for background noise recordings. (a) Shopping center. (b) Open-field area with power generator.

- Acquisition module.
- Detection module.

Figure 5.6 shows the general architecture of the system. As can be observed, the acoustic signals recorded by the different sensors are collected by the acquisition module, which provides a data frame to the detection

module. This module is the central processing unit and it has the responsibility of deciding when some acoustic event has taken place in the area under study. Furthermore, it must be noticed that the detection and the acquisition module take the initialization parameters set by the initialization module.

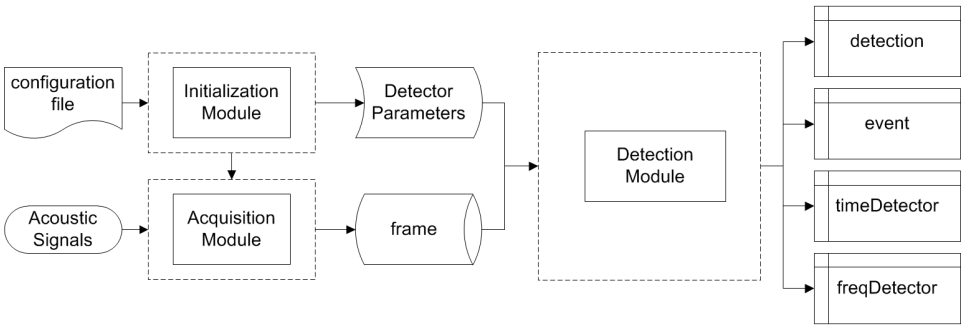


Figure 5.6: Block diagram of the acoustic event detection system.

In the following sections the main features of the modules involved in the acoustic detection system will be described.

5.1.3.1 Initialization module

In order to ensure proper functioning of the system, an initialization of the parameters and components is required to configure the modules of the system. The initialization module is responsible for setting all the internal parameters needed for the acquisition and processing unit (e.g. the sample frequency, number of channels, PFA, frame size, data buffer length, etc.). It also offers the possibility of selecting some of them by means of a configuration file which can be accessed and modified outside the system, depending on the particular scenario or situation.

5.1.3.2 Acquisition module

This module is responsible for acquiring the data from the digital devices, such as the audio card, and organizing them in an appropriate format that can be easily accessed by the further processing modules. For that purpose, the input data is stored in circular buffers for later use. The circular buffers are data structures that use a single fixed-size buffer, as if it were connected end-to-end. This type of structure is widely used for buffering data streams. This type of structure is managed by pointers and the FIFO (First In, First Out) criterion. When storing data, it finds the first free slot; to extract data, it eliminates the oldest one in the buffer. The data are stored with

this procedure in successive memory positions and thus, the space required remains constant, with the benefits this entails.

5.1.3.3 Detection module

The detection module reports an event when it is detected in the presence of a background noise. This module is of great importance since it is responsible for providing a good PD for the PFA (set in advance). Such a detection system can be efficiently used for the automatic detection of anomalous audio events in public or private spaces. It also enables the possibility of combining it with further processing stages (acoustic source identification and localization) if more than one sensor is used. Therefore, it has a considerable effect on the performance of other modules, as will be discussed in Chapter 6.

It is important to notice that the detection module manages a detector for each of the signal sources. If the system uses multiple microphones, this module will manage them by initializing an independent detector for each of them. The fact of using more than one microphone can be advantageous: introducing this redundancy allows the possibility of combining the decisions of each of them to reach a more reliable final decision. The use of multiple channels of acquisition is not only advantageous for detection: obtaining more reliable results and enabling the localization of sound sources when several microphones are necessary may also prove useful to improving the classification of events. In our case, the presence of an event is decided when at least half plus one of the detectors determine it. Other more complex techniques of fusion decision could be applied [7, 31], but this is not relevant to the objectives of this work.

Figure 5.7 presents a more detailed description of the detection module, and its main functionalities and characteristics are found below:

- **Background noise estimation**: The detection module integrates a training phase in which the background noise characteristics are learned in order to estimate the whitening matrix $\mathbf{R_w}$ for the implementation of the PED; and the transformation $\mathbf{U}$ and the non-linear function $g(\cdot)$ for both the EED and the PEED detectors. Due to the possible changing noise conditions in real acoustic scenarios it is also necessary to reestimate the noise characteristics to maintain a general applicability of the system to any type of situation. This procedure is based on a *reestimate* flag, activated by the event manager submodule after taking into account the result of the energy detector submodule.

- **Pre-processing data**: When necessary, this submodule is responsible for adapting the acoustic signal by applying the appropriate preprocessing functions commented in the previous submodule. In the case of non-Guassian or non-independent noise, the system will thereby

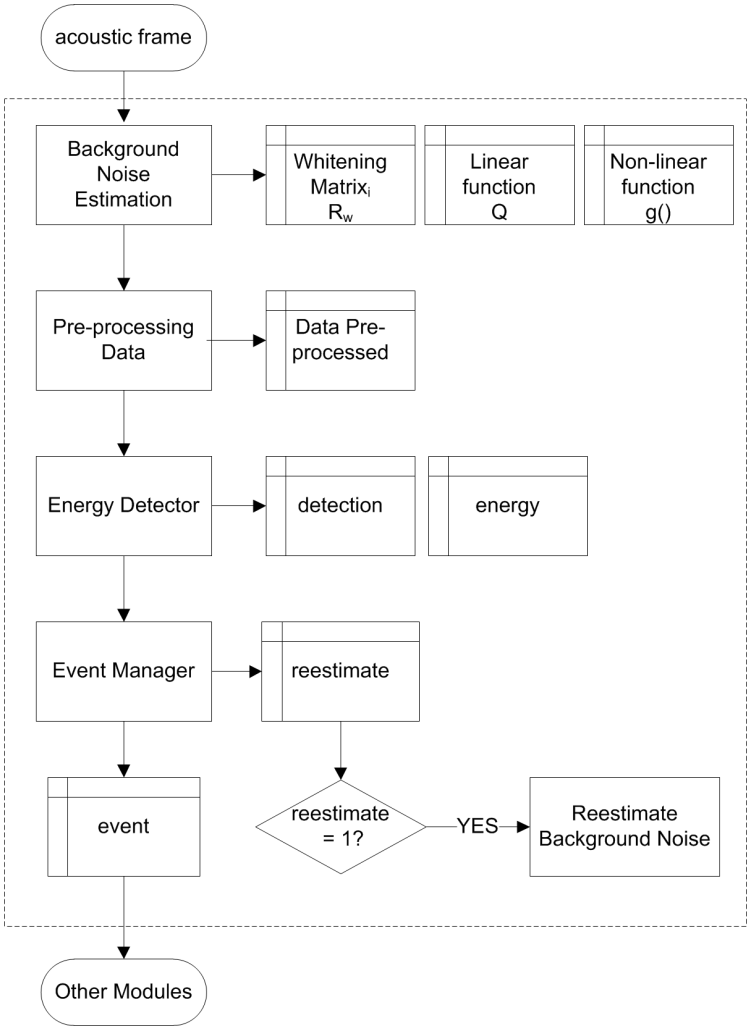


Figure 5.7: Description of the detection module.

be in condition to reach the most appropriate decisions in the detection of events by applying function $g(\cdot)$ and matrix $\mathbf{Q}$.

- **Energy detector**: This is the main core of the module and it implements any of the energy detector types and extensions described in Chapters 3 and 4, depending on the background noise conditions and on the available *a priori* knowledge of the sound sources. In all cases, the amplitude of the acoustic signal is observed and the energy is computed and compared, frame by frame, with an appropriate threshold. The resulting test provides meaningful evidence of the presence of an acoustic event when the energy is above the threshold.

- **Event manager**: The *detection* result of the previous submodule indicates the presence of a signal at a sampling interval determined by the frame size selected. Due to the computational constraints, this range of samples can sometimes be very limited, and thus insufficient by itself, to determine the presence of an event. For this reason, it is necessary to consider several consecutive intervals in order to establish the presence of an event, and then activate the output flag, denoted by *event*. This approach leads to a new generalization of the energy detectors that will be described in greater detail in Section 5.3.2. In addition, this submodule is responsible of activating the *reestimate* flag in case the environment conditions change. This fact permits the recalculation of the background noise characteristics and the subsequent adaptation of the ED to either sudden or slow variations of the noise level or variance.

5.2 Acoustic database

With the objective of studying the performance of the energy detector and the extensions presented in the previous chapters, an acoustic database was created. This database is composed of different acoustic signals. On the one hand, various sound events were recorded indoors. On the other hand, two types of real background noises were considered (both are analyzed and described below). The equipment and scenarios used for the recordings were previously described in Section 5.1.2.

5.2.1 Analysis of real background noises

In many practical problems of interest, it is entirely appropriate to assume stationary uncorrelated Gaussian noise: it results in easily implemented detectors. However, in real-world acoustic scenarios, it is not always possible to characterize the noise under these conditions due to its variability over time. For this reason, this section offers an analysis of the recorded background noises based on Gaussianity and independence. The study of the noise characteristics is essential for the detection problem since they can considerably influence the detector performance and can determine some specific parameters, such as the values N (observation vector size) and K (buffer size), as will be shown in Section 5.3.1.

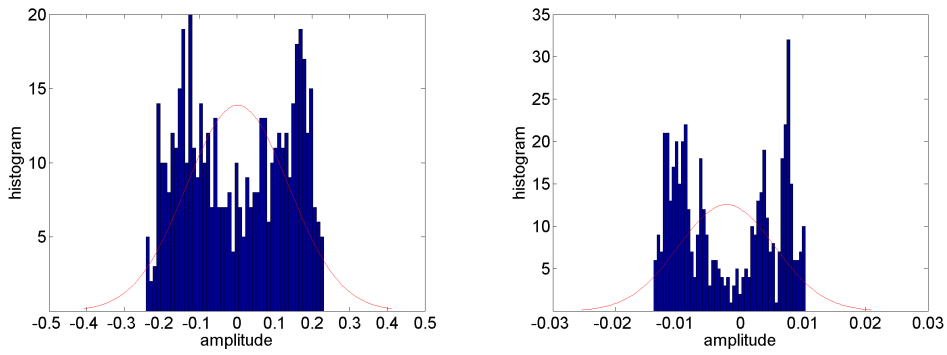
Two real ambient noises were studied. The first was a crowd of people present in a *shopping center*, and the second consisted of the noise from a continuous *power generator* that can typically be found in office air conditioners. These are good examples of acoustic noises with non-Gaussian distribution, non-independent samples, and non-stationary evolution over time. In the first case, approximately 16 minutes of ambient noise was recorded at dif-

ferent periods of time; in the second, approximately 42 minutes of ambient noise was acquired. For the recordings of the background noise sources, the stereo microphones from a digital video camera were utilized.

5.2.1.1 Gaussianity

Section 3.2 included an examination of how the presence of non-Gaussian noise can lead to poor detection performance. Therefore, it is very important to study the non-Gaussian characteristics of the real acoustic noises considered.

In this case, the two methods to measure the non-Gaussianity of a given process (explained in Section 3.1.1) are applied: histogram plots and kurtosis estimation. Histograms are widely used as a simple, but informative, method of approximating the PDF of the data displayed, as they provide a visual impression of the distribution shape. In Figure 5.8, the histograms of the *shopping center* and the *power generator* noises are illustrated and compared with a Gaussian PDF of the same standard deviation and mean, respectively. In both cases, the objective is to observe how the noise PDF differs from the Gaussianity assumption.



(a) (b)

Figure 5.8: Estimated PDF of two real acoustic noises versus a Gaussian PDF (solid line) of the same standard deviation and mean ($N = 512$). (a) *Shopping center*. (b) *Power generator*.

The degree of non-Gaussianity can be more specifically calculated by estimating the kurtosis, which measures the degree of peakedness or flatness of a distribution. The kurtosis was calculated for the two types of noises and was found to be 3.53 for the *shopping center* noise, and 3.97 for the *power generator* noise (where 0 stands for the Gaussian PDF).

5.2.1.2 Statistical dependence

The statistical dependence problem for the non-Gaussian case is quite complex (as was demonstrated in Section 3.2), and there has not been much effort devoted to it within the field of detection, especially with regards to high dimensional data vectors. Mathematically, statistical independence is defined in terms of probability densities. The random variables x and y are said to be independent if, and only if,

$$p_{x,y}(x,y) = p_x(x)p_y(y). \tag{5.1}$$

This means that the joint density $p_{x,y}$ of x and y must factorize into the product of their marginal densities $p_x(x)$ and $p_y(y)$, respectively. In other words, the information about the value of x does not yield any information about the value of y.

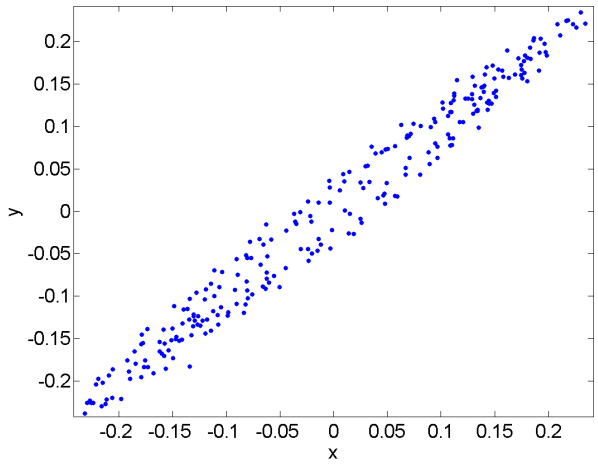


Figure 5.9: Joint distribution of two dependent components of the *shopping center* noise with non-Gaussian densities.

To illustrate the statistical dependence of the background noise, two random variables x and y are considered. They correspond to the even and odd samples extracted from the *shopping center* noise, respectively. In Figure 5.9, the joint density of the two variables is plotted for 256 samples. As observed, they are not independent: it is possible to predict the value of one of them, say x, from the value of the other, y. Similar results were obtained for the *power generator* noise. Furthermore, it is also possible to experimentally observe the statistical dependence of the noise by comparing the results achieved with real noises versus the equivalent results obtained for the Gaussian independent simulated noise, as will be shown in Section 5.4.2.

5.2.2 Description of the acoustic events

A complete sound database of different acoustic events was recorded and used for the experiments reported in Sections 5.4 and 5.5. In this case, the sounds were recorded in an indoor room like the one described in Section 5.1.2. It is possible to divide the sounds into two main types, according to their respective time duration:

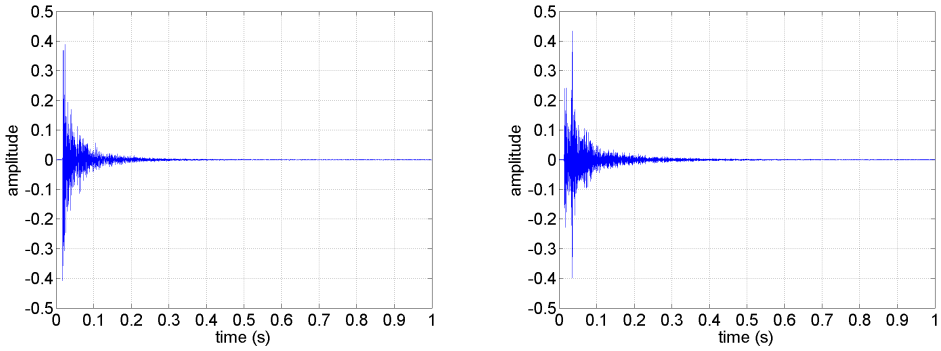
- **Impulsive (IM)**: This type of sound is characterized by a short impulse response in time, which consists of an initial and high intensity peak of energy followed by a transition in which the energy decreases slowly. This means that most of the energy of the impulsive sounds is concentrated at the beginning of the event. In our case, the acoustic events that fulfill these characteristics, and have a time duration shorter than 1 second, were considered as impulsive sounds.

- **Non-impulsive (NIM)**: The non-impulsive sounds have a longer duration in time. They are characterized by a lower amplitude than the impulsive ones and a more uniform response (in terms of energy distribution) over a specific period of time. We considered as non-impulsive sounds those acoustic events with the aforesaid characteristics and with a time duration equal to or greater than 1 second.

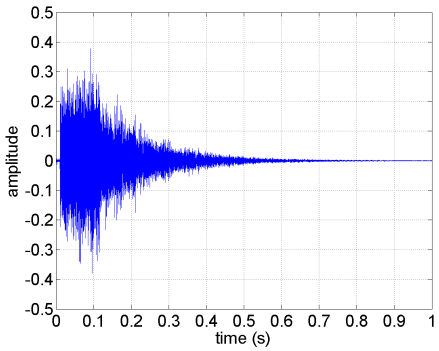
It must be noticed that one of the reasons for selecting a variety of impulsive and non-impulsive sounds is related to the fact that they occupy different frequency bands in the spectrum. This issue is of particular interest since the detection process can be challenging not only in the time but also in the frequency domain as will be studied in Section 5.5.

Due to the difficulty of recording sound events commonly found in dangerous situations (such as *gunfire* or *explosions*), the measurement campaign involved recording a number of alternative sounds which are easier to generate and offer similar behavior in terms of acoustic response. Therefore, impulsive sound sources such as *bursting balloons*, *claps* and *breaking glass* were generated and recorded. Additionally, *metallic sounds* and *human speech* were also analyzed as non-impulsive sound sources. Approximately 3 minutes of data (corresponding to 25 events) were acquired for each sound source in 3 different room positions. The room and microphones placement layout was the same as that shown in Figure 5.4. The structure of 8 microphones led to the recording of a total of 600 acoustic events for each sound source. Figure 5.10 illustrates the time realization of three impulsive sounds, while Figure 5.11 depicts the waveforms of the non-impulsive examples.

In order to study the detector performance in these situations, it is necessary to obtain the experimental results of PD and PFA. The events of the recordings then have to be automatically annotated using a tool which specifies the exact location (start and end time) of each audible event within

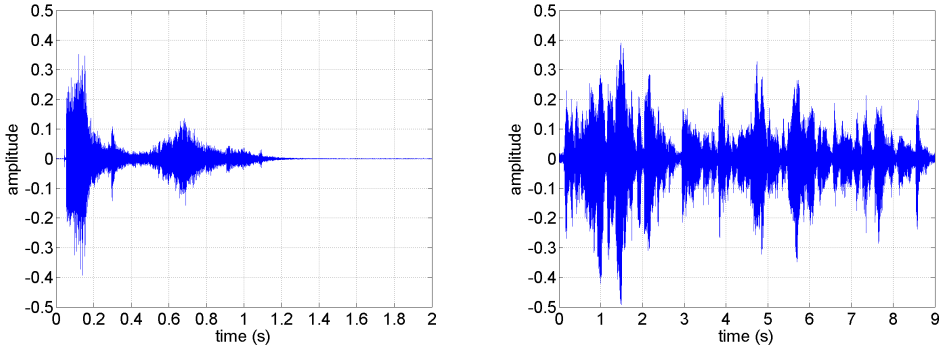


(a) (b)



(c)

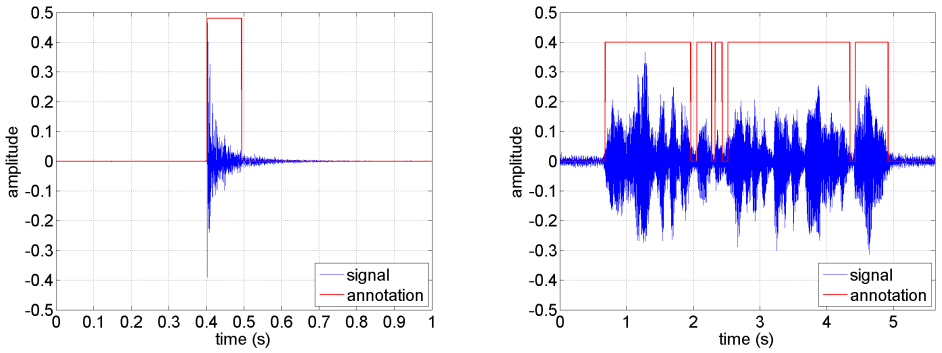
Figure 5.10: Time realization of impulsive sound events. (a) *Balloon*. (b) *Clap*. (c) *Breaking glass*.



(a) (b)

Figure 5.11: Time realization of non-impulsive sound events. (a) *Metallic sounds*. (b) *Speech*.

the audio file. Figure 5.12 presents an example of the annotation process applied to the *bursting balloon* and the *speech* recordings. It must be taken into account that this annotation method provides an estimated window in which the sound event is assumed to be present, and this includes the decreasing part of the signal. This fact can be better appreciated in the impulsive sound example. It affects the evaluation process of the different detectors when comparing them with their theoretical results as the impulsive sounds are obtained assuming that the signal energy is uniformly distributed through the whole observation interval.



(a) (b)

Figure 5.12: Annotation result. (a) *Balloon*. (b) *Speech*.

5.3 Practical background noise considerations

One of the most problematic aspects in the application of the proposed detectors is the necessity for *a priori* knowledge about the noise autocorrelation matrix $\mathbf{R_w}$. In practice, it can be estimated using (2.24) with a set of training samples of observation noise vectors (K). Obviously, good estimates of $\mathbf{R_w}$ require a relativy great value of K. This can be obtained by using a large interval of training noise samples (let us call T the total number available) and/or reducing the observation vector size N. However, the value T cannot be made arbitrarily large as there are some limiting factors: namely, the computational burden and time interval duration allowed for training, as well as the possible medium and long-term non-stationarity of the noise (which imposes a limit on the maximum interval where the obtained estimate is valid). On the other hand, N cannot be made arbitrarily small since this would produce an observation time scale that is too small, thereby leading to unstable behavior in the detectors. All these factors influence the estimate of $\mathbf{R_w}$ and can result in a much greater PFA than that designated.

5.3.1 Experimental parameter setting

Selecting the best combination of K and N becomes an important problem of practical interest. Unfortunately, it is not easy to find an analytic closed equation giving us the best values. Instead, an experimental fitting which may be considered part of the overall calibration of the detector is considered. We describe a procedure, which, although adapted to our particular case, may have general applicability to other scenarios. The choice of N and K was carried out from among a predefined range of values that took into account the aforementioned aspects and the applicability to different types of detectors. The two available background noise recordings were processed using the ED. As only noise was present, setting a very restrictive PFA (i.e. 10^{-12}) no detections were expected. Therefore, we selected the combination of N and K that resulted in the minimum number of measured false alarms.

N \ ***K***	512	1024	2048	4096
128	0.0153	0.0147	**0.0122**	0.0135
256	0.0308	**0.0302**	0.0310	0.0310
512	0.0577	0.0577	0.0579	0.0527
1024	0.1007	0.0995	0.0856	0.0874

Table 5.1: PFA of the ED in the presence of *shopping center* noise.

Four different vector sizes (N) were used: 128, 256, 512, and 1024; and for each one of them, four buffer sizes (K) were tested: 512, 1024, 2048, and 4096. In Tables 5.1 and 5.2, the results of PFA are presented when using the ED with the *shopping center* and the *power generator* noises, respectively. Considering the smallest PFA obtained as an indicator of better performance, the best combinations are marked in bold. Notice that the experimental PFA obtained differs from the designed one. This is due to the non-Gaussian and non-independent real noise characteristics, and to the sample estimate effects of $\mathbf{R_w}$. For the *shopping center* noise, the best PFA was achieved for a vector size of 128 and 256 using a buffer size of 2048 and 1024, respectively. In this case, since both combinations are possible solutions, one of them was selected bearing in mind the computational requirements. Since the *power generator* noise has different stationarity characteristics, the best results were achieved for a 1024 vector length and a 1024 buffer size. We used other types of energy detectors (EED and PEED) and, as expected, similar results were obtained as the adjustment of N and K depend primarily on the noise characteristics.

N \ K	512	1024	2048	4096
128	0.0522	0.0503	0.0511	0.0501
256	0.0531	0.0512	0.0536	0.0526
512	0.0317	0.0323	0.0321	0.0307
1024	0.0210	**0.0207**	0.0208	0.0208

Table 5.2: PFA of the ED in the presence of *power generator* noise.

5.3.2 Generalization of the energy detector

The previous section outlined the procedure to select the best combination of K and N. In such cases, where N can be assumed to be lower than the signal time duration, unstable detector behavior is expected. Therefore, the following generalization of the detectors is proposed in an effort to compensate for this degradation. This generalization will be evaluated as well when it is applied to the ED, but similar procedure can be derived for the other detectors.

Considering (2.27), the ED output, denoted *detection*, is the result of comparing the energy of the observation vector $\mathbf{y}_p$ (size N) with the corresponding threshold λ. It is set to 1, in case the energy overcomes the threshold H_1, and to 0 for the H_0 hypothesis. The generalization is based on this procedure, but in addition, a consecutive number of *detections* (M) will be required to activate the output, termed hereafter as *event*. The resulting *event* energy detector (EDe) is capable of reducing the PFA by ignoring the outlying *detections* that do not correspond to real acoustic events and only detecting the presence of those desired.

Taking into account the previous statements, the behavior of the EDe can be modeled following a binomial distribution. Considering the probability of getting exactly M *detections* in p trials, the corresponding PFAe and PDe of the EDe can be derived as follows:

$$PFAe = \binom{p}{M} PFA^M (1 - PFA)^{p-M} \tag{5.2}$$

$$PDe = \binom{p}{M} PD^M (1 - PD)^{p-M} \tag{5.3}$$

where the PFA and PD are obtained from the ED described in (2.17) and (2.18). The performance of this new generalization is illustrated in Figure 5.13, where the ROC curves of the ED and the EDe are derived from the theoretical expressions described before. In this example, the minimum resolution time for acoustic events is set to 40 ms, corresponding to 4 consecutive detections using 256 as the observation vector size (N) and a sample frequency of 24 kHz. This means that 4 active *detections* are required across

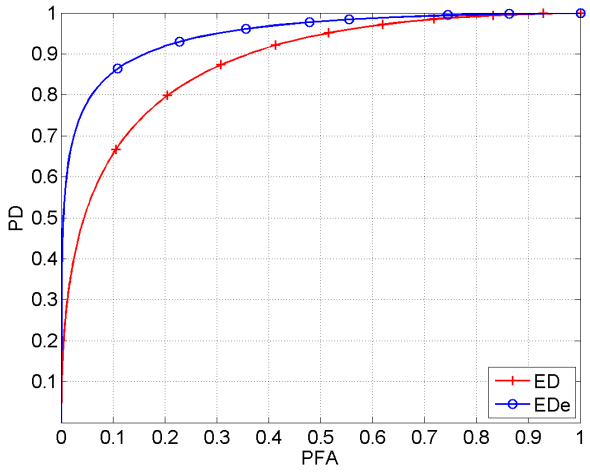


Figure 5.13: ROC curves of the ED and the EDe with independent Gaussian noise, $SNR = -8$ dB and $N = 256$.

four trials to obtain an activated EDe output ($event = 1$). The EDe curve obtains an improvement in PD up to 25.9% over the ED for a PFA of 10^{-3} thanks to the considerable reduction in false alarms arising in the ED. In the following, this generalization will be extended to the detectors presented in Chapter 3, thus yielding the EEDe and PEEDe.

5.4 Performance of the ED and its generalization with acoustic signals

In this section, the results of our experiments are presented and discussed. First, our attention turns to the evaluation of the ED and the EDe with different sound sources and in the presence of Gaussian noise. Then, the performance of the EDe is studied with non-Gaussian and non-independent real noise and then compared with the results of the generalized extensions presented in Chapter 3.

For all experiments (and considering the computational limitations), the reestimate time of the noise characteristics was 1.5 seconds, the PFA range studied was from 10^{-8} to 1 and the minimum resolution time event was set to 40 ms. Different SNR were considered, and the sizes of the observation vector N and the noise buffer K were set according to the results obtained in Table 5.1.

5.4.1 Acoustic event detection in Gaussian noise using the ED and the EDe

We tested both the ED and the proposed EDe by means of the probability of detection (calculated for different adverse SNR), and using Gaussian noise and the five types of acoustic events: *balloon*, *hammer*, *clap*, *glass*, and *speech*. It is worth noting how, as expected, there was an improvement in PD when using the EDe versus the ED. Table 5.3 shows the absolute percentage of improvement reached (in terms of PD) across all measurements, for a PFA of 10^{-3}. For low SNR, the number of missdetections increased; the ED cannot avoid them, unlike the EDe. In the case of *balloon* event and $SNR = -6$ dB, an absolute improvement in PD of up to 30.5% was attained.

SNR(dB)	***balloon***	***hammer***	***clap***	***glass***	***speech***
-2	28.6	29.9	26.3	11.8	5.3
-4	27.8	30.3	26.7	10.6	11.5
-6	30.5	30.4	22.9	10.7	9.1
-8	27.1	28.9	19.7	9.8	6.1
-10	23.3	28.4	14.8	8.6	3.4

Table 5.3: Improvement percentage (%) in PD of the EDe versus the ED for $PFA = 10^{-3}$.

Now, let us highlight this improvement by comparing the graphed outputs of the ED (*detection*) and the EDe (*event*), using the corresponding threshold for a PFA of 10^{-8}. Figure 5.14 shows three acoustic events of type *clap* in the presence of Gaussian noise and with $SNR = -4$ dB, where 0 stands for no detection and 1 for sound detection. As can be observed, the number of false detections is considerably reduced when 4 consecutive active *detections* are required to generate one *event* in the EDe. This circumstance ensures reliable detection results for all sound sources and leads to an improvement of performance. In the following this generalization will also be used with the extended detectors seen in Chapter 3.

5.4.2 Comparison of EDe, EEDe and PEEDe with real background noises

Since the particular acoustic event does not influence the comparison results of the different energy detectors, the experiments were conducted using one of the sound sources of the database, in this case the *hammer blows*. In Figure 5.15, the resulting ROC curves of the EDe are shown for the *shopping center* and the *power generator* noises (dashed lines), and they are compared with the Gaussian noise case (solid line). The SNR is set to -4 dB and the observation vector length N to 256. As expected, a loss in detectability can

be appreciated when using the EDe with non-Gaussian and non-independent noises.

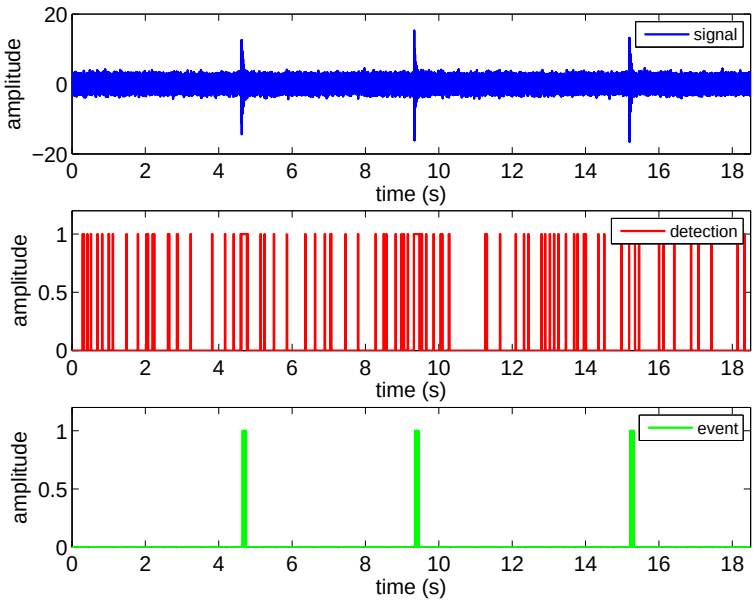


Figure 5.14: Output of the ED (*detection*) and EDe (*event*) when detecting an acoustic signal of type *clap* in white Gaussian noise and $SNR = -4$ dB.

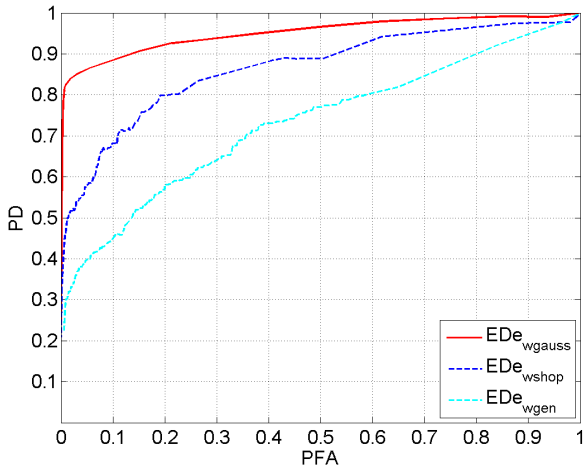


Figure 5.15: ROC curves of the EDe with Gaussian and non-Gaussian noises (*shopping center* and *power generator*) with $SNR = -4$ dB and $N = 256$.

However, in real non-Gaussian noise scenarios, the detectability can be significantly improved by using the EEDe and the non-parametric transformation proposed in Section 3.2.3. In Figure 5.16, the ROC curves of the EEDe and the PEEDe are set against the EDe when detecting *hammer blows* in the presence of the *shopping center* noise. Only the results of this noise are presented since it is more non-stationary than the *power generator* one. The SNR was approximately -6 dB while the observation vector length (N) was set to 100 (due to the highly time-consuming effort required to estimate the ICA transformation and the inexistent convergence of the algorithm for $N = 128$).

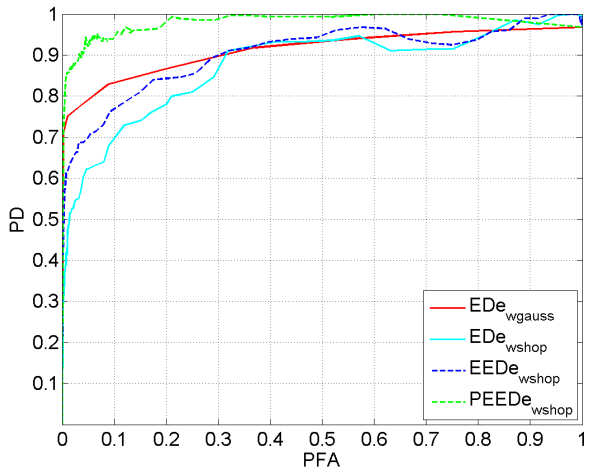


Figure 5.16: ROC curves of the EDe, the EEDe and the PEEDe in the presence of the *shopping center* noise with $SNR = -6$ dB and $N = 100$.

The results show that, since the EEDe considers the non-Gaussianity of the noise, it performs better than the EDe achieving an improvement in PD up to 12.5% for a PFA of 10^{-3}. However, and especially for low PFA values, the EEDe does not achieve the same results of the EDe for uncorrelated Gaussian noise, as it does not take into account the dependence of the real noise. In order to deal with the non-independence of the noise, it was necessary to use the PEEDe, which resulted in an improved overall performance in the EEDe, evidenced in Figure 5.16. For a PFA of 10^{-3}, the PEEDe achieves an absolute improvement in PD of up to 8%. Notice that the PEEDe operating in non-Gaussian and non-independent noise performs better than the EDe operating in Gaussian and independent noise, given the same SNR. This fact is discussed in [64] and is related to the change in SNR experienced after applying the non-linear transformation $g(\cdot)$.

5.5 Performance of the MED in acoustic event detection

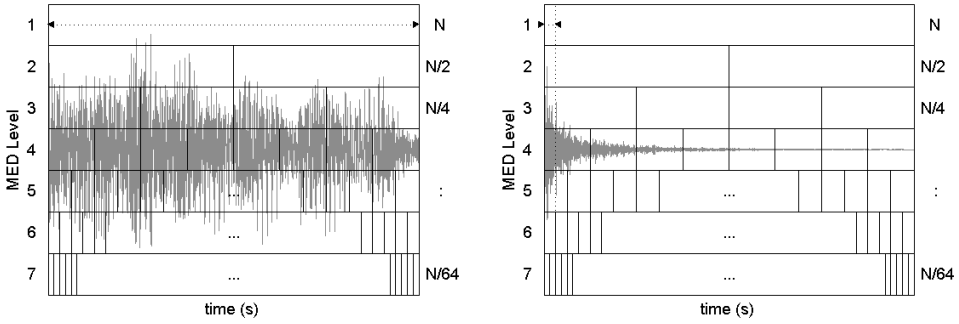
The theoretical problem of detecting unknown signals with unknown duration was previously presented in Chapter 4, where a MED structure was implemented in order to match the different possible signal durations. No evaluation of this novel approach was performed in any real-world scenario or practical application; but the improvements achieved by the theoretical results justify the utility and applicability to acoustic scenarios. Therefore, the focus of this section is to present a detailed evaluation within the framework of acoustic scene analysis, where the multiple energy detector can provide significant improvement in the detection of acoustic events. In particular, the performance of the MED is studied in the time and frequency domains since detection of signals with unknown bandwidth is also of particular interest [60].

5.5.1 Experimental setup and analysis of acoustic events

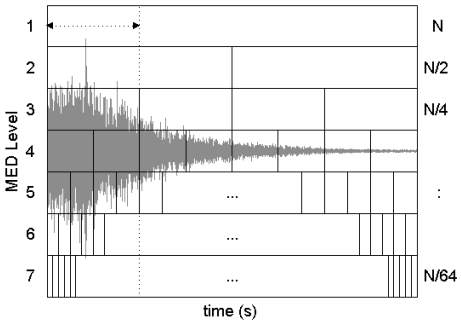
These experiments were performed with the objective of studying the performance of the MED in real-world acoustic scenarios. Various acoustic events of varying nature and duration were recorded in a typical office room, using a multichannel audio data acquisition unit with a sampling frequency of 24 kHz (as previously described in Section 5.1.2 and 5.2). As an example, a time and frequency MED structure of 7 layers was used with an original observation vector in the highest layer of $N = 16384$ samples, leading to a total duration of 0.68 seconds and a total bandwidth of 12 kHz, respectively. The PFA of the individual EDs was set to 10^{-8} since this is a reasonable value used in detection systems.

In order to contemplate the largest set of cases, three acoustic events (arranged according to time duration) were considered. Impulsive sound sources (*claps* and *breaking glass*) were analyzed, along with *human speech* as a non-impulsive sound source. In Figure 5.17, a time realization of the acoustic events is superimposed on the time MED structure. As can be observed, the *speech* signal extends its energy uniformly across the whole initial observation vector. On the contrary, the *clap* signal concentrates its main energy across the first segment of layer 6, corresponding to $N/32$ samples. Finally, as an intermediate example between the two aforementioned cases, the *breaking glass* signal extends its energy primarily throughout the first segment of the third layer.

Figure 5.18 shows the absolute value of the normalized frequency response of each acoustic signal. Again the frequency MED structure used is superimposed. As we can see, the acoustic signals selected have different spectrums and characteristics. The most important spectral information



(a) (b)



(c)

Figure 5.17: Realizations of the acoustic signals superimposed on the time MED structure of $L = 7$. (a) *Speech*. (b) *Clap*. (c) *Breaking glass*.

about the *speech* signal is concentrated mainly between 0 and 1 kHz. The *clap* example presents a low frequency spectrum mainly concentrated between 500 Hz and 2.2 kHz. The third reveals a totally different frequency response and corresponds to the *breaking glass* signal. In this case, the spectrum has higher frequency components and a wider bandwidth, mainly spanning between 3 and 6 kHz.

5.5.2 Acoustic detection results

This section presents and discusses the experiment results. The time and frequency ROC curves of the MED are obtained for different sound signals. In both cases, several $SNRN_o$ defined in the first layer of the structure ($SNRN_o = SNR/\sqrt{2N}$) were generated by adding WGN; but only those that permitted a reasonable comparison between the different ROC curves under study were presented for each sound source. In addition, in the time as well as in the frequency domain, the ROC curves comparison is studied

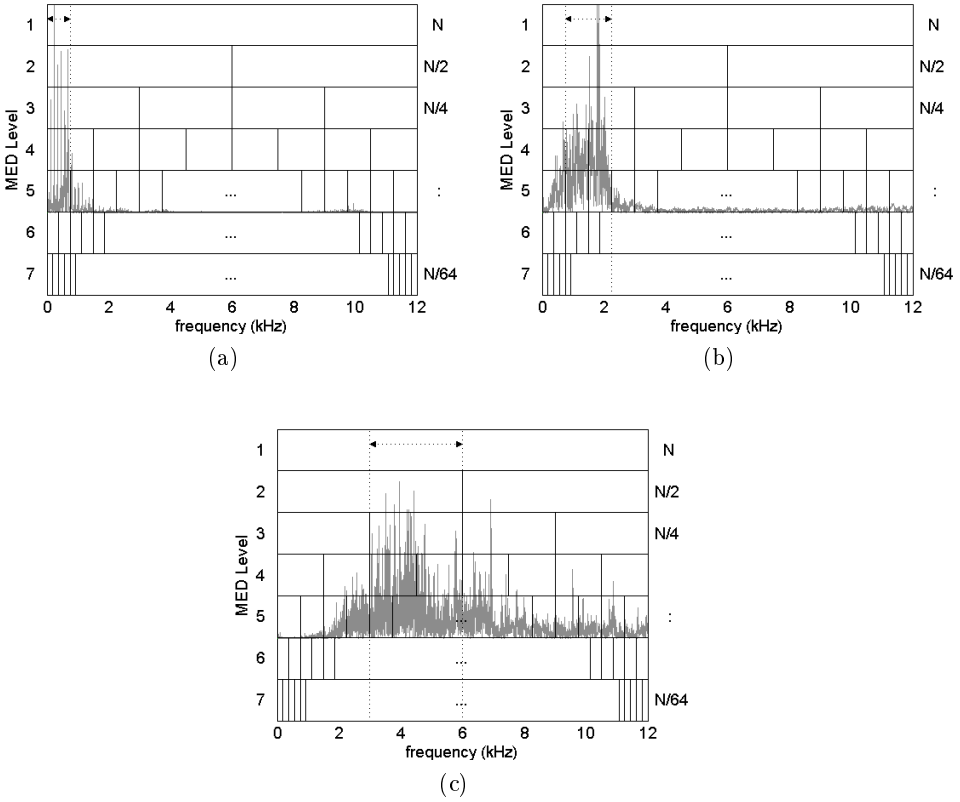


Figure 5.18: Normalized frequency response of the acoustic signals superimposed on the frequency MED structure of $L = 7$. (a) *Speech*. (b) *Clap*. (c) *Breaking glass*.

for several partitions L (number or layers used in the MED structure) of the original observation vector (with dimension N).

The principal objective is to show and validate the interest in using the MED instead of the ED (only one layer, $L = 1$) when the signal duration or bandwidth is unknown. Several cases are thus considered using the MED to detect signals with a time duration and bandwidth comparable to the segment length of different layers.

5.5.2.1 Signals with unknown time duration

Figure 5.19 presents the time MED performance when detecting the three acoustic signals of different time duration. Figure 5.19a shows the ROC curves for the *speech* signal, which extends uniformly across the whole initial observation vector. In this case, the best performance is reached for $L = 1$, equivalent to the single ED. As expected, some degradation is experienced

in the ROC when using unnecessary partitions ($L > 1$). On the contrary, as observed in Figure 5.19b (where a *clap* signal is to be detected), an improvement in MED performance is achieved by increasing the number of layers used, versus the single ED ($L = 1$). The best results are obtained for $L = 7$ since the signal duration is more comparable to the observation segment of the bottom layer.

Figure 5.19c illustrates an intermediate example where the acoustic signal has a time duration similar to that of the segment used in layer 3. As expected, the best ROC curve corresponds to $L = 3$. However, note how the degradation suffered when using more layers ($L > 3$) is not as significant as the degradation experimented when using no subdivisions ($L = 1$). Therefore, it is advisable to use as many layers as possible in the MED, observed by comparing the degradation in Figures 5.20c for $L = 7$ and $L = 1$.

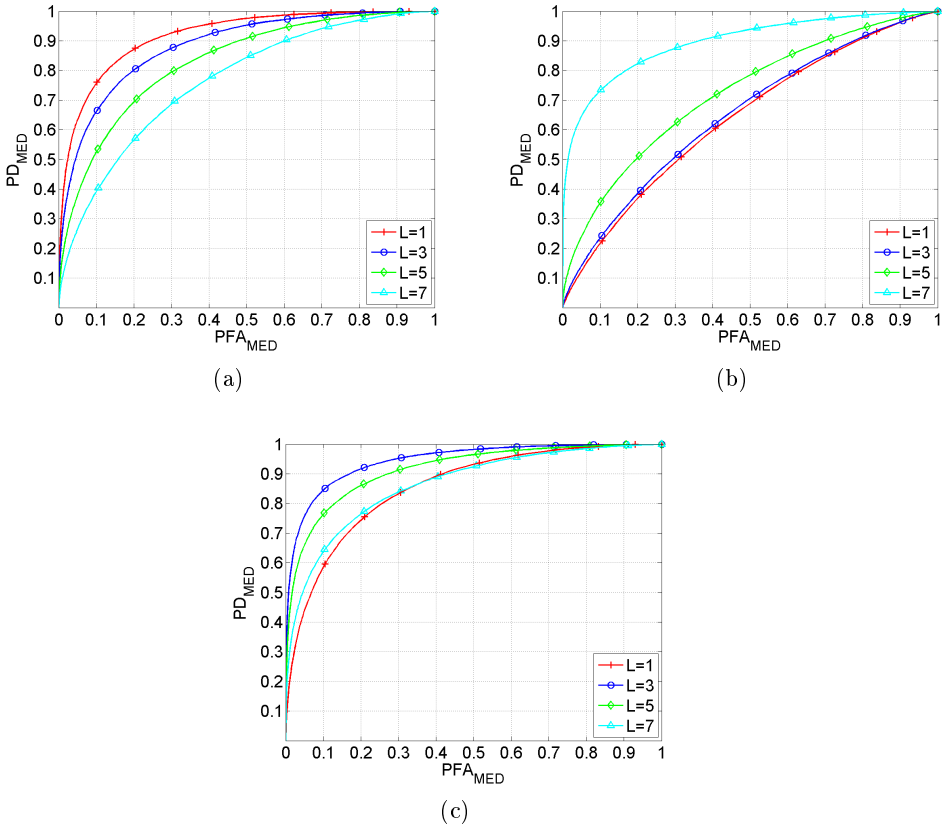
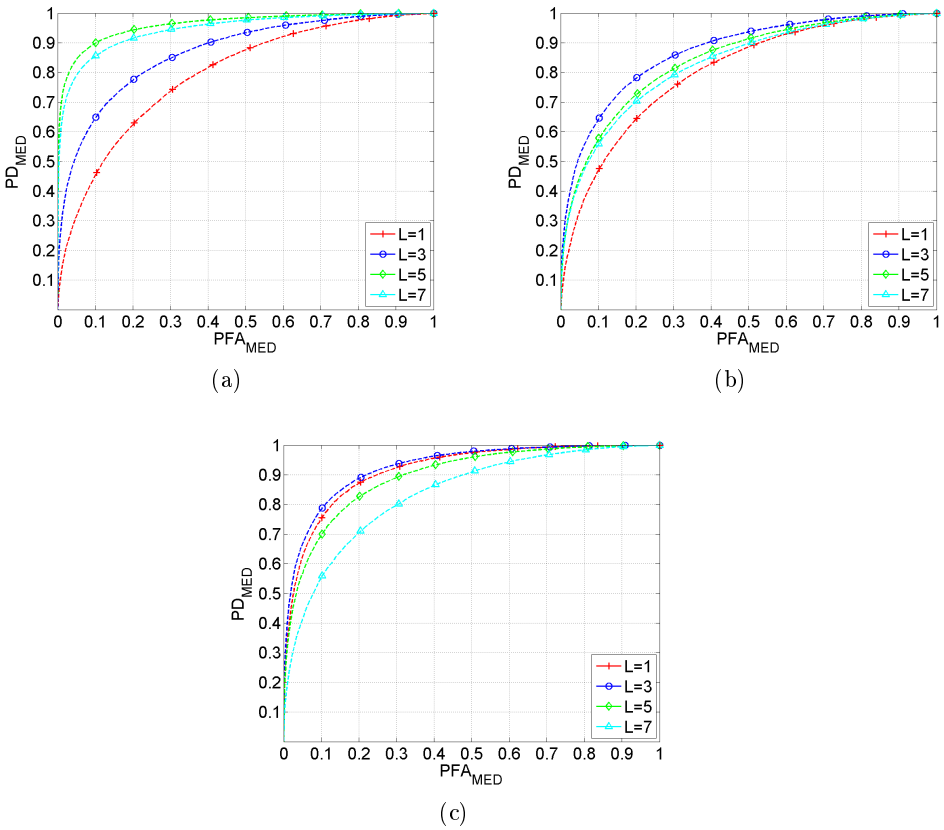


Figure 5.19: ROC curves of time MED structure using different layers (L). (a) *Speech* ($SNRN_o = 2$). (b) *Clap* ($SNRN_o = 0.5$). (c) *Breaking glass* ($SNRN_o = 1.5$).

5.5.2.2 Signals with unknown bandwidth

Figure 5.20 shows the experimental ROC curves obtained when detecting acoustic signals with the frequency MED structure. The first example is shown in Figure 5.20a. The best detection results are obtained for $L = 5$ since the *speech* bandwidth has a similar length to the observation segment of layer 5. The performance decreases when using more layers ($L > 5$), but as expected, the degradation is more significant when less layers ($L < 5$) are used, especially for $L = 1$, which corresponds to the ED.



(a) (b)

(c)

Figure 5.20: ROC curves of frequency MED structure using different layers (L). (a) *Speech* ($SNRN_o = 1$). (b) *Clap* ($SNRN_o = 1.5$). (c) *Breaking glass* ($SNRN_o = 2.5$).

The remaining two examples are presented in Figures 5.20b and 5.20c, where a *clap* and a *breaking glass* signals were detected. In both cases, the optimal number of layers to be employed in the frequency MED structure is equal to 3, as observed. Comparing the spectrums of both signals (Figure 5.18b and Figure 5.18c), we can observe how the *breaking glass* signal has a broader bandwidth; therefore, a smaller number of optimal layers would

be expected. This fact can be explained: in the *breaking glass* example, the second best result is obtained for $L = 1$, while for the *clap* case it was $L = 5$, corresponding to its actual bandwidth.

It must be pointed out that the frequency noise vectors are obtained taking the absolute value of the white Gaussian noise spectrum; therefore, they are Rayleigh distributed [43]. When the noise follows a non-Gaussian distribution, it is possible to implement other extensions of the ED, as presented in Chapter 3. This would lead to a general improvement of PD for all ROC curves in Figure 5.20, although the relative comparison among the different layers L would remain the same, and similar results would be achieved.

5.6 Conclusions

In this chapter, the ED and the possible extensions viewed in previous chapters have been evaluated in real-world acoustic applications where the sound sources are not entirely known. In particular, our study focused on surveillance applications where the acoustic detection of sounds events can help to prevent dangerous situations and to compensate some of the technical limitations inherent in video systems. First, we described the equipment and scenarios selected to generate a suitable acoustic database that permitted the evaluation of the different acoustic detectors. Then, the general detection system was described, and a detailed study of the background noises and acoustic events was presented. Furthermore, the problem of estimating the autocorrelation matrix in real noise conditions was studied. As a result, a new generalization applicable to all the energy detectors seen before was presented and evaluated, showing the improvements reached with this novel approach.

A complete set of experiments was performed with real acoustic events and background noises. First, a detailed evaluation of the EDe, EEDe, and PEEDe showed that, as expected, for the case of non-Gaussian and non-independent noise, the EEDe obtains better results in terms of PD, reaching a significant improvement of 12.5% for a PFA of 10^{-3} (compared with the EDe). Furthermore, since the EEDe does not consider independence, the PEEDe achieves the best results: a PD improvement of 8% versus the EEDe. Second, 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.

Chapter 6

Application of the ED and the MED to the acoustic scene analysis

Knowledge is a treasure,
but practice is the key to it.

Thomas Fuller

The detection of novel sounds is a challenging task, as shown in Chapter 5, but it can considerably improve the performance of further systems. In this chapter, the focus is placed on the enhancements achieved through other acoustic processing techniques when they are used in combination with a previous detection phase. Therefore, a detector is efficiently used to advertise an automated system that an event has occurred, and at the same time, to enable further processing (e.g. acoustic source localization or classification). Part of the research work presented in this chapter is the fruit borne from a cooperative exchange program of Ph.D. students from the "Grupo de Tratamiento de Señal (GTS)" of "Universitat Politècnica de València (UPV)" and the acoustic scene analysis group of "Universität Karlsruhe (TH)".

6.1 Improvement of localization algorithms by means of an adaptive energy detector approach

6.1.1 Introduction

One of the most important areas in which the acoustic scene analysis is required is in the interaction between man and machine. Appropriate situ-

ations occur in scenarios where a human cooperates with a *humanoid robot*, or is assisted by one [2]. 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 unpredictable waveforms of the signals and the presence of background noise. Thus, to resolve the two aforementioned issues, two novel localization approaches are presented in this section. They combine the information provided by an energy detector with the well-known localization method SRP-PHAT [24].

In the first case, a microphone array 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, as described in [59].

In the second case, it must be noted that for a complete acoustic scene analysis, especially for surveillance applications or interaction with a humanoid robot, it is necessary to localize and detect all types of sound events which can occur in the proximity. Basically, two types of sound sources can be 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 is required. Therefore, a novel approach using an ED in a temporal event alignment and a pre-classification is proposed [53].

The detection and the localization techniques in which the new approaches are based will be described. Subsequently, we will describe the procedure followed to combine them, before obtaining the modified localization algorithm. Finally, some results of these novel techniques are presented.

6.1.2 Current techniques

The detection problem is inversely related to the knowledge of both the signal to be detected and the characteristics of the background noise. The easiest case would be to detect known events in a stationary white Gaussian background noise environment. But when the sound sources are not completely known, designing the appropriate detector becomes more difficult [49]. In this case, energy detection can be useful to collect more information about the actual event and improve the localization step.

6.1.2.1 Adaptive energy detector

Since real audio signals are completely unpredictable, the energy detector is of special interest in detecting the events in the presence of a background noise which, in this case, is assumed to be Gaussian distributed. However, as adjacent audio samples are highly correlated, the noise signals cannot be considered to have white properties. In this case, some additional pre-processing is required and a PED is used (described in Section 2.2.3). Figure 6.1 depicts the complete energy detector procedure used in this chapter. The acoustic signal is divided into frames $\mathbf{y}$ of size N; then these observed vectors are linearly transformed ($\hat{\mathbf{R}}_{\mathbf{w}}$), so that a new white vector $\mathbf{y}_p$ is obtained. After that, the energy of the pre-whitened data is calculated ($\mathbf{E}_p$) and compared with a threshold λ fixed by the PFA.

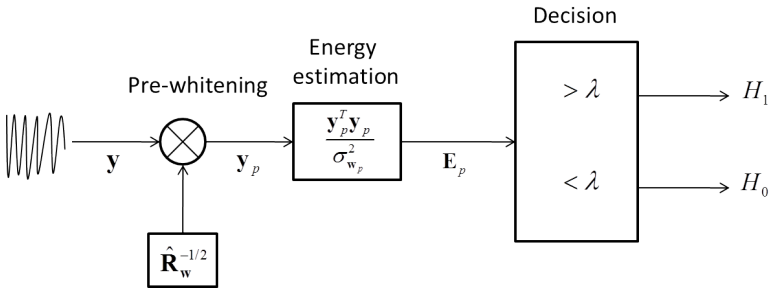


Figure 6.1: Block diagram of an energy detector.

This whitening process assumes stationarity in the background noise environment. This assumption is not always correct and could considerably reduce the performance of the energy detector when used in real scenarios where the characteristics of the noise change considerably over time. For this reason, knowledge about the noise which alters the desired sound sources at any time is important in order to dynamically adapt the estimation of $\hat{\mathbf{R}}_{\mathbf{w}}$, making our decision even more robust in the presence of background noise. The initial estimation of the noise covariance matrix must therefore be computed, and then the last K noise vectors (according to our energy detector decision), are used to reestimate $\hat{\mathbf{R}}_{\mathbf{w}}$ every τ seconds. This period of time will depend on the characteristics of the noise, particularly on its stationarity.

6.1.2.2 Localization algorithm

Nowadays, two approaches for acoustic localization are mainly used. The first is based on the estimation of the time difference of arrival (TDOA) of sound signals in a pair of spatially separated microphones. The other well-known technique for acoustic localization is the so-called Power Field (PF), also known as steered response power (SRP) [24]. They are described below.

Time delay estimation

The estimation of the TDOA of sound signals is achieved by correlating the sound signals of two paired microphones. The correlation function $R_{x_i x_j}(\tau)$ in the frequency domain can be defined as

$$R_{x_i x_j}(\tau) = \int_{-\infty}^{+\infty} X_i(\omega) X_j(\omega)^* e^{j\omega\tau}\, d\omega \tag{6.1}$$

where X_i is the Fourier transformation of the given microphone signal x_i. In theory, the signals x_i and x_j in a given pair should be an exact copy of each other, with a time delay τ. However, in real-world environments noise and reverberation effects arise. This leads to the following system model:

$$x_i(t) = h_i(t) * s_i(t) + n_i(t) \tag{6.2}$$

$$x_j(t) = h_j(t) * s_j(t) + n_j(t), \tag{6.3}$$

where $h_i(t)$ is the acoustic impulse response of the room from the source to the i^{th} microphone; the additive term $n_i(t)$ summarizes the channel noise in the microphone system as well as the environmental noise for the i^{th} sensor, and $s_j(t)$ represents a delayed version of the signal $s_i(t)$ by τ_{ij} seconds.

In order to make the correlation more stable, the so-called phase transform (PHAT) $\psi_{x_i x_j}$ is additionally used to weight the correlation function. This leads to the well-known generalized cross correlation (GCC) function [46]:

$$R^{(g)}_{x_i x_j}(\tau) = \int_{-\infty}^{+\infty} \psi^{PHAT}_{x_i x_j}(\omega) X_i(\omega) X_j(\omega)^* e^{j\omega\tau}\, d\omega, \tag{6.4}$$

with the PHAT weighting function defined by

$$\psi^{PHAT}_{x_i x_j}(\omega) = \frac{1}{|X_i(\omega) X_j(\omega)^*|}, \tag{6.5}$$

which can also be regarded as a whitening filter.

SRP-PHAT

Based on the time delay estimation, the spatial position of a sound source can be calculated. Therefore, the SRP technique can be used. In this approach, beamforming is used to focus a microphone array on a specific spatial area. To find the exact position of a sound source, the entire environment is scanned, searching for the spatial position with the highest acoustic power.

The combination of SRP and the TDOA methods mentioned before leads to a called SRP-PHAT [24], which combines the stability of the SRP against reverberation with the efficiency of the GCC method, thereby giving us the possibility to build a real-time system.

SRP-PHAT is computed as

$$A(\mathbf{v}) = \frac{1}{|M_p|} \sum_{(i,j)\in M_p} R^{(g)}_{s_i s_j}(\tau_{ij}(\mathbf{v}), \tag{6.6}$$

where $\tau_{ij}(\mathbf{s})$ denotes the theoretical delay between the microphones in pair (i, j) for the assumed spatial source position $\mathbf{v} = (v_x, v_y, v_z)$. M_p represents a given set of microphone pairs. In order to estimate the source position $\hat{\mathbf{v}}$, the position of the maximal value in $A(\mathbf{v})$ has to be found in a given search space $\mathbf{V}$:

$$\hat{\mathbf{v}} = \arg\max_{\mathbf{v}\in\mathbf{V}} A(\mathbf{v}). \tag{6.7}$$

6.1.3 Modified localization algorithms by means of an ED

Based on the ED and the localization algorithm previously described, two possible combinations of these techniques are presented in this section. The modified localization approaches that are yielded serve to improve the localization results of the standard methods. Let us now describe them in detail.

6.1.3.1 Background noise suppression

In acoustic scenarios, background noise (for example, the fan of an air conditioner or long-lasting sounds) is ubiquitous. Depending on the type of sound, the correlation of the background noise could be higher than the correlation of the desired sound source to be localized. This problem can commonly lead to a high degree of mislocalizations.

In order to improve the localization accuracy in noise conditions, common spectral subtraction techniques described in the literature could be used for the noise reduction purpose, especially in speech signal applications [8, 65]. In these scenarios, spectral subtraction method can suppress noise effectively by subtracting its spectral magnitude from that of the noisy signal. Using a voice activity detector (VAD) [14, 82], an estimate of the noise signal is measured during silence, or non-speech activity, in the signal. However, there are several aspects that must be taken into account regarding these techniques. On the one hand, since the subtraction is implemented on the signal itself, the phase could change, and consequently, the localization algorithm would be affected. On the other hand, the complexity of these methods is considerably increased in order to obtain a robust performance of the VAD under different noise environments. In addition, most of the algorithms assume that the noise is stationary, uncorrelated, additive and characterized by Gaussian distributions, which differs from real-world applications.

To avoid this, a method to suppress these background noise sources has been developed by using the adaptive ED described in Section 6.1.2.1. In

this case, the acoustic signal is kept unaltered and thus does not affect the localization algorithm. Furthermore, the simple structure of the ED employed permits adaptation to different types of noise conditions, as is shown below. Figure 6.2 shows this new approach. As can be observed, to suppress the

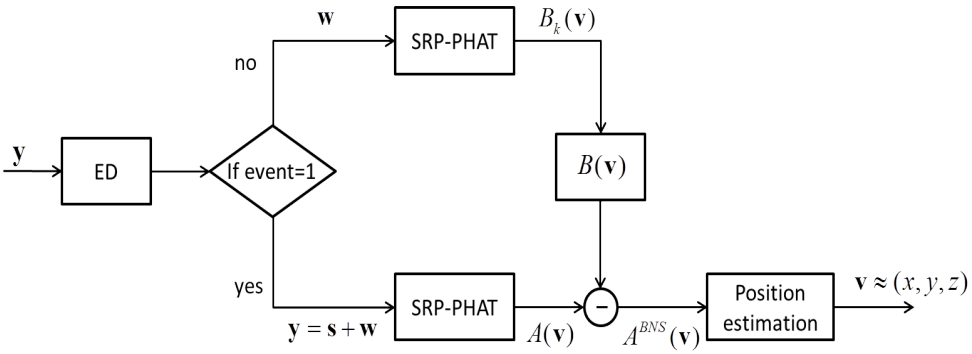


Figure 6.2: Block diagram description of the new localization approach, based on the background noise suppression.

background noise, it is first necessary to collect information about it. Using the adaptive energy detector makes it possible to distinguish between a stationary background noise source and another active source. Assuming a fully correlated noise and an ideal room with a Dirac impulse response, a noise-free estimation of the generalized cross correlation can be easily achieved in a given microphone pair by subtracting the GCC of the noise from the GCC of the received signal, as follows:

$$R^{(g)}_{s_i s_j}(\tau) = R^{(g)}_{x_i x_j}(\tau) - R^{(g)}_{n_i n_j}(\tau). \tag{6.8}$$

The estimation of $R^{(g)}_{n_i n_j}(\tau)$ for each microphone pair is achieved by calculating (6.4) during phases of no activity in the sound source, detected by using the PED. When no desired sound source is detected, our system estimates the SRP-PHAT of the current background noise $B_k(\mathbf{v})$, as indicated in (6.6), where $k = 1, \ldots, K$ and K represents the number of realizations. If a desired sound source is detected, the mean of the last K SRP-PHAT computations of the noise is computed and subtracted from the current SRP-PHAT estimation in the following manner:

$$A^{(BNS)}(\mathbf{v}) = A(\mathbf{v}) - B(\mathbf{v}) \tag{6.9}$$

where

$$B(\mathbf{v}) = \frac{1}{K} \sum_{k=1}^{K} B_k(\mathbf{v}) \tag{6.10}$$

and the resulting Power-Field $A^{(BNS)}(\mathbf{v})$ is estimated using background noise suppression (BNS). The novel localization approach yielded is termed SP-BNS. To estimate the sound source position $\hat{\mathbf{v}}$, we use the same maximum search described in (6.7).

Results

To evaluate both the performance of this combined technique and the improvement introduced in the localization phase (in comparison to the case without background noise suppression), two recording sets were tested in a kitchen scenario where people and a robot interact [59]. In the first, the performance of the presented localization system was assessed when a *speaker* was active in several room positions. To better evaluate the proposed system, a second signal source was conceived using a *toaster* as an impulsive sound source. Three kinds of typical kitchen background noise sources were studied in each setup in order to simulate different signal-to-noise ratios and to evaluate the system's ability to ignore the noise source (once localized). In the first scenario (S1), only the desired sound source (without any additional background noise) was recorded. In the second (S2), a *fan* was additionally included as a background noise. In the third and final scenario (S3), sound emitted by a kitchen *grinder* was added to the *fan* in S2. The microphone array used in our experiments was built according to the head geometry of a humanoid robot (Figure 6.3) and consisted of four omni-directional electret condenser microphones. It was roughly an inverse T-shape geometry, with a total width of 20 cm and a height of 5.5 cm.

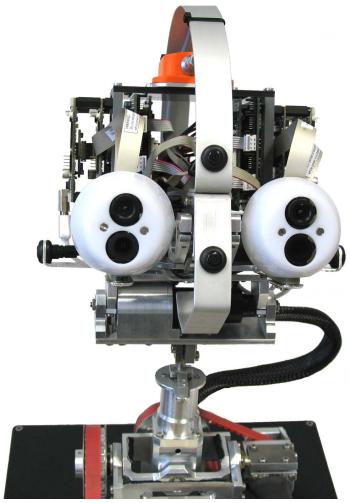


Figure 6.3: Head of the humanoid robot ARMAR III.

Table 6.1 shows the averaged results of all measurements for both sound source types, *speech* (a) and the impulsive sound source (b), using the SRP-PHAT method and the presented background noise suppression approach, SP-BNS (SRP-PHAT plus BNS). The absolute percentage of the correct localization rate is given for all combinations of active sound sources (noise and desired) in the corresponding scenario (S). In order to highlight the performance of the SP-BNS method, normalized values corresponding only to the correct localization of active sound sources are presented in brackets.

S	**Method**	***Fan***	***Grinder***	***Speech***	**Wrong**
S1	SRP-PHAT	off	off	**95.0** **(100.0)**	5.0 (–)
	SP-BNS	off	off	**93.2** **(100.0)**	6.8 (–)
S2	SRP-PHAT	19.1 (20.7)	off	**73.0** **(79.3)**	7.9 (–)
	SP-BNS	0.0 (0.0)	off	**95.4** **(100.0)**	4.6 (–)
S3	SRP-PHAT	0.0 (0.0)	83.7 (84.0)	**15.9** **(16.0)**	0.4 (–)
	SP-BNS	0.0 (0.0)	1.2 (1.4)	**88.3** **(98.6)**	10.5 (–)

(a)

S	**Method**	***Fan***	***Grinder***	***Toaster***	**Wrong**
S1	SRP-PHAT	off	off	**39.5** **(100.0)**	60.5 (–)
	SP-BNS	off	off	**38.7** **(100.0)**	61.3 (–)
S2	SRP-PHAT	23.3 (48.1)	off	**25.1** **(51.9)**	51.6 (–)
	SP-BNS	0.0 (0.0)	off	**46.6** **(100.0)**	53.4 (–)
S3	SRP-PHAT	0.2 (0.2)	40.0 (56.8)	**30.2** **(43.0)**	29.6 (–)
	SP-BNS	0.0 (0.0)	0.0 (0.0)	**35.8** **(100.0)**	64.2 (–)

(b)

Table 6.1: Percentage of averaged measurement results. (a) *Speech*. (b) Impulsive sound source (*toaster*).

In the case of *speech*, the correct localization rate was about 95% for both methods in the scenario without any background noise (S1). When using the SRP-PHAT method without BNS, we found, not surprisingly, that the localization rate for the desired source decreases dramatically for the scenarios with active background noise (S2 and S3) However, nearly 100% suppression of the background noise was achieved when the SP-BNS method

was used. This led to an improvement in the correct localization rate from 16% to 88% in S3. For the impulsive sound sources (i.e. a *toaster*), SRP-PHAT did not reach the high accuracy obtained with *speech*. This is the reason why the mislocalization rate already amounts to about 60% in the scenario, without any background noise (S1). Since the purpose of this section is to study the improvement of the acoustic localization SRP-PHAT in comparison to the novel approach (SP-BNS), the normalized values are given in brackets to avoid the influence of mislocalization on the results. Then it can be seen that, in the case of the combination of SRP-PHAT and BNS, for example, we observe an improvement in the correct localization rate from 43% to 100% for S3. In scenarios S2 and S3, background noise is totally suppressed, since there are no localizations of these noises.

6.1.3.2 Temporal event alignment

For the impulsive sound sources, SRP-PHAT does not reach the high accuracy obtained with non-impulsive events (shown in the previous section). This is the reason why even for scenarios without any background noise, the mislocalization rate is very high. To localize both impulsive and non-impulsive sound sources, the standard SRP-PHAT technique was modified (described in Section 6.1.2.2). The basic idea is to distinguish between the different types of sound sources, and adapt the localization algorithm accordingly. Therefore, in a pre-classification phase, the sound source which should be localized is classified as an impulsive or non-impulsive event. This is done by measuring the length of the event (counting detections of the energy detector) in a specific time interval. An event is handled as impulsive if the total time duration of all detections in the time interval amounts to less than one second.

Mislocalizations of impulsive events can be attributed mainly to reverberation. Because all possible reflected paths of the sound are longer than the direct path, the first wave which arrives at the microphone pair is not influenced by reverberation. In order to exploit this knowledge, it is necessary to exactly align the correlation window and the event. This is done by positioning the beginning of an event to the middle of the correlation window, using the detections of the adaptive energy detector. This alignment is done for both sound source types. In the case of an impulsive event, however, the window is additionally decreased to a quarter of its original size in order to gain more influence of the first wave. After the first localization, the localization algorithm has to handle two different event types: for the impulsive, it terminates and is waiting for the next event. For a non-impulsive event type, it waits for a half-correlation window and the pre-classification is then repeated. This whole procedure is iterated until the classification result is impulsive again. That means that the ongoing event is finished, and the algorithm is awaiting the next event.

Results

In order to evaluate the localization of impulsive and non-impulsive sound sources using this algorithm, recordings were done with and without background noise. Different signal-to-noise ratios are of particular interest because there are various noise sources which can arise, for example, in the proximity of a humanoid robot [53]. In this application, a typical case was represented by the cooling *fans* of the robot. For the evaluation, the following impulsive sound sources were analyzed: putting a *cup* on the table, opening and closing a *door*, dropping a *spoon* on a table, and a *toaster*. A *mixer* and human *speech* were used as non-impulsive sound sources. The microphone array used for our experiments was the same as that described in the previous section.

As a baseline localization method, the standard SRP-PHAT approach was utilized. The sound data were divided into windows of a specific size and with an overlap factor of 0.5. Each time an event inside of such a window was detected by the energy detector, the data was first multiplied by a Hamming window and then passed to the localization algorithm described in Section 6.1.2.2. The window size used in this case was 8192 samples and corresponded to 170 ms, leading to 11.7 localizations per second.

source	baseline method		modified method	
	correct [%]	RMS [°]	correct [%]	RMS [°]
cup	29.23	42.08	93.41	10.91
door	60.81	39.77	79.77	17.46
spoon	48.18	37.22	100.00	3.26
toaster	73.39	23.78	97.87	6.35
mixer	96.89	7.12	98.75	4.58
speech	94.98	8.90	97.78	6.28

(a)

source	baseline method		modified method	
	correct [%]	RMS [°]	correct [%]	RMS [°]
cup	30.38	42.08	93.26	6.19
door	65.43	31.29	74.97	26.06
spoon	27.28	34.28	97.78	4.02
toaster	55.64	30.53	80.22	17.94
mixer	79.44	16.04	80.22	12.65
speech	96.09	9.38	95.65	9.04

(b)

Table 6.2: Comparison between the baseline and the modified localization method based on percentage of correct localizations and the corresponding RMS (in degrees). (a) Without background noise. (b) With background noise.

In Table 6.2, it can be observed that for non-impulsive sound sources like *speech* or a *mixer*, this setup delivers high correct localization rates (over

95%) with a small root mean square error (RMS), under both conditions (i.e. with and without background noise). But the disadvantages of this setup can be clearly seen in the results for impulsive sound sources. In this case, the localization rate occasionally drops under 50% and the RMS also increases significantly. For non-impulsive sources, the modified algorithm results in a slight improvement of 1-2%. However, the localization rate and the RMS can be improved significantly for impulsive sound sources. In this case, an absolute improvement of up to 71% is attained, with a significant decrease in the RMS. For example, by using the modified localization method, the RMS decreases from 37° to 3° in the *spoon* case. Figure 6.4 highlights this fact by comparing the baseline and the modified localization method for the azimuth estimation.

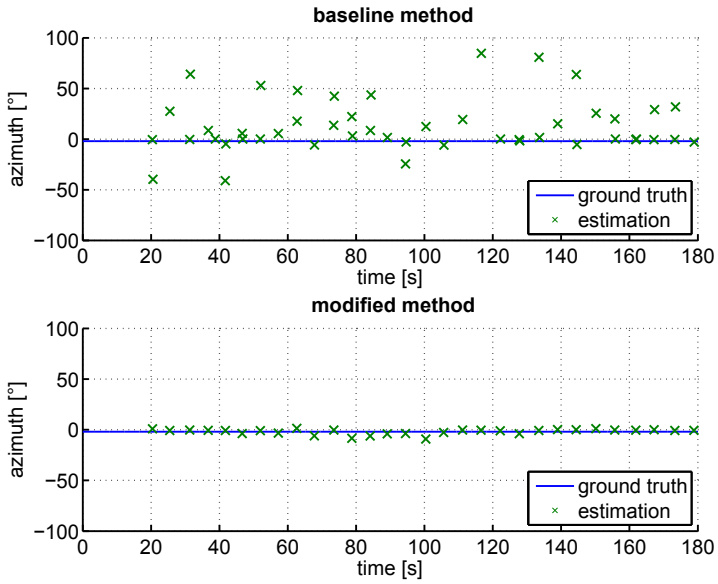


Figure 6.4: Localization azimuth results for the case of a dropping *spoon* in a noisy environment; comparison between the baseline and the modified localization method.

6.2 Acoustic classification using time and frequency MED features

6.2.1 Introduction

Within the framework of surveillance applications, the identification of possible dangerous events that might occur in the scene under analysis is of paramount importance. In that sense, the classification of sound sources

is the process of assigning an observed event to the most similar of several previously defined classes. The classes are characterized in a training stage, using sound events that are known to belong to each class. This is frequently called supervised learning [26]. Usually, this task is not performed directly with the signal, but rather with some data extracted from it, which forms the feature vector. This data is then compared with the information extracted from the training data on the basis of its statistical properties, e.g., probabilistic models [26], linear discriminants [57], support vector machines [12], etc. It is important, therefore, to identify the features that distinguish these classes.

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 [27]. MFCC usually offers good performance, but the vulnerability of these features to noise degrades their recognition performance. Thus better features are generally desired for noisy environments. In [92], for example, a noise-robust feature extraction method was presented to deal with this problem. In this section, however, we propose the use of the MED, presented in Chapter 4, to obtain a new set of features. The MED is capable of determining the presence of an acoustic event within a background noise. Furthermore, it can also provide information about it, which can then be employed for classification. On the one hand, a modified MFCC extraction method is presented; on the other, some appropriate novel features are extracted by using the signature an event produces when it is processed by the MED [61].

6.2.2 Classification method

The proposed method adopts a hierarchical classification approach to assign a label to an event in a given audio frame, as shown in Figure 6.5. First, a time and frequency MED detector is required to distinguish between an acoustic event and the background noise. In a second step, the acoustic sounds are pre-classified into impulsive and non-impulsive events, thereby allowing better classification in a later stage. Finally, further classification is performed in the third level using a Bayesian classifier to label the events.

6.2.2.1 Pre-classification

Two kinds of acoustic events have been taken into account. The first are the impulsive events, which are characterized by a short duration. In contrast, the non-impulsive events have a longer duration. Therefore, the hierarchical classification used here requires an initial first step in which both types of sounds are differentiated to better facilitate classification later on, as will be demonstrated.

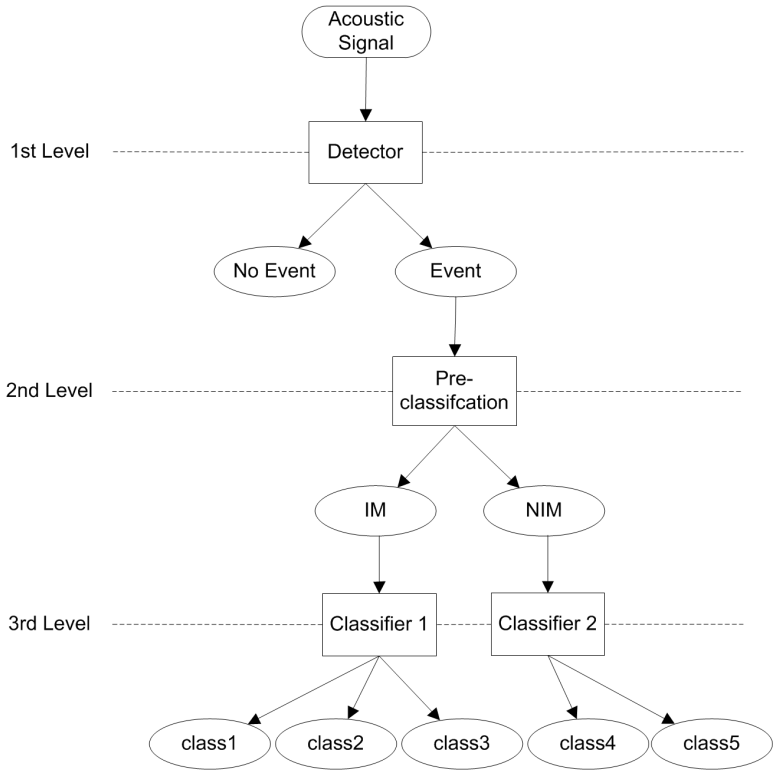


Figure 6.5: Hierarchical classification scheme.

The sound source is pre-classified as an impulsive (IM) or non-impulsive (NIM) event. This is done by measuring the length of the event: counting the number of detections in the lowest level of the temporal MED. An event is handled as impulsive if the total time duration of all detections in the time interval amounts to less than a given duration (configurable).

The importance of the pre-classifier must be pointed out, since further classification steps can take advantage of this decision. When a new event is detected, only the categories belonging to the pre-classified class will be considered, and a different parameter setup will be adjusted. Furthermore, due to the different nature of the signals, different parameter setups can be set for IM and NIM events to enhance the subsequent classification step.

6.2.2.2 Feature extraction

Prior to the recognition algorithm, an analysis is required of the signal to be classified to extract the feature vector. This is a critical step since it produces the parameters upon which the recognition algorithm is based. In this work, two types of feature are evaluated to compare the performance of

the classifier under bad SNR conditions:

1. A modified version of the well-known MFCC features.
2. The novel features extracted from the MED structure.

The MED is able to determine the presence of an acoustic event in a background noise and also provide information about it, useful for classification. Thus, a modified MFCC extraction method will first be presented. Then some appropriate novel features will be extracted using the signature an event generates when it is processed by the MED.

Modified MFCC features

MFCC features are derived from a cepstral transformation and they represent the power spectrum of an acoustic segment [19]. These features can attain high recognition accuracy in controlled environments, but the performance in real-world applications tends to degrade significantly due to the mismatch between distorted input signal and the pre-trained acoustic models. It is therefore important to exclude segments with no event information since the recognition rate can considerably decrease.

When classification follows the detection phase, it must be noted that the beginning and duration of the signal buffer used in classification are both important, especially for low SNR. The correct buffer size should be chosen to fit the event time duration, but as signals of different time durations can arise in real-world applications, selecting this value is not an easy task. Setting the buffer size to the long term events will result in a considerable decrease in the recognition rate of IM events as noise will be also present in the buffer. On the contrary, setting the buffer size to the short term events will generate a loss of information when classifying the NIM events.

To address this problem, a robust approach is used to extract features only from active segments that contain the acoustic event. Thus, taking advantage of the temporal MED structure, only the active detection segments in the lowest level will be considered in extracting the MFCC features for the classification phase (although other levels of the structure could be also used). Then, once the activated parts of the event are selected, the resulting signal is divided into frames of 21 ms and then characterized by a 13-dimensional MFCC vector. The first component of the feature vector corresponds to the DC component and is not used for classification [52].

MED features

The information provided by the MED structures can be used not only for detecting new sound sources but also for classifying them. Therefore, novel features are extracted from the time as well as the frequency structures of the

MED. These features are based on energy and offer the advantage of being noise-independent, as the MED continuously adapts to changes in noise over time. They are calculated in the following way:

$$H(l) = \sum_{m=1}^{2^{l-1}} u(l,m); \quad \forall l = 1,\ldots,L-1 \tag{6.11}$$

$$V(m) = \sum_{l=1}^{L} u\left(l, \left\lceil \tfrac{m}{2^{L-l}} \right\rceil\right); \quad \forall m = 1,\ldots,2^{l} \tag{6.12}$$

where $H(l)$ and $V(m)$, are the horizontal and vertical energy distribution of the event in the detection structure, respectively.

Afterwards, the coefficients $H(l)$ and $V(m)$ are concatenated into a single feature vector for each MED structure. PCA is applied to reduce the dimensionality [32, 39]. The features extracted from the time domain are called TMED features, and those extracted from the frequency domain are denoted FMED features. Similarly, the combination of these two feature sets through simple concatenation of both vectors and subsequent application of PCA, leads to a new set, referenced as combined multiple energy detector (CMED) features. This feature extraction method requires much less time than the MFCC features since only sums in the two dimensions of the MED are required. Furthermore, depending on the pre-classification result, the size of the TMED and FMED feature vector can be adjusted for impulsive and non-impulsive sources.

In Figure 6.6, the vertical (V) and horizontal (H) TMED and FMED features of different sound sources have been plotted for a SNR of 20 dB. We can observe how IM events (*clap*, *hammer blow*, and *shot*) have similar TMED H and FMED H feature vectors, but highly differentiated FMED V and TMED V feature vectors. This observation enables their correct identification. In addition, the *breaking glass* event presents a particular TMED V feature vector that distinguishes it from the other events. With regards to the NIM events (*metallic sounds*, *siren*, and *speech*), note that all have similar features vectors except for those related to the FMED V; they have different bandwidths, but similar time durations.

6.2.2.3 Acoustic event modeling

The acoustic events are distinguished on the basis of their specific feature vectors (described above) and by using a Bayesian classifier [26]. Using a supervised approach for the training phase, a parametric model is calculated for each class. Then, a Bayes decision rule is applied to assign one of the pre-trained classes to each new sound source.

Assuming K classes (noted c_i, $i = 1,\ldots,K$), the posterior probability that a feature vector $\mathbf{x}$ belongs to a certain class c_i can be calculated using

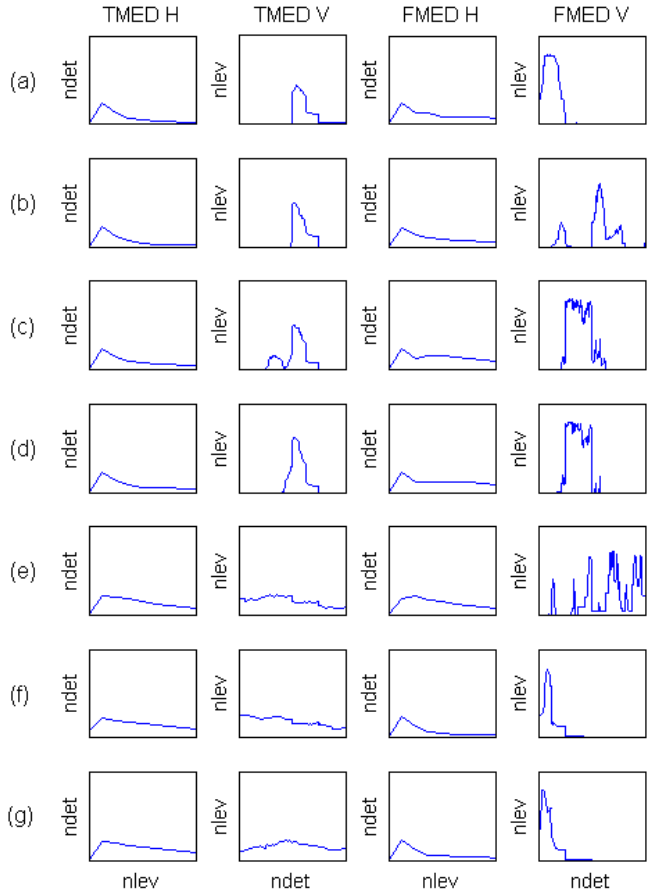


Figure 6.6: Vertical (V) and horizontal (H) TMED and FMED features for $SNR = 20$ dB. For H features, *ndet* is the number of activated detectors for each level; for the V features, *nlev* is the number of activated levels for each detector. (a) *Clap*. (b) *Hammer blow*. (c) *Breaking glass*. (d) *Shot* (*bursting balloon*). (e) *Metallic sounds* (*keys*). (f) *Siren*. (g) *Speech*.

the Bayes theorem:

$$P(c_i|\mathbf{x}) = \frac{p(\mathbf{x}|c_i)P(c_i)}{\sum_{j=1}^{K} p(\mathbf{x}|c_j)P(c_j)} \tag{6.13}$$

where $P(c_i)$ is the class prior to c_i. Then, the acoustic event is assigned to the class c_i that satisfies

$$i = arg\left\{\max_i P(c_i|\mathbf{x})\right\}. \tag{6.14}$$

In this work, the prior classes $P(c_i)$ are assumed to be equal for all classes since there is no knowledge about the event occurrence.

The feature vectors have been modeled using a multivariate Gaussian probability density function, where the sample mean and the sample covariance matrix are the only parameters calculated for every class during the training process:

$$p(\mathbf{x}|c_i) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^T C_i^{-1}(\mathbf{x}-\mu_i)\right]}{(2\pi)^{\frac{D}{2}}\sqrt{det(C_i)}} \tag{6.15}$$

where μ_i and C_i are respectively the sample mean vector and sample covariance matrix of the class c_i, estimated during the training process.

6.2.3 Evaluation results

Acoustic events of varying natures and durations that can be indicative of dangerous situations in surveillance applications, were recorded. Impulsive sound sources such as *hammer blows*, *claps*, *breaking glass*, and simulated *shots* (*bursting balloons*) were generated. Additionally, *human speech*, *sirens* and *metallic sounds* (e.g. *keys*) were also analyzed as non-impulsive events. Furthermore, different SNR were performed by adding correlated Gaussian noise.

Real recordings were carried out in a typical office room using a multichannel audio data acquisition unit with a sampling frequency of 24 kHz (as described in Section 5.1.2). Approximately 3 minutes of data were acquired for each sound source and each room position, cumulating in a total of 600 acoustic events for each SNR. The parameter setting, such as the number of levels of the MED structure (L) or the observation vector length (N), is influenced by the computational requirements fixed by the real-time processing of the signals. Taking into account these considerations and the sample frequency previously selected, a MED structure of 9 levels was implemented with a total duration of 5.46 seconds in the highest level and an energy detector of 256 samples ($\approx$10 ms) in the lowest. The PFA was set to 10^{-8}, since this is an acceptable value used in detection systems.

To calculate the MFCC features, the signal was divided into frames of 21 ms and characterized by a 13-dimensional vector. The first element of the vector corresponded to the DC component, and it was not used for classification. The number of features used for TMED and FMED depended on the pre-classification result. After applying PCA, a 14-dimensional and a 40-dimensional vector for the impulsive and the non-impulsive events are used, respectively.

Several experiments were carried out in order to test the performance of the detection and classification of acoustic events in the presence of simulated Gaussian noise [61]. The database described in the experimental setup

section was utilized and the results of each phase are presented below for several SNR conditions.

6.2.3.1 Event detection

Table 6.3 shows the PD for impulsive and non-impulsive sounds. The detections obtained for the SNR (equal to 20 dB) are used as ground-truth and, as expected, it can be observed how the PD decreases with SNR. Furthermore, there are some impulsive events, like *shot* and *glass*, which are more robust against the noise and present a high PD even in low SNR. However, for the non-impulsive events, the PD decreases considerably since the beginning and ending of the events are masked by the noise and are not detected.

source	**SNR(dB)**						
	10	0	-2	-4	-6	-8	-10
clap	99.3	99.3	99.3	98.1	90.1	82.1	72.9
hammer	98.1	94.6	93.3	92.4	90.9	84.6	74.6
glass	100	100	100	100	100	100	100
shot	99.4	99.4	99.4	99.4	99.4	99.4	99.4
metal	95.3	93.6	93.1	92.4	92.7	87.7	80.2
siren	90.3	83.8	78.1	73.3	64.9	52.1	36.9
speech	94.8	92.7	90.4	89.1	84.8	77.2	65.9

Table 6.3: PD(%) of impulsive and non-impulsive events in several SNR conditions.

6.2.3.2 Pre-classifier

The results of the pre-classifier are presented in Table 6.4, where the percentage of the correct classification for impulsive and non-impulsive events is shown. It must be noticed that it functions better in IM events, since the tails of the NIM sounds can be misclassified. This fact can be appreciated particularly in low SNR, where there are fewer detections. Therefore, the pre-classifier is more likely to decide IM.

source	**SNR(dB)**					
	20	10	0	-4	-8	-10
IM	97.6	97.6	97.5	97.5	99.1	99.0
NIM	89.9	90.4	89.8	87.2	84.1	80.7

Table 6.4: Pre-classification results (%) of impulsive and non-impulsive events in several SNR conditions.

6.2.3.3 Feature evaluation

In this section, the classification results are presented assuming that the detection and the pre-classification stages are correct. A comparison between four audio features sets (MFCC, TMED, FMED and CMED) are evaluated on their ability to differentiate acoustic events. Features extracted from detections in 20 dB are used in the training step, while the other SNRs are used for testing.

Figure 6.7 shows the results achieved when classifying the IM and NIM sounds using the different feature sets. In both cases, the results show that the TMED features present the poorest performance, especially for the non-impulsive events. This can be explained by the fact that these pyramid features only consider the temporal information of the pyramid, which makes more difficult to distinguish between events of the same time duration. Therefore, these features are not very robust against events of the same nature and they are greatly affected by the noise conditions. This

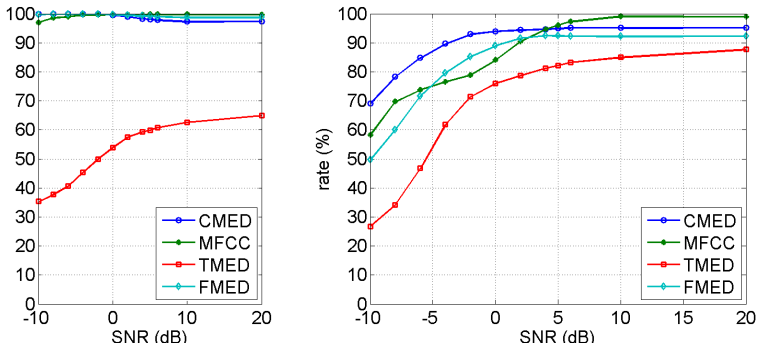


Figure 6.7: Probability of classification for several SNR conditions, $L = 9$, $N = 256$, and $PFA = 10^{-8}$. (a) NIM sounds. (b) IM sounds.

fact can be observed in Figure 6.7a: the recognition rates obtained for the non-impulsive events using FMED, CMED and MFCC are above 96% in all SNR, but the TMED features present the worst results. For the impulsive sounds, it can also be observed in Figure 6.7b that FMED performs better than TMED, and , they are complementary in the sense that they can be combined (CMED) to reach an even better classification rate. Comparing these new features against MFCC, we can see that MFCC behaves slightly better in SNRs higher than 5 dB; but as the noise conditions become worse, its performance decreases considerably. In this case, the CMED features reach better recognition rates, and an improvement up to 15% can be obtained. Therefore, it can be seen that they are more robust against the noise conditions.

We must take into account that the results shown are obtained using the modified MFCC features that use the information provided by the temporal MED structure. So, features are extracted from the pieces of signal corresponding to active detectors in the lowest level of the MED structure (modified MFCC). Therefore, the classification performance of these features is also compared with the ones extracted from the entire buffer in which the signal is present. The improvement obtained in classification by employing the modified MFCC (versus the standard ones) is shown in Table 6.5, for impulsive and non-impulsive events. The results show that the classification rates are very similar in the case of non-impulsive events. In this kind of detection there are usually a considerable number of active detections in the MED, and a substantial part of the buffer is involved in the classification. Meanwhile, for impulsive events, there is a big improvement when using the modified MFCC. There are numerous inactive detectors (signal is mainly noise) which can corrupt the classification if the standard ones are used. Therefore, although MFCC features have been usually used in the context of non-impulsive events, we have demonstrated that it is also possible to extend its use to the classification of impulsive events when using the additional information provided by the MED.

source	**SNR(dB)**					
	20	10	0	-4	-8	-10
IM	0.29	5.21	28.5	23.4	22.1	13.6
NIM	0.38	-0.04	-0.12	-0.21	-0.77	0.40

Table 6.5: Percentage (%) of improvement of the classification rate of the MFCC features after using the information provided by the temporal pyramid.

6.3 Application in surveillance scenarios

6.3.1 General system structure

This section presents the architecture, the main functionalities, and the issues related to dependability within a security system, based on the acoustic scene analysis. The system is able to detect impulsive as well as non-impulsive sounds in the presence of a background noise and then extract more information about it in order to classify and localize the source.

The system is highly heterogeneous in terms not only of the detection, classification, and localization technologies involved, but also in the programming architecture. This maintains its capacity to communicate with other systems and to provide them the information extracted from the acoustic signal processing. The design of the system has to deal with the following

types of sounds that can be indicative of dangerous situations: sounds generated by *breaking glass*, *sirens*, *metallic knocks*, *shots* and *speech*. Moreover, the system must be able to adapt to the changes in background noise as the scenario conditions can vary.

The surveillance prototype architecture is depicted in Figure 6.8. As we can observe, the system is composed of five main modules: the acquisition module, which is responsible of adapting the acoustic signal recorded by the microphones to the rest of the processing modules; the initialization module that will set all the internal and external parameters required for the correct functioning of the algorithms used by the other modules; and finally, the detection, classification, and localization modules, which are the processing units responsible of providing the results, and which will be described in more detail below.

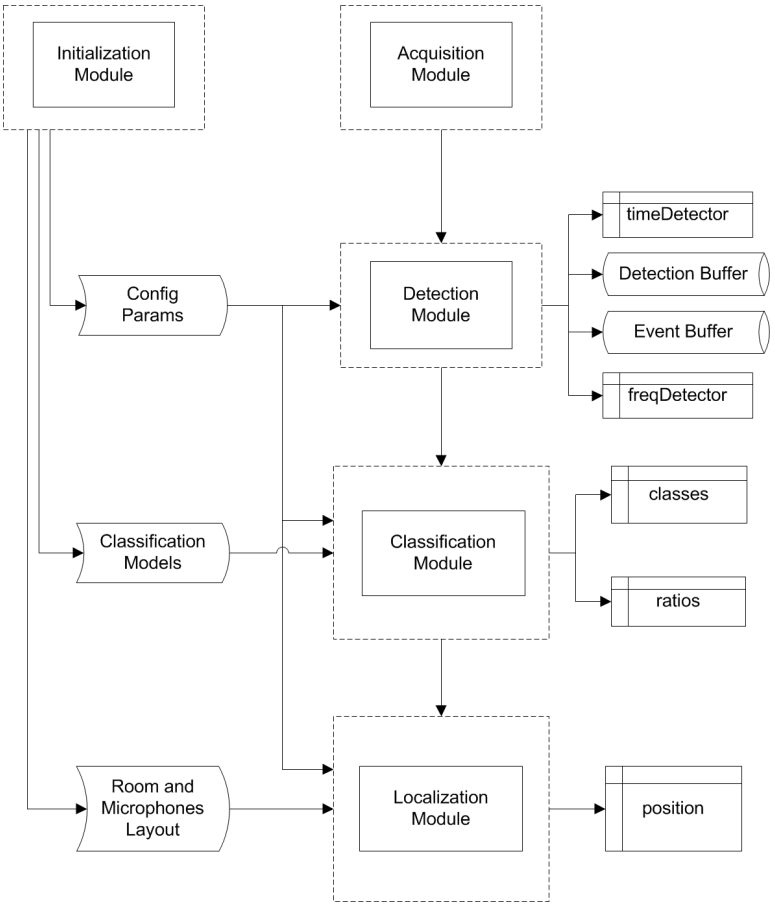


Figure 6.8: Block diagram of the entire system.

6.3.1.1 Detection module

The detection of impulsive and non-impulsive event was previously described in Chapter 5. This module can be used for security applications. Now, the most challenging task is to study the possible interaction with other modules that can provide more functionalities, and thus acquire more precise knowledge of the acoustic scene analysis. Its main objective is to detect any acoustic event in the environment and, in this case, it is responsible for activating the other two modules. That means that in case an acoustic event is detected, other processing stages (such as the classification and localization algorithms) will start collecting more information about it in order to decide which type of event it is and where it is located. For that reason, it is important to note that the detection results for each input frame analyzed are provided to the following modules. These detection results are:

- *detection buffer*: output of the energy detector stored over time.
- *event buffer*: output of the event manager submodule stored over time.
- *timeDetector*: structure with the time MED results.
- *freqDetector*: structure with the frequency MED results.

6.3.1.2 Classification module

Whenever the detection module finds anomalous events in the input signal, the recognition process is activated and a two stage classification process is initiated. The block diagram of this module is shown in Figure 6.9. First, in case a detection flag is activated in the detection module, a prior analysis of the signal is performed in order to pre-classify the detected sound and to determine whether the event is impulsive or non-impulsive. This procedure is implemented for each of the signals acquired by a single microphone, which uses the results of the corresponding detection module to determine the class. Then a combination of the results for all microphones is performed in order to obtain a more reliable decision. In a second classification step, and taking advantage of the pre-classification result and the detection structures of multiple energy detectors performed in the detection module, the class of the detected signal is determined after comparing it with different sound models, trained from a database (see Section 5.2). Again, the classification result is obtained for each microphone, and a final combined decision of the current event is provided. The mean of the probability for each class is obtained (accounting for all the microphones) and then the class with the maximum probability from among all of them (*classSecondLevel*) is selected. In addition, the ratios (*ratios2ndLevel*) representing the similarity of the event with all the classes available are also provided.

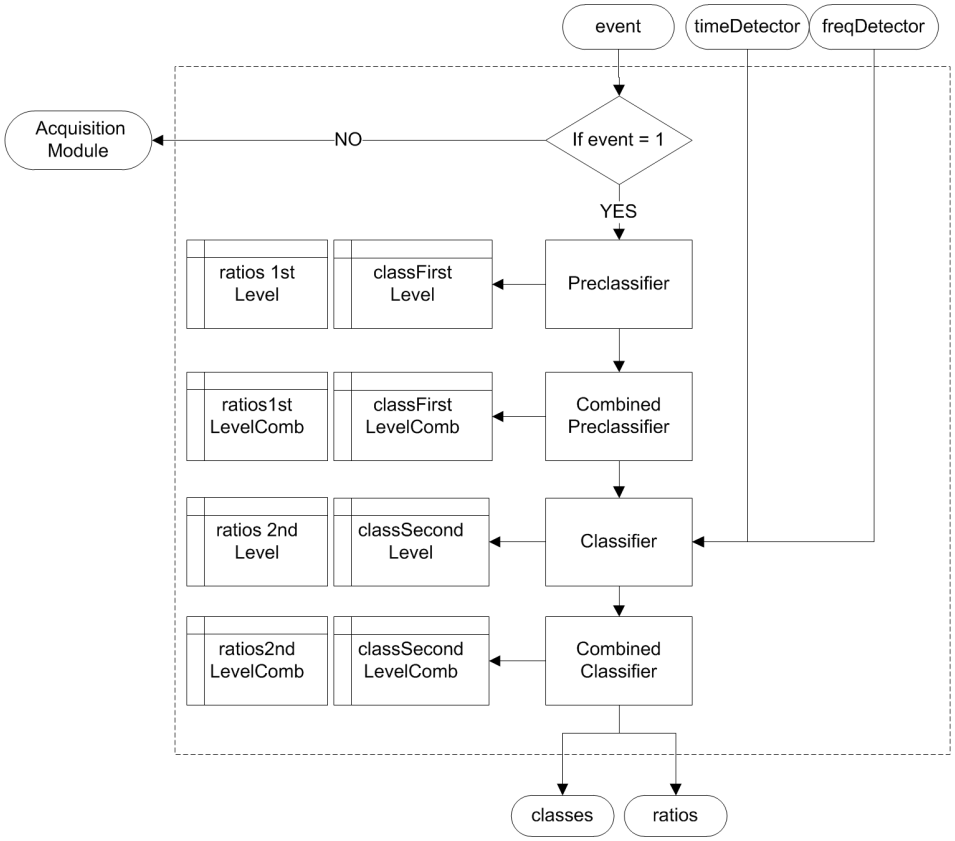


Figure 6.9: Block diagram description of the classification module.

6.3.1.3 Localization module

The final objective of sound localization in most surveillance systems consists of localizing the acoustic sound source position. Figure 6.10 depicts the block diagram of the localization implemented. It can be observed how, in addition to some configuration parameters, it is necessary to provide the event detection flag (indicating the presence of an event), as well as the prior classification results obtained in the classification module. These inputs are required since the particularity of this localization approach is mainly concentrated in the additional two functionalities included so as to deal with the localization of sounds in noisy environments (background noise suppression) and the localization of sounds of different nature (time event alignment). In the first case, the detection result allows the localization to start only when a sound occurs that is different from the noise. At the same time, this result collects localization information about the background noise in case no acoustic event occurs. In the second case, the pre-classification results provided by the classification module allows us to apply a different param-

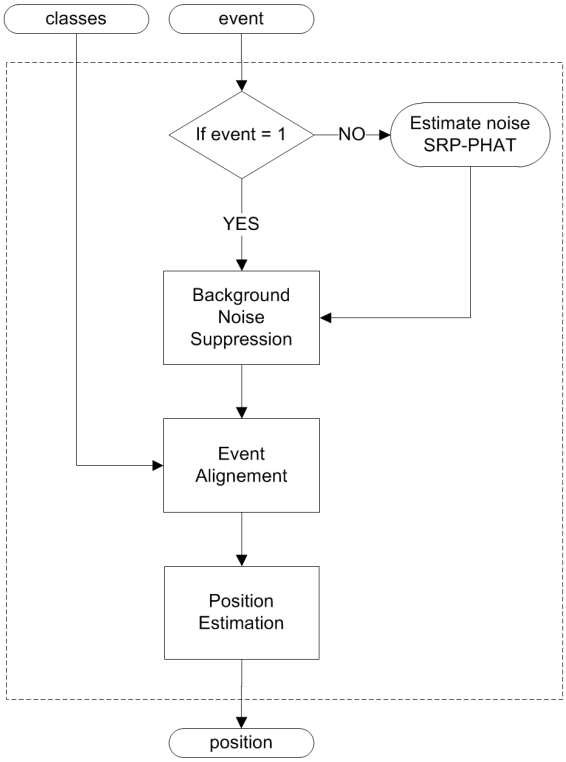


Figure 6.10: Block diagram description of the localization module.

eter setup and alignment procedure, depending of the nature of the event detected. The localization of impulsive and non-impulsive events is thereby improved. As a result, the position of the sound is yielded.

6.3.2 Graphical user interface

As the different algorithms used in the acoustic surveillance system have now been described, the graphical user interface (GUI) we developed is presented. It uses a combination of all those technologies providing a platform that the user can interact with. This tool allows us to present a summarized representation of the information and results from the different modules that make up the system. By using graphical icons and visual indicators (colored detection structures, column bars, superposition of results, and typed command labels), the tool presents the user with the information and the available actions: for example, configuring the parameter settings, performed by manipulating the graphical elements.

Designing the visual appearance and the behavior of the GUI included an important element of software programming. There were many types of programming languages that could be utilized to develop the GUI; but since

the algorithms were mainly implemented and tested using Matlab, the GUI programming environment from this program was selected [81]. Thus, it was possible to more easily adapt all the algorithms previously developed.

Let us now present the GUI of the acoustic surveillance system. The design of the tool takes into account the three main modules of the system and it is organized following this structure. Each of the modules has its corresponding screen on the GUI. In addition, a parameter setting is introduced to initialize and configure all of them. The screens that make up the application are as follows:

- Parameter setting.
- Detection results.
- Classification of each microphone and fusion decision of all rates.
- Combination of localization and classification results.

6.3.2.1 Parameter configuration

The first step is to configure all the parameters of the different modules involved in the system. Figure 6.11 shows the parameter setting screen. The parameters are grouped into different blocks, each of them responsible for initializing a particular module of the overall system. The main groups are listed here:

- **Acquisition parameters**: these parameters are used to initialize the selected audio source to be processed. There are two possible sources: the audio acquisition card or an audio file. Some of the most important parameters are the number of microphones used, the sample frequency, and the frame size.
- **Detection parameters**: in this case, these parameters are responsible for initializing the detection module. To do so, it is necessary to specify the PFA; the frame size used (which can differ from that used in the acquisition data phase) and the minimum resolution that is going to be considered (in ms) to detect an event.
- **Localization parameters**: the localization algorithm requires the specification of some internal parameters involved in the algorithm. At the same time, and in order to determine the exact position of the event, it is necessary to provide the exact microphone array location, as well as the area in which the localization of sound events is going to take place.
- **Classification parameters**: since the classification requires some predefined trained classes to compare the current events detected, it

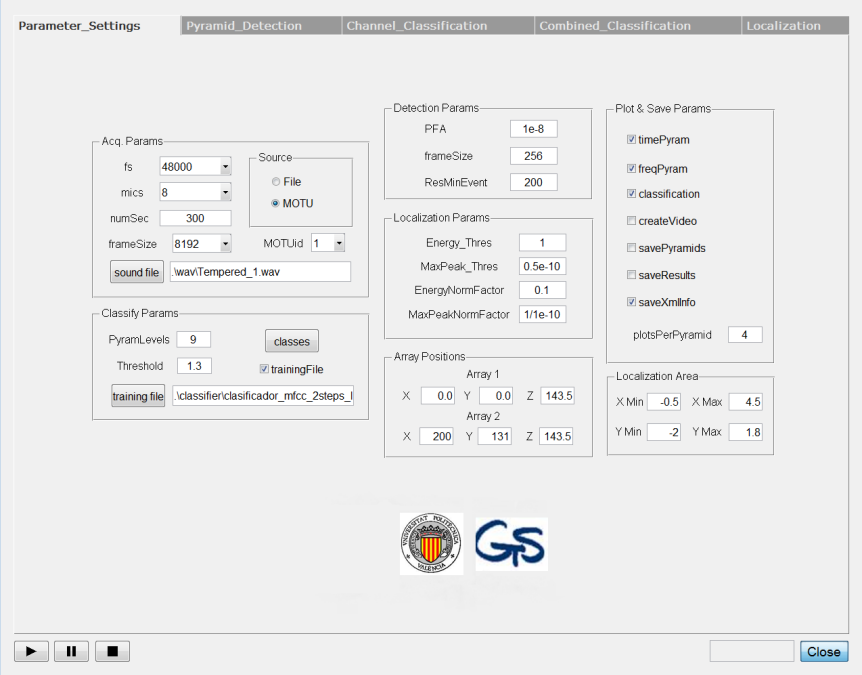


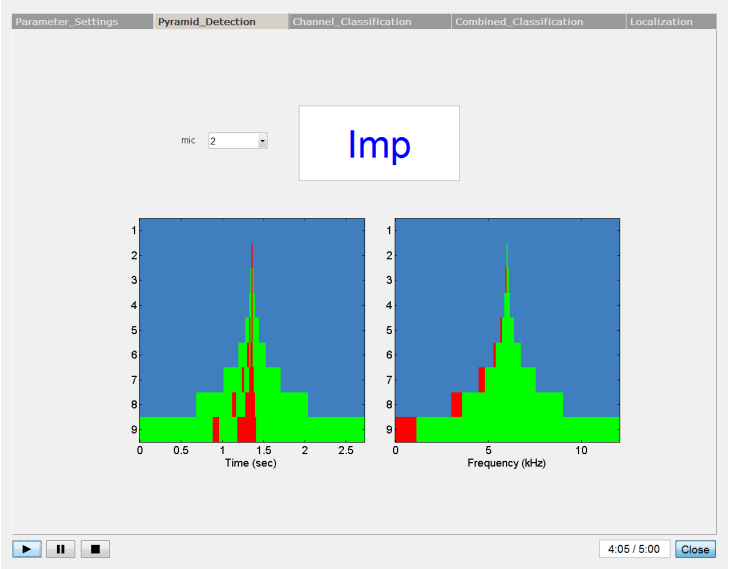
Figure 6.11: Menu of parameter settings from the GUI.

is necessary to specify the training file containing such classes. The number of MED layers must also be specified, since the classification phase extracts various novel features from this structure.

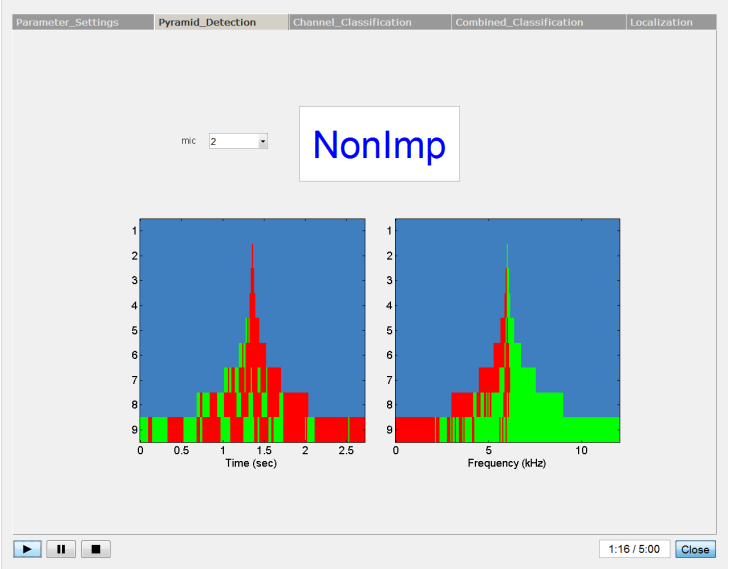
- **Plot and save parameters**: in order to select the diagrams and graphic results shown, there is a check-box menu that permits the use to represent the results of the different modules, as well as store the main results from the acoustic processing in a data file.

6.3.2.2 Detection results

Figure 6.12 displays the detection results from a particular microphone for two different acoustic events. In both cases, the detection results are represented using a MED structure of 9 layers in the time and frequency domains. Each layer is determined by a set of colored detectors: green (no detection is activated), and red (the energy of this detector exceeds the corresponding threshold). Due to the particular distribution of the detectors, the MED has a pyramidal shape, as can be observed. In addition, the pre-classification result is also shown. On the one hand, an impulsive event characterized by its short time duration and low bandwidth is detected by the MED (Figure 6.12a). On the other hand, the detection results of a non-impulsive event are depicted (Figure 6.12b). In this case, the time MED has many



(a)



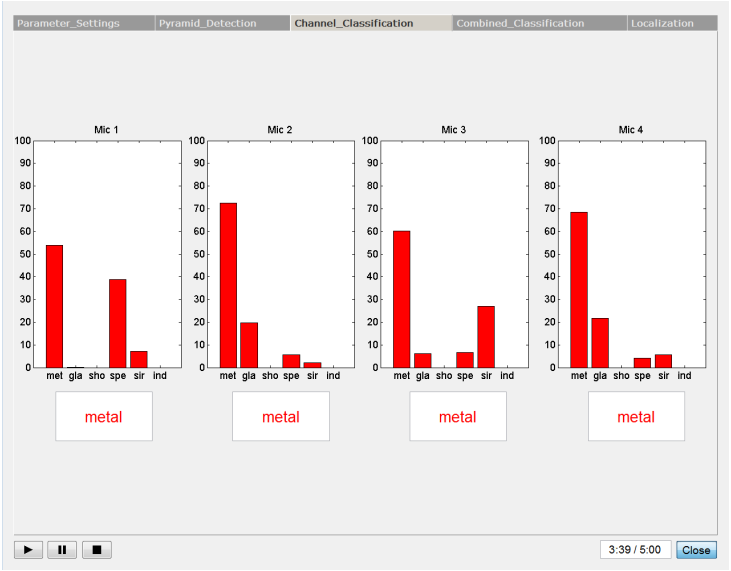
(b)

Figure 6.12: Detection MED structure and pre-classification results for two sound sources. (a) Impulsive sound. (b) Non-impulsive sound.

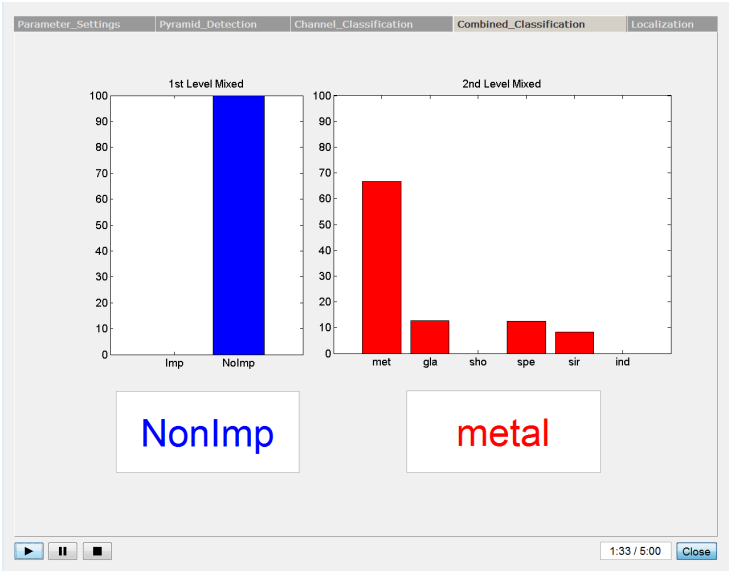
more detections throughout the entire structure. The frequency structure also reveals a wider bandwidth from the sound event than in the impulsive case.

6.3.2.3 Classification results

Due to the fact that the system is equipped with a microphone array, the redundancy of acquired audio signals can be exploited in order to improve the final classification decision. To do so, the classification algorithm is first applied separately to each channel.



(a)



(b)

Figure 6.13: Probability of classification. (a) Independent channel classification. (b) Combined classification of all microphones.

The resulting classification rates of each microphone are shown in Figure 6.13a, where the probability that the current acoustic event detected belongs to one of the pre-defined classes is depicted. As can be observed, the correct class decision is not so clear in all microphones since there are many factors that can influence each particular microphone (e.g. distance, noise level, etc.). Therefore, a channel combination approach provides a more reliable final decision. In Figure 6.13b, the global classification decision, calculated by combining the classification results from all channel-based classifiers, is shown for both the pre-classification step and the Bayes classifier.

6.3.2.4 Combination of localization and classification results

The final GUI screen displays the localization of the acoustic events in combination with the classification results. As can be observed in Figure 6.14, the result of the channel combination used for the classification of the detected acoustic event is plotted at the top. Then, in the bottom left-hand side the

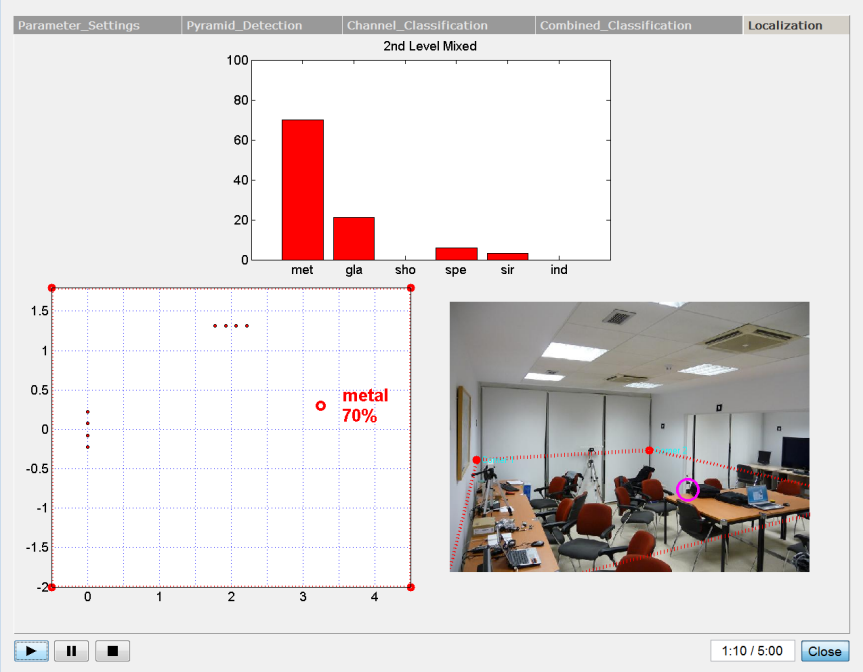


Figure 6.14: Block diagram description of the localization module.

area under study and its dimensions in combination with the exact location of the microphone arrays are represented. When an event is detected, its exact position in the room it is marked with a red circle, and the classification result is given. Finally, in the bottom right-hand side the real time image it

is shown, acquired with a video camera installed in the opposite corner where the microphones are placed. In this image, the limits of the area in which the localization is considered are overlayed. In addition, the exact position of the detected sound event is localized in the real image with a pink circle, using a matrix transformation that converts the x, y and z coordinates into the real image. This fact allows us to observe the area under surveillance in real time, and at the same time, to compare the position given by the localization algorithm with the actual location of the acoustic source.

6.4 Conclusions

This chapter has presented the promising results achieved by combining the localization and classification of acoustic sounds with energy detectors. Two main acoustic applications were evaluated which combined these techniques:

- First, the well-known SRP-PHAT localization algorithm was presented, and two modifications of this technique were proposed. The first was devoted to solving the problem of mis-localizations caused by the presence of background noise. In this case, an ED is used to collect the information about the background noise correlation to then substract it when a desired sound source is detected and must be localized. The localization of non-impulsive sound sources showed very stable results (achieving an improvement up to 72% with speech data), while the impulsive sound sources were not localized reliably. The second modification faced the problem of localizing unpredictable waveforms, which can profoundly influence localization accuracy. By using the information provided by an ED over time, it was demonstrated how it is possible to pre-classify the acoustic events into impulsive and non-impulsive. This enables a temporal alignment of the event and an adaptation of some parameters of the localization algorithm. In both cases, acoustic events extracted from a real kitchen scenario (where a man interacted with a humanoid robot) were considered. This permitted an evaluation of the two modifications and the improvement reached with the new techniques in comparison with the standard SRP-PHAT algorithm. In so doing, an absolute improvement of the localization accuracy of up to 71% could be achieved.

- Then, a Bayesian classifier, and features extracted from possible acoustic signals present in surveillance scenarios, were studied in order to characterize the sound sources present in the acoustic scene. In this case, a novel approach based on the utilization of a MED structure for detection and for extracting new features, was presented. By using the time and frequency MED structure, a new set of features were thereby extracted (TMED and FMED) and compared with the stan-

dard MFCC. Real experiments with acoustic signals of varying natures (impulsive and non-impulsive) were conducted. It was shown how the combination of TMED and FMED, named CMED, offered the best classification rates, with an improvement of up to 15% (in low SNR for the impulsive sounds) in comparison to the MFCC. No significant improvements were achieved for the non-impulsive ones.

Part IV

Conclusions

Chapter 7

Conclusions and future work

When one door closes,
another one opens.

Miguel de Cervantes

In closing this thesis, this chapter presents the conclusions reached in this work, summarizing the most relevant aspects concerning two principal points: the theoretical extensions proposed for the energy detectors and the results obtained from the simulations and experiments performed in real-world applications with acoustic signals. Taking into consideration these concluding remarks, future lines of research arise and are addressed to promote continued investigation into new techniques for fusion decision in multiple detectors and its application to other disciplines.

7.1 Conclusions

In this thesis, several strategies for designing detectors under different signal and noise conditions were presented. It was shown how the degree of difficulty of a detector is inversely related to the degree of knowledge about the signal to be detected and the background noise in which it is present.

We first reviewed the detection problem when the signal was deterministic, leading to the matched filter (MF) detector. Our attention was mainly focused on those cases where the signal was entirely unknown or it could be assumed to be present in a subspace. These problems were analyzed in detail in Chapter 2 and led us to the study of the energy detector (ED) and the matched subspace filter (MSF) for when the noise is Gaussian and uncorrelated. The existing dependence between the noise samples was also

examined in the energy detector, which led to the pre-processed energy detector (PED).

In Chapter 3, it was demonstrated that both the ED and the MSF degrade when the noise background is non-independent and non-Gaussian. Therefore, two extensions of the ED, denoted as extended energy detector (EED) and pre-processed extended energy detector (PEED) [64], and a generalization of the MSF, defined as generalized matched subspace filter (GMSF) [63], were presented in this chapter to deal with this situation. In both cases, independence was achieved by means of a linear matrix transformation derived from independent component analysis (ICA). In order to obtain a level of general applicability for the new proposed detectors, a non-parametric estimation of the noise PDF was used after the linear transformation. For each detector, the issue of non-Gaussianity was avoided in a distinct manner:

- In the EED, a study of the standard Rao test solution demonstrated the necessity of finding an alternative transformation to the one proposed by the Rao test. Therefore, two different transformations were analyzed: the Box-Cox transformation and a scalar non-linear function based on the Gaussianization process using a random variable with unknown PDF, which was estimated using a non-parametric approach. Various experiments were performed to compare both techniques. The non-linear function was selected for its higher performance (on the basis of the ROC curves) when given the same conditions. In this case, the generalized likelihood ratio test (GLRT) condition of the extended detector is not generally demonstrated. Thus, an SNR enhancement factor α was defined after the non-linear transformation and subsequently used as an indicator of the expected improvement to be derived from the proposed extension of the energy detectors.

- In the MSF, the non-linear transformation used to manage non-Gaussian noise conditions was inherent to the solution proposed by the Rao test, thus preserving its properties.

Several experiments were conducted with simulated examples in order to evaluate the EED in comparison to the ED, and the PEED in comparison to the PED. The results revealed improvements in performance gained when using the PEED with non-Gaussian and non-independent noise versus the ED and the PED. In a similar manner, the MSF was compared against the GMSF; a study of the resulting ROC curves showed similar performance enhancement with the generalized solution. More specifically, it was demonstrated that the degradation of the MSF in the presence of non-independent noise cannot be easily compensated for without the use of the GMSF.

In Chapter 4, an extension of the classical energy detector was proposed to deal with the scenario in which signal duration and bandwidth are un-

known. This can be especially interesting for the field of novelty detection, where specific parameters of the signal are entirely unknown [89]. Multiple energy detectors were applied to subintervals of the original observation interval, leading to a multiple energy detector (MED) structure where signal presence was decided if at least one of the EDs was in favor. We derived the corresponding probabilities of false alarm and detection as part of a particular strategy of successive half-segmentations of the original interval, thereby obtaining a layered structure of energy detectors. ROC curves were computed to show the resulting improvements in detection performance when the MED structure ($L > 1$) is employed (instead of the single ED ($L = 1$)), given that the signal duration or bandwidth is smaller than the original observation vector of length N. Some degradation is to be expected when the number of required layers is overestimated, but we showed that this factor is not significant when weighed against the expected improvements.

Chapter 5 offered a comparative study of the different detectors utilized in the framework of acoustic event detection, as part of surveillance applications. A detection system was introduced, and the performance of the ED and the corresponding extensions (EED and PEED) were evaluated for the more general case of non-Gaussian and non-independent noise. A new generalization was also proposed to deal with non-stationary noise conditions. The novel detection structure denoted by MED was evaluated using acoustic signals where the duration of the sound sources and their bandwidth were completely unknown. Subsequent data collection of real-world acoustic events and background noises was carried out whereby signals of varying natures and bandwidths were recorded. Several experiments were conducted in order to evaluate the ED and the MED under these adverse noise conditions with low SNR:

- Results from experimental testing of ED showed a considerable reduction in the number of false alarms, thereby demonstrating that the proposed generalization (EDe) performed better. Consequently, it was applied to other extensions. In real-world scenarios with non-Gaussian and non-independent noise, the study of the ROC curves revealed a loss in detectability when using the EDe; hence, the best results were obtained with the PEEDe.

- The MED was applied in both the time and frequency domain by using different observation intervals, respectively. The experimental results were illustrated with time and frequency ROC curves under adverse noise conditions with low SNR. The improvements obtained with the MED allowed us to validate the expected theoretical behavior presented in Chapter 4. In addition, as the signal duration and bandwidth are not known in advance, it was shown how this improvement can be worthwhile despite any possible degradation arising from the use of unnecessary layers of the MED structure [60].

Finally, the influence of background noise and the waveform of the acoustic signals on the localization and classification of sound sources was studied in Chapter 6. As a result, the PED and the MED were used in combination with these techniques in order to improve the localization accuracy of acoustic sounds and the classification rates:

- In the first case, an adaptive energy detector was used in combination with the common SRP-PHAT technique as part of the acoustic source localization. The adaptive PED was able to adjust to a given stationary background noise and it distinguished between the background noise and a real acoustic event. Several experiments were subsequently conducted; the resulting modified localization algorithm was evaluated and revealed considerable improvements [59]. In addition, the problem of localizing acoustic signals of different nature was addressed, and another modified localization algorithm was proposed and evaluated to enable reliable localization of both impulsive and non-impulsive sound sources. Again, based on the standard SRP-PHAT localization approach, we showed that a much higher correct localization rate, with and without background noise, can be obtained using an energy detector in the temporal alignment and the pre-classification of an event [53].

- In the second case, the detection and classification of acoustic events in noisy environments was considered using a MED structure and a two-stage classification approach, respectively. Mel-frequency cepstral coefficients (MFCC) features are typically used for the classification of sound sources. Nevertheless, a novel set of features was extracted from the time and frequency MED structure and were then compared. Some experiments with real acoustic signals were conducted, with the results demonstrating the performance of the detection structure and the importance of the pre-classifier. Furthermore, the new set of features, denoted by TMED, FMED, and CMED, were evaluated and compared with the MFCC. The experimental results showed that for non-impulsive events the TMED features offer the worst results. For the impulsive sounds, MFCC performs slightly better in good SNR conditions, but on the contrary CMED presents a significant improvement in the classification accuracy in low SNR [61].

7.2 Future lines of research

The results achieved in the thesis does not seek in any way to close lines of work, and they must be regarded as a rethinking of the problem, which allows future developments not only in the field of acoustic scene analysis, but in the field of any type of application in which a detection phase is required

in general. Thus, from the results achieved throughout this thesis the following possible future lines of research arise. With regards to the innovative extensions of the ED proposed in Chapter 3, it would be interesting to investigate other ICA algorithms that are capable to work with higher dimension vectors and that are less time consuming in obtaining i.i.d. components . In addition, other methods of Gaussianization could be tested and compared against the ones presented in this work to reduce the computation required at the moment due to the implementation of the non-parametric estimation of the noise PDF. Further investigations should consider possible extensions of the proposed MED structure that are described below:

- The case of non-Gaussian can be addressed by using extended energy detectors in each individual member of the layered structure, thereby dealing with the most general case of non-Gaussian background noise and unknown signal duration. In a similar way, non-independent noise samples could be dealt with by applying the pre-processed extended energy detector.

- Extension to 2-D domains, such as time-frequency or image processing, is straightforward if one simply considers multiple 2-D patches. Other possible partitions of the initial observation vector and various combined decisions of the time and frequency MED structures could be devised in order to improve the final detection process. Therefore, different subdivisions strategies can be attempted, but new derivations should be calculated to obtain the corresponding probabilities.

- A sequential version of the proposed detector is also possible by using a double threshold at every individual energy detector. As a result, the decision could be delayed until at least one of the energies is above the corresponding upper thresholds, or all energies are below the corresponding lower thresholds. However, derivation of the corresponding theoretical PD_{MED} and PFA_{MED} is not straightforward.

With regards to applying energy detectors with the goal of localizing and classifying sound sources within real acoustic scenarios, further research can be handled as follows:

- In relation to acoustic localization, different microphone constellations and various parameter settings could be considered in order to improve the reliability of the localization when using SRP-PHAT. In addition, a comparative evaluation should be performed between the modified SRP-PHAT algorithm and other localization algorithms. Furthermore, they could be tested by introducing a previous energy detector step. Other possible improvements are related to reducing mislocalization rates by means of a tracking algorithm.

- Focusing finally on the classification process, it was shown that the novel features of CMED behave better than the MFCC when dealing with impulsive sound sources. Therefore, future work must consider the possibility of combining both in order to improve classification reliability, In addition, this pair should also be compared with some other common acoustic features.

Furthermore, all the detection techniques could also be applied to other acoustic applications such as underwater detection, characterization of dolphins, radar and sonar detection, image segmentation, etc. In general, to all kind of application that requires a previous detection phase capable of discerning between interesting and not interesting events.

Part V

Appendix

Appendix A

Probability density functions

The detector performance depends largely on the ability to determine the Probability Density Function (PDF) of the noise samples. Therefore, two well-known PDFs, widely used and referenced throughout this work, are studied here: Gaussian and chi-squared distribution. The first is used to evaluate the different detectors with this type of noise distribution, and the second describes the distribution followed by the energy statistic in some detector cases.

A.1 Gaussian distribution

The PDF of a Gaussian random variable x (also known as normal distribution) is defined as:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]} \qquad -\infty < x < \infty \tag{A.1}$$

where μ is the mean and σ^2 the variance of x. It is denoted as $\mathcal{N}(\mu, \sigma^2)$ and we can say that $x \sim \mathcal{N}(\mu, \sigma^2)$ ($\sim$ signifiies "is distributed according to"). In Figure A.1, the representation of several Gaussian PDF for several variances and zero-mean is represented.
If $\mu = 0$, the moments are:

$$E(x^n) = \begin{cases} (n-1)\sigma^n!! & n \quad \text{even} \\ 0 & n \quad \text{odd.} \end{cases} \tag{A.2}$$

For non-zero mean:

$$E[(x+\mu)^n] = \sum_{k=0}^{n} \binom{n}{k} E(x^k)\mu^{n-k} \tag{A.3}$$

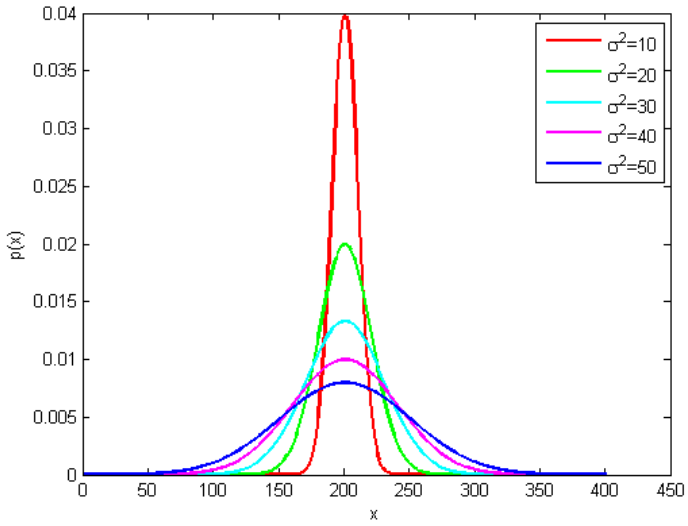


Figure A.1: Gaussian PDF for different values of variance.

where $E(x^n)$ is given by (A.2). For $\mu = 0$ and $\sigma^2 = 1$, the PDF is denoted as a normal standard PDF, and the cumulative distribution function (CDF) is defined as:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}t^2\right)} dt. \tag{A.4}$$

On the other hand, it is also possible to provide a more convenient description. Known as the right-tail probability, it is the probability to exceed a specified value and is defined as $Q(x) = 1 - \Phi(x)$, where

$$Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}t^2\right)} dt. \tag{A.5}$$

Function $Q(x)$ is also known as the complementary cumulative distribution function. Its approximation is normally employed and is given by:

$$Q(x) \approx \frac{1}{\sqrt{2\pi x}} e^{\left(-\frac{1}{2}x^2\right)}. \tag{A.6}$$

If we know that a probability is given by $P = Q(\gamma)$, then we can determine γ for a specific probability P. Thus, $\gamma = Q^{-1}(P)$, where Q^{-1} is the inverse function [43].

A.2 Non-central chi-squared distribution

A generalization of the chi-squared distribution (χ^2_ν) arises as a result of the sum of several random squared Gaussian variables which are Independent

and Identically Distributed (i.i.d.) with non-zero mean. More specifically, if $x = \sum_{i=1}^{n} x_i^2$, where samples x_i are independent and $x_i \sim N(\mu_i, 1)$, then x has a non-central chi-squared PDF with ν degrees of freedom and a non-centrality parameter denoted as $\lambda = \sum_{i=1}^{\nu} \mu_i^2$. The PDF is not obvious and it must be expressed with integrals or with finite series. It is possible to define it using integrals, as follows:

$$p(x) = \begin{cases} \frac{1}{2}(\frac{x}{\lambda})^{\frac{\nu-2}{4}} exp\left[-\frac{1}{2}(x+\lambda)\right] I_{\frac{\nu}{2}-1}(\sqrt{\lambda x}) & x > 0 \\ 0 & x < 0. \end{cases} \tag{A.7}$$

Here, $I_r(u)$ is the modified Bessel function of type one and order r, defined as:

$$I_r(u) = \frac{(\frac{1}{2}u)^r}{\sqrt{\pi}\Gamma(r+\frac{1}{2})} \int_0^{\pi} exp(u\ cos(\theta)\ sin^{2r}\theta\ d\theta) \tag{A.8}$$

and its series is given by:

$$I_r(u) = \sum_{k=0}^{\infty} \frac{(\frac{1}{2}u)^{2k+r}}{k!\Gamma(r+k+1)}. \tag{A.9}$$

This PDF tends to the Gaussian distribution when ν increases. By using the series expansion of (A.9), it can be expressed as a finite series:

$$p(x) = \frac{x^{\frac{\nu}{2}-1} exp\left[-\frac{1}{2}(x+\lambda)\right]}{2^{\frac{\nu}{2}}} \sum_{k=0}^{\infty} \frac{(\frac{\lambda x}{4})^k}{k!\Gamma(\frac{\nu}{2}+k)}. \tag{A.10}$$

Particularizing for $\lambda = 0$, the non-central chi-squared PDF is denoted as simply chi-squared. In general, the non-central chi-squared PDF, with ν degrees of freedom and non-centrality parameter λ, is denoted as $\chi'^2_\nu(\lambda)$. The mean and variance are:

$$\begin{aligned} E(x) &= \nu + \lambda \\ var(x) &= 2\nu + 4\lambda. \end{aligned} \tag{A.11}$$

It is possible to denote the right-tail probability as:

$$Q_{\chi'^2_\nu(\lambda)}(x) = \int_x^{\infty} p(t)dt \qquad x > 0. \tag{A.12}$$

In the following figures, it can be observed how the chi-squared PDF changes when varying either the degrees of freedom (Figure A.2) or the non-centrality parameter (Figure A.3).

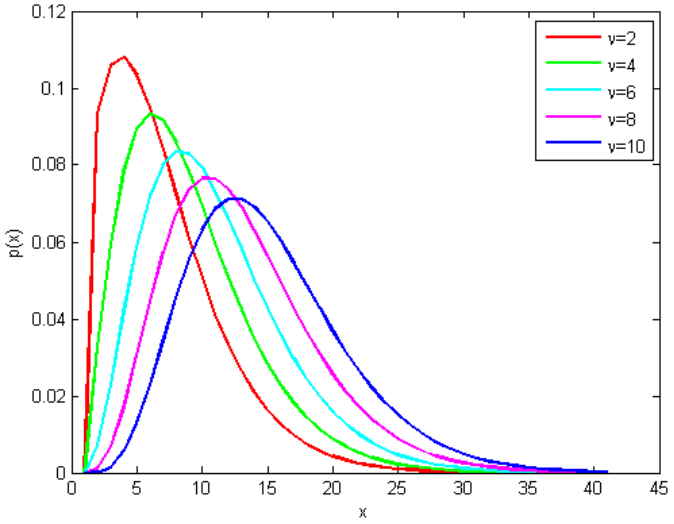


Figure A.2: Chi-squared PDF for different degrees of freedom (ν).

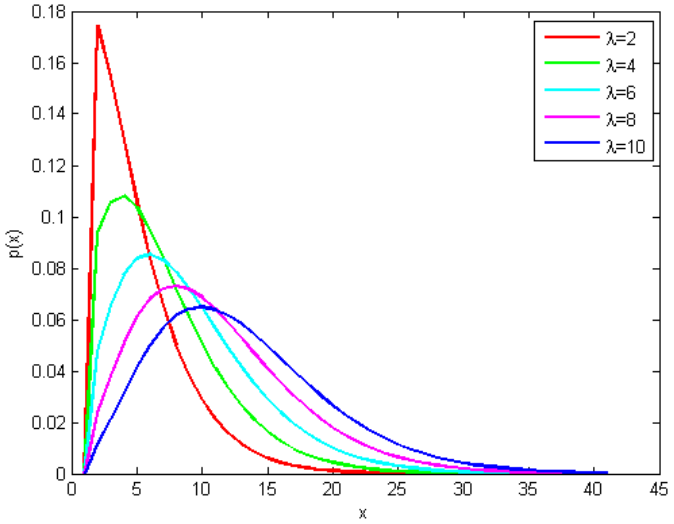


Figure A.3: Chi-squared PDF for different parameter λ.

Appendix B

Derivation of P_{lm}

In order to obtain the PFA_{MED} and the PD_{MED}, it is necessary to obtain the P_{lm}. It can be defined as the probability that the *m-th* detector at layer l decides $\mathbf{H}_0$ (conditioned to $\mathbf{H}_0$), and that the detectors at the lower layers have also decided $\mathbf{H}_0$.

Let us call E_{lm} the observed energy corresponding to the *m-th* detector at layer l. The probability P_{lm} can be defined in the form:

$$P_{lm} = P(E_{lm} < \lambda_l / H_0; \quad E_{l'm'} < \lambda_{l'}, l' > l, m' = 1, \ldots, 2^{l'}) \tag{B.1}$$

but it must be taken into account that

$$\begin{aligned} E_{lm} &= E_{l+1,2m} + E_{l+1,2m-1} \\ &= (E_{l+2,4m} + E_{l+2,4m-1}) + (E_{l+2,4m-2} + E_{l+2,4m-3}) \\ &= \ldots . \end{aligned} \tag{B.2}$$

For $l' > l$, we can then write:

$$E_{lm} = E_{l',2^{l'-l}m} + E_{l',2^{l'-l}m-1} + \ldots + E_{l',2^{l'-l}m-2^{l'-l}+1}. \tag{B.3}$$

The conditional probability P_{lm} defined in (B.1) is conditioned only to a subset of the EDs at the lower levels, defined by (B.3). This leads to

$$\begin{aligned} P_{lm} &= P(E_{lm} < \lambda_l / H_0; \quad E_{l'm'} < \lambda_{l'}, \\ &\qquad l' > l, m' = 2^{l'-l}m - k, k = 0 \ldots 2^{l'-l} + 1). \end{aligned} \tag{B.4}$$

Considering (B.3), the conditioned probability in (B.4) can be expressed in terms of E_{lm}:

$$P_{lm} = P(E_{lm} < \lambda_l / H_0; \quad E_{lm} < 2^{l'-l}\lambda_{l'}, l' > l). \tag{B.5}$$

Let us now evaluate the relationship between the threshold $\lambda_{l'}$ (used in a specific layer l' of the MED) and the threshold λ_l (corresponding to another layer l). As all EDs are operating with the same PFA, from (4.3) it is possible to derive the following equality:

$$\frac{\lambda_l - N_l}{\sqrt{2N_l}} = \frac{\lambda_{l'} - N_{l'}}{\sqrt{2N_{l'}}}. \tag{B.6}$$

Taking into account (B.6) and the existing relation between the length of the observation vector at two different layers (l and l') $N_{l'} = N_l/2^{l'-l}$, it is possible to obtain threshold λ_l as a function of $\lambda_{l'}$:

$$\lambda_l = \lambda_{l'}\sqrt{2^{l-l'}} + N_l(2^{l-l'} - \sqrt{2^{l-l'}}). \tag{B.7}$$

Subsequently, the threshold used in (4.3) to compare the estimated energy can be obtained:

$$\begin{aligned} 2^{l'-l}\lambda_{l'} &= \lambda_l\sqrt{2^{l'-l}} + N_l(1 - \sqrt{2^{l'-l}}) \\ &= (\lambda_l - N_l)\sqrt{2^{l'-l}} + N_l. \end{aligned} \tag{B.8}$$

However, it must be taken into account that for $PFA < 0.5$ (which holds for any practical implementation of the detector), $\lambda_l = N_l$ is satisfied. This is because, as was previously explained, energy E_{lm} is assumed to be a Gaussian random variable having mean N_l. Hence, the expression $2^{l'-l}\lambda_{l'}$ is an increasing function on l', and the condition in (B.5) can be expressed only for $l' = l + 1$ as follows:

$$P_{lm} = P(E_{lm} < \lambda_l/H_0; \quad E_{lm} < 2\lambda_{l+1}). \tag{B.9}$$

Note that from (B.8) the relation between the thresholds of two successive layers of the MED can be derived:

$$\begin{aligned} 2\lambda_{l+1} &= (\lambda_l - N_l)\sqrt{2} + N_l \\ &= \lambda_l + (\sqrt{2} - 1)(\lambda_l - N_l) > \lambda_l. \end{aligned} \tag{B.10}$$

Using the Bayes's rule we can now rewrite (4.7) as follows:

$$\begin{aligned} P_{lm} &= P(E_{lm} < \lambda_l/H_0; \quad E_{lm} < 2\lambda_{l+1}) \\ &= \frac{P(E_{lm} < 2\lambda_{l+1}/H_0; E_{lm} < \lambda_l)P(E_{lm} < \lambda_l/H_0)}{P(E_{lm} < 2\lambda_{l+1}/H_0)} \\ &= \frac{1 - PFA}{1 - \Delta_l} \end{aligned} \tag{B.11}$$

where

$$\Delta_l = Q\left(\frac{2\lambda_{l+1} - N_l}{\sqrt{2N_l}}\right) \qquad l < L-1. \tag{B.12}$$

The function Δ_l is dependent on λ_{l+1}, whose value can be obtained with the expression in (4.3):

$$Q^{-1}(PFA) = \frac{\lambda_{l+1} - N_{l+1}}{\sqrt{2N_{l+1}}}. \tag{B.13}$$

Its value is then derived as follows:

$$\lambda_{l+1} = \sqrt{2N_{l+1}}Q^{-1}(PFA) + N_{l+1}. \tag{B.14}$$

Substituting (B.12) and noting that $2N_{l+1} = N_l$, the following expression is obtained:

$$\Delta_l = Q(\sqrt{2} \cdot Q^{-1}(PFA)) \qquad l < L-1. \tag{B.15}$$

Finally, this expression can be used in (B.11) to obtain the P_{lm}:

$$P_{lm} = \frac{1 - PFA}{1 - Q\left(\sqrt{2} \cdot Q^{-1}(PFA)\right)} \qquad l < L-1. \tag{B.16}$$

List of Acronyms

BNS Background Noise Suppression

BSS Blind Source Separation

CMED Combined Multiple Energy Detector

DFT Discrete Fourier Transform

ED Energy Detector

EED Extended Energy Detector

FMED Frequency Multiple Energy Detector

GMM Gaussian Mixture Model

GCC Generalized Cross Correlation

GED Generalized Energy Detector

GLRT Generalized Likelihood Ratio Test

GMSF Generalized Matched Subspace Filter

GTS "Grupo de Tratamiento de Señal"

ICA Independent Component Analysis

i.i.d. Independent and Identically Distributed

IM Impulsive

MED Multiple Energy Detectors

MF Matched Filter

MFCC Mel-Frequency Cepstral Coefficient

MLE Maximum Likelihood Estimate

MSF Matched Subspace Filter

NIM Non-Impulsive

PCA Principal Component Analysis

PDF Probability Density Function

PD Probability of Detection

PED Pre-processed Energy Detector

PEED Pre-processed Extended Energy detector

PFA Probability of False Alarm

PHAT Phase Transform

ROC Receiver Operating Characteristic

SNR Signal-to-Noise Ratio

SRP Steered Response Power

SVM Support Vector Machine

TDOA Time Difference of Arrival

TMED Time Multiple Energy Detector

UMP Uniformly Most Powerful

WGN White Gaussian Noise

Bibliography

[1] A. E. T. V. Annamaria Mesaros, Toni Heittola. Acoustic event detection in real life recordings. In *Proceedings of the* 18^{th} *European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010.

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

[3] P. Atrey, N. Maddage, and M. Kankanhalli. Audio based event detection for multimedia surveillance. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 813–816, Toulose, France, 2006.

[4] K. P. Balanda and H. L. MacGillivray. Kurtosis: A critical review. *The American Statistician*, 42(2):111–119, 1988.

[5] H. B. Barlow. Unsupervised learning. *Neural computation*, 1(3):295–311, 1989.

[6] P. J. Bickel and K. A. Doksum. An analysis of transformations revisited. *Journal of the American Statistical Association*, 76(374):296–311, 1981.

[7] R. Blum, S. Kassam, and H. Poor. Distributed detection with multiple sensors i. advanced topics. *Proceedings of the IEEE*, 85(1):64–79, Jan. 1997.

[8] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(2):113–120, Apr. 1979.

[9] R. Boscolo, H. Pan, and V. Roychowdhury. Independent component analysis based on nonparametric density estimation. *IEEE Transactions on Neural Networks*, 15(1):55–65, 2004.

[10] G. Box and D. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B.General*, 26(2):211–252, 1964.

[11] A. Brutti, M. Omologo, and P. Svaizer. Comparison between different sound source localization techniques based on a real data collection. In *Hands-Free Speech Communication and Microphone Arrays, 2008. HSCMA 2008*, pages 69–72, May 2008.

[12] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2)(121-167-706), 1998.

[13] Y. Chan, Q. Yuan, H. So, and R. Inkol. Detection of stochastic signals in the frequency domain. *IEEE Transactions on Aerospace and Electronic Systems*, 37(3):978–988, Jul. 2001.

[14] J.-H. Chang, N. S. Kim, and S. Mitra. Voice activity detection based on multiple statistical models. *IEEE Transactions on Signal Processing*, 54(6):1965–1976, Jun. 2006.

[15] Y. Chen. Improved energy detector for random signals in gaussian noise. *IEEE Transactions on Wireless Communications*, 9(2):558–563, Feb. 2010.

[16] E. Cheng, M. Piccardi, and T. Jan. Stochastic boats generated acoustic target signal detection in time-frequency domain. In *Proceedings of the* 4^{th} *IEEE International Symposium on Signal Processing and Information Technology*, pages 429–432, Dec. 2004.

[17] C. Clavel, T. Ehrette, and G. Richard. Events detection for an audio-based surveillance system. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 1306–1309, Amsterdam, Netherlands, 2005.

[18] P. Comon and C. Jutten. *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Academic Press (Elsevier), Feb. 2010.

[19] S. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(4):357–366, Aug. 1980.

[20] A. De Maio and S. Iommelli. Coincidence of the rao test, wald test, and glrt in partially homogeneous environment. *IEEE Signal Processing Letters*, 15:385–388, 2008.

[21] A. De Maio, S. Kay, and A. Farina. On the invariance, coincidence, and statistical equivalence of the glrt, rao test, and wald test. *IEEE Transactions on Signal Processing*, 58(4):1967–1979, Apr. 2010.

[22] M. Desai and R. Mangoubi. Robust gaussian and non-gaussian matched subspace detection. *IEEE Transactions on Signal Processing*, 51(12):3115–3127, Dec. 2003.

[23] P. Devijver and J. Kittler. *Pattern recognition: a statistical approach.* Prentice Hall, 1982.

[24] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein. *"Robust localization in reverberant rooms"*, chapter 8, pages 157–180. Springer, Berlin, 2001.

[25] N. R. Draper and D. Cox. On distributions and their transformation to normality. *Journal of the Royal Statistical Society.Series A.General*, 31(3):472–476, 1969.

[26] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis.* Wiley Interscience, 1973.

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

[28] R. Fletcher. *Practical methods of optimization.* Wiley, 2nd, 1987.

[29] L. Gerosa, G. Valenzise, F. Tagliasacchi, F. Antonacci, and A. Sarti. Scream and gunshot detection in noisy environments. In *Proceedings of the* 15^{th} *European Signal Processing Conference (EUSIPCO)*, Poznan, Poland, Sept. 2007.

[30] J. Gosálbez, A. Salazar, I. Bosch, and J. Moragues. Noise control system for building noise areas. In *Proceedings of the* 1^{o} *Workshop en Tecnologías de audio cognitivo para aplicaciones en seguridad y acústica forense*, pages 23–26, Valencia, Spain, Jun. 2007.

[31] D. Hall and J. Llinas. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23, Jan. 1997.

[32] W. Härdle and L. Simar. *Applied multivariate statistical analysis.* Springer-Verlag, 1st edition, 2006.

[33] A. Härmä, M. McKinney, and J. Skowronek. Automatic surveillance of the acoustic activity in our living environment. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2005.

[34] R. Hippenstiel. *Detection theory: Applications and digital signal processing.* Academic Press, 2001.

[35] W. Huang, T. K. Chiew, H. Li, T. S. Kok, and J. Biswas. Scream detection for home applications. In *Proceedings of the* 5^{th} *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 2115–2120, Jun. 2010.

[36] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[37] D. N. Joanes and C. A. Gill. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society*, 47(1):183–189, 1998.

[38] J. A. John and N. R. Draper. An alternative family of transformations. *Journal of the Royal Statistical Society.Series A.General*, 29(2):190–197, 1980.

[39] I. Jolliffe. *Principal component analysis*. Springer-Verlag, 1st edition, 1986.

[40] S. A. Kassam. *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1st edition, 1988.

[41] S. Kay and J. Gabriel. An invariance property of the generalized likelihood ratio test. *IEEE Signal Processing Letters*, 10(12):352–355, Dec. 2003.

[42] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. NJ: Prentice-Hall, 1st edition, 1993.

[43] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*. NJ: Prentice-Hall, 1st edition, 1998.

[44] S. M. Kay. Exponentially embedded families - new approaches to model order estimation. *Aerospace and Electronic Systems, IEEE Transactions on*, 41(1):333–345, Jan. 2005.

[45] S. M. Kay. The multifamily likelihood ratio test for multiple signal model detection. *IEEE Signal Processing Letters*, 12(5):369–371, 2005.

[46] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. on Acoustics*, 24(4):320–327, Aug. 1976.

[47] V. Kostylev. Energy detection of a signal with random amplitude. In *Proceedings of IEEE International Conference on Communications (ICC)*, volume 3, pages 1606–1610, 2002.

[48] S. Kraut, L. Scharf, and L. McWhorter. Adaptive subspace detectors. *IEEE Transactions on Signal Processing*, 49(1):1–16, 2001.

[49] K. Kroschel. *Statistische Informationstechnik: Signal- und Mustererkennung, Parameter- und Signalschätzung.* Springer, Berlin, 4th edition, 2004.

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

[51] E. Lehmann. *Testing statistical hypothesis.* Springer, 2nd edition, 1986.

[52] B. Lilly. *Robust speech recognition in adverse environments.* Ph.D. dissertaion, Faculty of Engineering, Griffith University, Nathan Campus., 2000.

[53] T. Machmer, J. Moragues, A. Swerdlow, A. Serrano, L. Vergara, and K. Kroschel. Robust impulse sound source localization by means of an energy detector for temporal alignment and pre-classification. In *Proceedings of the* 17^{th} *European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, Aug. 2009.

[54] B. F. J. Manly. Exponential data transformations. *Journal of the Royal Statistical Society.Series A.General*, 25(1):37–42, 1976.

[55] K. Mardia and T. Kent. *Multivariate analysis.* Academic Press, 2nd edition, 2002.

[56] M. Markou and S. Sameer. Novelty detection: a review - part 1: Statistical aproaches. *Signal Processing*, 83:2481–2497, Nov. 2003.

[57] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition.* Wiley-Interscience, 2004.

[58] E. Menegatti, E. Mumolo, M. Nolich, and E. Pagello. A surveillance system based on audio and video sensory agents cooperating with a mobile robot. In *Proceedings of the* 8^{th} *International Conference on Intelligent Autonomous Systems IAS-8'04)*, pages 335–343, Amsterdam, Netherlands, Sept. 2004.

[59] J. Moragues, T. Machmer, A. Swerdlow, L. Vergara, J. Gosálbez, and K. Kroschel. Background noise suppression for acoustic localization by means of an adaptive energy detection approach. In *Proceedings of* 33^{th} *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2421–2424, Las Vegas, USA, Mar. 2008.

[60] J. Moragues, A. Serrano, L. Vergara, and Gosálbez. Improving detection of acoustic signals by means of a time and frequency multiple energy detector. *IEEE Signal Processing Letters*, 18(8):458–461, Aug. 2011.

[61] J. Moragues, A. Serrano, L. Vergara, and J. Gosálbez. Acoustic detection and classification using temporal and frequency multiple energy detector features. In *Proceedings of the* 36^{th} *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1940–1943, Praga, Czech Republic, May. 2011.

[62] J. Moragues, L. Vergara, and Gosálbez. Novelty detection based on energy detectors. In *Proceedings of the 1º Workshop en Tecnologías de audio cognitivo para aplicaciones en seguridad y acústica forense*, pages 27–30, Valencia, Spain, Jun. 2007.

[63] J. Moragues, L. Vergara, and J. Gosálbez. Generalized matched subspace filter for non-independent noise based on ica. *IEEE Transactions on Signal Processing*, 59(7):3430–3434, Jul. 2011.

[64] J. Moragues, L. Vergara, J. Gosálbez, and I. Bosch. An extended energy detector for non-gaussian and non-independent noise. *Signal Processing*, 89(4):656–661, 2009.

[65] W. Mwema and E. Mwangi. A spectral subtraction method for noise reduction in speech signals. In *IEEE AFRICON 4th*, volume 1, pages 382–385, Sept. 1996.

[66] J. Neyman and E. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 231:289–337, 1933.

[67] P. Nicolas and D. Kraus. Detection and estimation of transient signals in coloured gaussian noise. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2821–2824 vol.5, 1988.

[68] S. Ntalampiras, I. Potamitis, and N. Fakotakis. On acoustic surveillance of hazardous situations. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 165–168, Apr. 2009.

[69] J. Ogg, F. A note on bayes detection of signals. *IEEE Transactions on Information Theory*, 10(1):57–60, Jan. 1964.

[70] A. Papoulis. *Probability, random variables and stochastic processes.* MacGraw-Hill, 1st edition, 1984.

[71] L. R. Pericchi. A bayesian-approach to transformations to normality. *Biometrika*, 68(1):35–43, 1981.

[72] R. Radhakrishnan, A. Divakaran, and A. Smaragdis. Audio analysis for surveillance applications. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 158–161, Oct. 2005.

[73] J.-L. Rouas, J. Louradour, and S. Ambellouis. Audio events detection in public transport vehicle. In *Proceedings of IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 733–738, Sept. 2006.

[74] R. Sakia. The box-cox transformation technique: a review. *Journal of the Royal Statistical Society. Series D, The statistician*, pages 169–178, 1992.

[75] L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Transactions on Signal Processing*, 42(8):2146–2157, Aug. 1994.

[76] L. L. Scharf. *Statistical Signal Processing: detection, estimation, and time series analysis.* Addison Wesley, 1991.

[77] P. Schultheiss and L. Godara. Detection of weak stochastic signals in non-gaussian noise: a general result. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 4:361–364, 1994.

[78] D. Scott and S. Sain. Multidimensional density estimation. *in Handbook of Statistics*, 23: Data Mining and Computational Statistics, Aug. 2004.

[79] H. sEcurity: teconologíaS Para la sEguridad integRal en espacIos públicos e infraestructuras. Cenit-2005, 2005.

[80] M. R. Sheldom. *Introduction to probability models.* Academic Press, 9th edition, 2007.

[81] S. T. Smith. *Matlab: advanced GUI development.* Dog Ear, 1st edition, 2006.

[82] J. Sohn, N. S. Kim, and W. Sung. A statistical model-based voice activity detection. *IEEE Signal Processing Letters*, 6(1):1–3, Jan. 1999.

[83] J. V. Stone. Independent component analysis: an introduction. *Trends in cognitive sciences*, 6(2):59–64, 2002.

[84] J. W. Tukey. One degree of freedom for nonadditivity. *Biometrics*, 5(3):232–242, 1949.

[85] V. Tuzlukov. *Signal detection theory.* Birkhäser, 2001.

[86] H. Urkowitz. Energy detection of unknown deterministic signals. *Proceedings of the IEEE*, 55(4):523–531, Apr. 1967.

[87] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti. Scream and gunshot detection and localization for audio-surveillance systems. In *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 21–26, Sept. 2007.

[88] L. Vergara. On the equivalence between likelihood ratio tests and counting rules in distributed detection with correlated sensors. *Signal Processing*, 87(7):1808–1815, 2007.

[89] L. Vergara, J. Moragues, J. Gosálbez, and A. Salazar. Detection of signals of unknown duration by multiple energy detectors. *Signal Processing*, 90(2):719–726, 2010.

[90] P. Willett, P. Swaszek, and R. Blum. The good, bad and ugly: distributed detection of a known signal in dependent gaussian noise. *IEEE Transactions on Signal Processing*, 48(12):3266–3279, Dec. 2000.

[91] W. Zajdel, D. M. Gavrila, J. D. Krijnders, and T. Andringa. Cassandra: Audio-video sensor fusion for aggression detection. In *Proceedings of IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, pages 200–205, London, UK, Sept. 2007.

[92] N. Zheng, X. Li, H. Cao, T. Lee, and P. Ching. Deriving mfcc parameters from the dynamic spectrum for robust speech recognition. In *Chinese Spoken Language Processing, 2008. ISCSLP '08. 6th International Symposium on*, pages 1–4, Dec. 2008.

[93] C. Zieger, A. Brutti, and P. Svaizer. Acoustic based surveillance system for intrusion detection. In *Proceedings of IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, pages 314–319, Genoa, Italy, 2009.

Publications

International Journals

- **J. Moragues**, A. Serrano, L. Vergara, and J. Gosálbez. Improving detection of acoustic signals by means of a time and frequency multiple energy detector. *IEEE Signal Processing Letters*, 18(8):458-461, Aug. 2011.

- **J. Moragues**, L. Vergara, and J. Gosálbez. Generalized matched subspace filter for non-independent noise based on ICA. *IEEE Transactions on Signal Processing*, 59(7):3430-3434, Jul. 2011.

- L. Vergara, **J. Moragues**, J. Gosálbez, and A. Salazar. Detection of signals of unknown duration by multiple energy detectors. *Signal Processing*, 90(2):719-726, 2010.

- **J. Moragues**, L. Vergara, J. Gosálbez, and I. Bosch. An extended energy detector for non-Gaussian and non-independent noise. *Signal Processing*, 89(4):656-661, 2009.

- **J. Moragues**, L. Vergara, J. Gosálbez and M.A. Gomis. Acoustic event detection based on robust energy detectors. *Sent to Digital Signal Processing.*

International Conferences

- **J. Moragues**, A. Serrano, L. Vergara, and J. Gosálbez. Acoustic detection and classification using temporal and frequency multiple energy detector features. *In Proceedings of the* 36^{th} *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1940-1943, Praga, Czech Republic, May 2011.

- T. Machmer, **J. Moragues**, A. Swerdlow, A. Serrano, L. Vergara, and K. Kroschel. Robust impulse sound source localization by means of an energy detector for temporal alignment and pre-classification. *In*

Proceedings of the 17^{th} *European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, Aug. 2009.

- A. Swerdlow, **J. Moragues**, T. Machmer, L. Vergara, J. Gosálbez, and K. Kroschel. Acoustic detection and classification of sound sources using temporal multiple energy detector features. *In Proceedings of the* 17^{th} *European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, Aug. 2009.

- **J. Moragues**, T. Machmer, A. Swerdlow, L. Vergara, J. Gosálbez, and K. Kroschel. Background noise suppression for acoustic localization by means of an adaptive energy detection approach. *In Proceedings of* 33^{th} *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2421-2424, Las Vegas, USA, Mar. 2008.

- L. Vergara, J. Gosálbez and **J. Moragues**. Audio signal processing for surveillance applications. *In Proceedings of International Conference on Emerging Security Information, Systems and Technologies (SECURWARE)*, Valencia, Spain, Oct. 2007.

Other Conferences

- **J. Moragues**, L. Vergara, and Gosálbez. Novelty detection based on energy detectors. *In Proceedings of the 1º Workshop en Tecnologías de audio cognitivo para aplicaciones en seguridad y acústica forense*, pages 27-30, Valencia, Spain, Jun. 2007.

- J. Gosálbez, A. Salazar, I. Bosch, and **J. Moragues**. Noise control system for building noise areas. *In Proceedings of the 1º Workshop en Tecnologías de audio cognitivo para aplicaciones en seguridad y acústica forense*, pages 23-26, Valencia, Spain, Jun. 2007.