Escuela Politécnica Superior de Linares

Grado en Ingeniería de Tecnologías de Telecomunicación

UNIVERSIDAD DE JAÉN
Escuela Politécnica Superior de Linares

Trabajo Fin de Grado

SISTEMA PARA LA LOCALIZACIÓN DE LA DIRECCIÓN DE DISPAROS PROCEDENTES DE ARMAS DE FUEGO UTILIZANDO TÉCNICAS DE LOCALIZACIÓN DE FUENTES SONORAS

Junio, 2014

Escuela Politécnica Superior de Linares

Grado en Ingeniería de Tecnologías de Telecomunicación

UNIVERSIDAD DE JAÉN
Escuela Politécnica Superior de Linares

Trabajo Fin de Grado

SISTEMA PARA LA LOCALIZACIÓN DE LA DIRECCIÓN DE DISPAROS PROCEDENTES DE ARMAS DE FUEGO UTILIZANDO TÉCNICAS DE LOCALIZACIÓN DE FUENTES SONORAS

RESUMEN

La localización acústica de disparos con armas de fuego tiene muchas aplicaciones en el ámbito militar y civil. Concretamente, un sistema de localización de fuentes sonoras de disparos es fundamental para una correcta protección de personas o incluso para salvar la propia vida en situaciones de atentados terroristas. En este trabajo fin de grado, el alumno deberá implementar un sistema, basado en un array de micrófonos, que sea capaz de localizar la dirección de la cual procede sonidos de armas de fuego. El proyecto se implementará utilizando el entorno MATLAB para su optimización y evaluación.

NOMENCLATURAS

STFT	Short Time Fourier Transform
DFT	Discrete Fourier Transform
IDFT	Inverse Discrete Fourier Transform
FFT	Fast Fourier Transform
x	Señal de entrada temporal
X	Espectro de la señal de entrada
Y	Espectro de la señal factorizada Y=WH
F	Número de bins de frecuencia del espectro
T	Número de frames del espectro
NMF	Non- Negative Matrix Factorization
$_d$	Matriz no negativa de activaciones de la señal de disparo
W_d	Matriz no negativa que contiene información de la señal de disparo
W_n	Matriz no negativa que contiene información del ruido
$_n$	Matriz no negativa de activaciones de ruido
N_d	Matriz de normalización de $_d\ y\ W_d$
N_n	Matriz de normalización de $_n\ y\ W_n$
K	Bases de activación de la etapa de entrenamiento
K'	Bases de activación de la etapa de separación
λ	Parámetro de Sparse
TDOA	Time Delay of Arrival
DOA	Direction of Arrival
GCC-Phat	Generalized Cross Correlation
SRP-Phat	Seteered Response Power

$R_{x_i x_j}$	Correlación cruzada resultado del método GCC-Phat
$\tau_{s,ij}$	Retardo de la fuente sonora para una pareja de micrófonos ij.
P_s	Sumatorio de las correlaciones cruzadas para todos los pares de micrófonos
SNR	Relación señal a ruido en decibelios
TP	Positivo cierto
FN	Falso negativo
FP	Falso Positivo
E_{abs}	Error absoluto
$RMSE$	Raíz de error cuadrático medio
MSE	Error cuadrático medio

ÍNDICE GENERAL

Pág.

CAPÍTULO 1. INTRODUCCIÓN 9

1.1. Motivación 11

1.2. Organización de la memoria 11

CAPÍTULO 2. OBJETIVOS 13

CAPÍTULO 3. ESTADO DEL ARTE 15

3.1. Sonido de un arma de fuego. El disparo 15

3.1.1. Caracterización del sonido de un disparo 15

3.1.2. Tipos De Armas De Fuego 18

3.2. Detección del disparo 19

3.2.1. Métodos usados en detección de disparo 20

3.2.2. Non-negative Matrix Factorization (NMF) 22

3.3.Localización de fuentes sonoras basada en TDOA (time delay of arrival) 25

3.3.1. Fundamentos TDOA 25

3.3.2. CC (Correlación Cruzada) 29

3.3.3. GCC(Correlación Cruzada Generalizada) 30

3.3.4. GCC-PHAT (Correlación Cruzada Generalizada con transformada de fase) 31

3.3.5. SRP-PHAT (Steered Response Power Con Transformada De Fase) 32

3.3.6. Multiple Signal Classification 33

3.4.Plataformas y/o lenguajes de programación para la implementación de localización de fuentes sonoras en tiempo real 34

3.4.1. GPU (Graphics Processing Unit) 34

3.4.2. FPGA (Field-Programmable Gate Array) 35

3.5. Aplicaciones comerciales 36

CAPÍTULO 4. MATERIALES Y MÉTODOS 39

4.1. Implementación de etapa de detección 40

4.1.1. Etapa De Entrenamiento 41

4.1.2. Etapa De Separación 49

4.2. Implementación de etapa de localización 56

4.2.1. GCC-Phat 59

4.2.2. SRP-Phat 67

4.2.3. SRP-Phat modificado 70

CAPÍTULO 5. RESULTADOS Y DISCUSIÓN 77

5.1. Bases de datos 77
5.1.1. Detección 78
5.1.2. Localización 82
5.2. Setup 90
5.2.1. Consideraciones en la etapa de detección 90
5.2.2. Consideraciones en la etapa de localización 92
5.3. Métricas 93
5.3.1. Métricas de detección 93
5.3.2. Métricas de Localización 95
5.4. Análisis de resultados 96
5.4.1. Resultados de etapa de detección 97
5.4.2. Resultados de etapa de localización 111
5.4.3. Coste computacional 119
CAPÍTULO 6. CONCLUSIONES 123
CAPÍTULO 7. LINEAS DE FUTURO 125
CAPÍTULO 8. BIBLIOGRAFÍA 127
ANEXO. Manual de usuario 131

ÍNDICE DE FIGURAS

Pág.

Figura 1. Esquema general del sistema de localización de fuentes sonoras propuesto. 10
Figura 2. Evento acústico presente captado por un micrófono: 'shockwave' [6].....17
Figura 3. Eventos acústicos presentes en señales capturadas por dos micrófonos (30 cm de separación) de un disparo (rifle Winchester 308) a una distancia de aproximadamente 9 metros [2]. 17
Figura 4. Típica señal de un disparo en tiempo. Pistola de 9mm a 30 metros del micrófono [4].....19
Figura 5. Estructura básica del algoritmo de detección basado en umbralización de energía. 21
Figura 6. Representación de un modelo sencillo de localización de una fuente sonora mediante un array de dos micrófonos. 26
Figura 7. Estimación del DoA en el caso de dos micrófonos separados una distancia d. 27
Figura 8. Geometría de un array circular uniforme con M=8 micrófonos. [18].....28
Figura 9. (a) Localizaciones posibles para un determinado TDOA entre dos micrófonos, (b) aproximación cónica.....29
Figura 10. Tabla de diferentes pesados encontrados en la literatura.....31
Figure 11. Esquema del funcionamiento de CUDA [29].....35
Figura 12. Esquema general de la etapa de detección del disparo.....40
Figura13. Curva de divergencia Kullback-Leibler para una señal producida por una señal de un disparo del arma Vickers de calibre .303. 43
Figura 14. Espectro de la señal ocasionada por el disparo de una M4A2 de calibre 5,6. 46
Figura 15. A) Espectro de reconstrucción de la señal de disparo usando distancia Euclidea. B) Espectro de reconstrucción usando divergencia Kullback-Leibler. C) Espectro de reconstrucción de la señal de disparo usando divergencia Itakura-Saito.....47
Figura 16. Gráficas de convergencia de las funciones de costo aplicada a una señal de disparo de una M42A de calibre 5,6: A) Distancia Euclidea. B) Divergencia Kulback-Leibler. C) Divergencia Itakura-Saito. 48
Figura 17. Señal de disparo de arma semiautomática Cz 75 de 9 mm mezclado con ruido de calle. A) Espectrograma de la señal original de entrada (X). B) Espectrograma de la señal reconstruida (Y). C) Espectrograma de Hd. D) Matriz de activación Hd.....52
Figura 18. Convergencia de la Divergencia Kullback-Leibler para 70 iteraciones.....53
Figura 19. Suma de la matriz de activación con la aplicación de suavizado. 54
Figura 20. A) Matriz de activaciones y B) Sumatorio de matriz de activaciones para una señal producida por un arma semiautomática de tipo Beretta 92 Fs de 9mm. Sparse con $\lambda=0$ para ambas etapas.....56
Figura 21. A) Matriz de activaciones y B) Sumatorio de matriz de activaciones para una señal producida por un arma semiautomática de tipo Beretta 92 Fs de 9mm. Sparse con $\lambda=1$ para ambas etapas. 56
Figura 22. Esquema general de la etapa de localización de la fuente sonora originada por un arma de fuego.....57
Figura 23. Esquema general del método GCC-Phat.....61

Figura 24. Representación geométrica del array de micrófonos. 64
Figura 25. Localización de un disparo usando el método GCC-Phat. A) Espectrograma de la señal de entrada. B) Energía de la señal de entrada. C) Representación polar de la localización de la fuente sonora.....66
Figura 26. A) Espectro de la señal capturada por unos de los micrófonos y B) Localización de la fuente sonora evaluando en cada trama. 69
Figura 27. DOA estimado por el método SRP-Phat para una fuente sonora situada a 10° con respecto al centro del array de micrófonos, evitando información irrelevante. 70
Figura 28. Recorte de la matriz de retardos que devuelve la función tau() usando el método GCC-Paht original. Columna1: Micrófono i. Columna 2: Micrófono j. Columna 3: θi. Columna 4: θj. Columna 5: Ángulo candidato. Columna6: Retardo inicial. Columna 7: Retardo final. 72
Figura 29. Recorte de la matriz de retardos que devuelve la función tau() usando el método SRP-Phat modificado con un sector de 20°. Columna1: Micrófono i. Columna 2: Micrófono j. Columna 3: θi. Columna 4: θj. Columna 5: Ángulo candidato. Columna6: Retardo inicial. Columna 7: Retardo final. 72
Figura 30. Implementación del algoritmo SRP-Phat modificado con un sector de 20° para una localización de la fuente sonora de 10°. A) Espectrograma de entrada. B) Localización trama a trama. C) Energía de la señal de entrada. D) Ángulo estimado de disparo.....74
Figura 31. Implementación del algoritmo SRP-Phat modificado con un sector de 2° para una localización de la fuente sonora de 10°. A) Espectrograma de entrada. B) Localización trama a trama. C) Energía de la señal de entrada. D) Ángulo estimado de disparo.....75
Figura 32. Tablero para la simulación de un array circular de 1 metro de diámetro formado por seis micrófonos. 83
Figura 33. Vista general del entorno de simulación. 84
Figura 34. Tarjeta multicanal Octa-Capture de Roland. 86
Figura 35. Edición de preferencias en la plataforma AUDITION. 87
Figura 36. Asignación de canales a pistas de audio en AUDITION. 87
Figura 37. Equipos encargados de la reproducción de la fuente sonora producida por un arma de fuego.....88
Figura 38. Interfaz de audio M-Audio para la reproducción del disparo. 89
Figura 39. Equipos utilizados en la reproducción del disparo. Altavoz a la derecha y subwoofer a la izquierda. 89
Figura 40. Diagrama de barras para λ=0 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.1.....97
Figura 41. Diagrama de barras para λ=0.3 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de BR.1. 98
Figura 42. Diagrama de barras para λ=0.5 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.1.....98
Figura 43. Diagrama de barras para λ=0.7 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.1.....99
Figura 44. Diagrama de barras para λ=1 en la etapa de entrenamiento y λ= [0, 0.3, 0.5, 0.7, 1] en la etapa de separación. Base de datos de ruido BR.1. 99
Figura 45. Diagrama de barras para λ=0 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.....100

Figura 46. Diagrama de barras para λ=0.3 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.....101
Figura 47. Diagrama de barras para λ=0.5 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.....101
Figura 48. Diagrama de barras para λ=0.7 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.....102
Figura 49. Diagrama de barras para λ=1 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.....102
Figura 50. Diagrama de barras para λ=0 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos BI.1.....104
Figura 51. Diagrama de barras para λ=0.3 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos BI.1.....105
Figura 52. Diagrama de barras para λ=0.5 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos BI.1.....105
Figura 53. Diagrama de barras para λ=0.7 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos BI.1.....106
Figura 54. Diagrama de barras para λ=1 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos BI.1.....106
Figura 55. Diagrama de barras para un número de bases K=2, 3, 5 y 7 para diferentes SNR.....110
Figura 56. Error absoluto para GCC-Phat, SRP-Phat y SRP-Phat modificado, usando la base de datos de localización. 111
Figura 57. Desviación estándar para GCC-Phat, SRP-Phat y SRP-Phat modificado, usando la base de datos de localización.....112
Figura 58. Error absoluto resultante de la evaluación de todos los métodos en los casos de motor apagado, arrancado y acelerado. 113
Figura 59. Error absoluto medio de cada método para diferentes relaciones señal a ruido. 114
Figura 60. RMSE de cada método para diferentes relaciones señal a ruido.....114
Figura 61. Desviación estándar media de cada método para diferentes relaciones señal a ruido. 115
Figura 62. Diagrama de barras del error absoluto medio de cada método para diferentes relaciones señal a ruido.....116
Figura 63. Diagrama de barras del RMSE de cada método para diferentes relaciones señal a ruido. 116
Figura 64. Diagrama de barras del desviación típica media de cada método para diferentes relaciones señal a ruido.....117
Figura 65. Vista de la pantalla principal de la aplicación para el usuario. 131
Figura 66. Selección de la relación señal a ruido como parámetro de entrada. 132
Figura 67. Selección de la carpeta que contiene los archivos de audio a analizar.....133
Figura 68. Vista del análisis previo de la señal de entrada. 133
Figura 69. Parámetros de la etapa de detección. 134
Figura 70. Mensaje de aviso para introducir debidamente el parámetro K. 134
Figura 71. Obtención de resultados de la etapa de detección de disparo.....135
Figura 72. Aparición de paneles para la etapa de localización al detectar la señal como disparo.....136

Figura 73. Resultados obtenidos en la etapa de localización con el método SRP-Phat modificado y un sector de 10°, para una fuente situada en 50°.....136
Figura 74. Detección y Localización de una señal de disparo situada a 50°.....137
Figura 75. Mensaje de aviso por no haber seleccionado el método de localización. .137
Figura 76. Disparo no detectado con una señal con SNR=-5 dB y la fuente situada en 50°.....138

ÍNDICE DE TABLAS

Pág.

Tabla 1: Esquema general de Bases de Datos para la etapa de detección. 78

Tabla 2. Resultados del sistema de detección de señales de disparo usando SPARSE con λ=0.7 en la etapa de entrenamiento y λ=1 en la etapa de separación.....103

Tabla 3. Métricas para configuración óptima para entornos ruidosos. 103

Tabla 4. Resultados para un sistema de localización de disparos insertando otros sonidos impulsivos usando SPARSE con λ=0.7 en la etapa de entrenamiento y λ=1 en la etapa de separación. 107

Tabla 5. Métricas para el sistema óptimo elegido usando señales de diferentes sonidos impulsivos. 107

Tabla 6. Porcentajes de detección para diferentes técnicas con Kern lineal y SVMs.108

Tabla 7. Porcentajes de detección para diferentes técnicas con RBF Kern y SVMs. .108

Tabla 8. Porcentajes de detección para NMF con la configuración óptima y la peor en entornos de ruido. 109

Tabla 9. Error absoluto medio y desviación típica media para los métodos implementados. 112

Tabla 10. Resumen de métodos óptimos según la SNR del entorno. 119

Tabla 11. Coste computacional de la etapa de detección. 120

Tabla 12. Coste computacional para la etapa de localización 121

CAPÍTULO 1. INTRODUCCIÓN

Las armas de fuego están presentes entre los ciudadanos, aunque pocos países reconozcan la posesión de armas un derecho constitucional. La ley española, por su parte, no contempla la tenencia de armamento que no se halle relacionada con tiro deportivo o bien con alguna práctica relacionada con la caza, excepto que la autoridad contemple lo contrario en casos excepcionales. Sin embargo, países como Estados Unidos recoge en su Segunda Enmienda el derecho del ciudadano a la posesión de armas de fuego. Contemplados estos diferentes puntos de vista sobre el derecho a la posesión de armamento, no se entrará en debate sobre cuál es la mejor elección por parte de un Estado, sino que, se optará por elaborar una solución a problemas concretos desencadenados del uso nocivo de estos peligrosos objetos.

Hoy día, aunque estemos en pleno siglo XXI, los atentados terroristas por arma de fuego siguen siendo uno de los mayores problemas a los que se enfrenta el servicio de seguridad que se dedica a la protección de personalidades relevantes de la sociedad, por ejemplo policías, soldados o guardaespaldas privados. Imagine el siguiente escenario en el cual, una personalidad importante del ámbito político internacional se encuentra realizando un viaje de Estado en su coche oficial. Si se produjese un intento de atentado mediante un francotirador en un lugar con difícil visibilidad y el coche quedase bloqueado en la carretera sin poder avanzar momentáneamente, dicha personalidad sabe que debe agacharse lo más abajo posible en el centro de la parte trasera del vehículo. La razón se debe a que dicho lugar es el más seguro ya que se desconoce el lugar desde donde el francotirador o terrorista se encuentra disparando. En los instantes iniciales de un atentado de estas características, la primera reacción del personal de seguridad es protegerse de los disparos eligiendo por intuición, en décimas de segundo, una posición segura. Sin embargo, la intuición suele ser mala consejera y consecuentemente, convierte a dicha persona en un blanco fácil.

Pues bien, en estas situaciones sería interesante poder disponer de un sistema que fuese capaz de determinar la dirección desde donde proceden los disparos ya que podría salvar bastantes vidas si la información proporcionada por dicho sistema fuese lo suficientemente fiable y en tiempo real. De este modo, la protección del personal de seguridad, continuando con el ejemplo anteriormente citado, se basaría en un sistema que calculara la localización del arma y no en intuición

La localización de una fuente sonora puede ser determinada utilizando un array de micrófonos y calculando los retardos que existen entre dichas señales capturadas. Igual de importante es el elaborar un sistema de detección de disparos, haciendo así que la localización de la fuente sonora sea exclusivamente para sonidos producidos por armas de fuego. De manera que, este Trabajo Fin de Grado pretende implementar un sistema que determine, mediante técnicas de localización de fuentes sonoras, la dirección desde la cual proceden los disparos. Para cumplir con el requisito de localizar la dirección de una fuente sonora producida por un arma de fuego, el Trabajo Fin de Grado se ha visto ampliado, de manera que se ha diseñado un sistema que permita la detección de señales procedentes de un arma de fuego, sistema que no estaba en los objetivos iniciales del proyecto.

El Trabajo Fin de Grado a citar tiene una beca de colaboración asociada en la cual se extrapola dichos sistemas a su implementación en tiempo real, para ello se recurrirá a un potente hardware. Así, la localización deberá ser lo suficientemente rápida para que la información proporcionada sea considerada útil y fiable evitando detectar el mayor número de falsos positivos y falsos negativos. En la Figura 1 se muestra un esquema del sistema de localización a ejecutar. Dicho sistema estará formado por una primera etapa que se encargará de detectar cuando se ha producido un sonido de disparo con arma de fuego, evitando así que ruido ambiente sea detectado como disparo. Una vez detectado dicho disparo, la segunda etapa del sistema se activará e intentará localizar en tiempo real la dirección desde la cual el arma de fuego ha disparado.

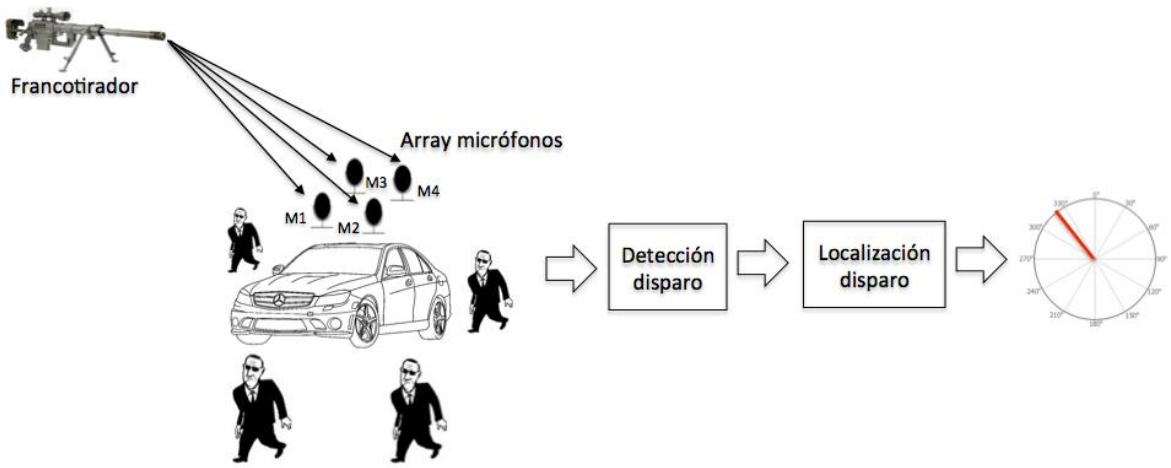


Figura 1. Esquema general del sistema de localización de fuentes sonoras propuesto.

1.1. Motivación

La tecnología avanza a un ritmo que es difícil de alcanzar, siempre surge un nuevo proyecto capaz de desarrollar un servicio que cubra una necesidad o simplemente nos haga la vida más confortable. Para eso está la tecnología, hecha por el hombre para el hombre. He ahí la razón, si tenemos en la mano la posibilidad de, ya no cubrir una necesidad vital, sino conservar la vida, se debe poner la tecnología en marcha para proteger al hombre.

Como se ha visto en la sección anterior, las armas de fuego son capaces de matar, dicho esto, cabe señalar que las armas de fuego son la principal causa de muerte en Estados Unidos por encima de cualquier otra arma u objeto nocivo. Las personalidades relevantes en la sociedad, como políticos, sufren continuas amenazas de este tipo y por ello cuentan con personal que se encarga de su seguridad, enfrentándose a este riesgo. Son un claro ejemplo de la necesidad de máxima protección frente a armas de fuego. Por ello se pretende diseñar un sistema de localización de una fuente sonora ocasionada por un disparo, que permita así protegerse y anticiparse a un nuevo ataque de un francotirador.

De modo que, a la vez que se trabaja en un sistema capaz de salvar vidas se plantea el hecho de su posible comercialización dada su utilidad en el mundo real, ya que se trata de una línea de investigación abierta.

Este proyecto de investigación asociado a un Trabajo Fin de Grado aprovecha conceptos estudiados en el Grado de Ingeniería de Telecomunicaciones como procesado de señal, nociones sobre acústica, codificación y transmisión de señales y programación. Como bien se ha señalado anteriormente se trabaja en una beca de investigación a partir de este proyecto en la cual se va a obtener dicho sistema en tiempo real, de manera que pueda resultar útil a nivel comercial.

1.2. Organización de la memoria

Los diferentes capítulos indicados en el índice marcan la organización del Trabajo Fin de Grado a desarrollar y se detallan a continuación:

Capítulo 2: Detalla los objetivos que enmarcan el proyecto y que deben ser asumidos y llevados a cabo para la correcta finalización del sistema a implementar.

Capitulo 3: Se hace una revisión del estado del arte de las investigaciones científicas en el campo de la localización y detección de fuentes sonoras, indicando que se puede encontrar actualmente y partiendo de estos conocimientos.

Capítulo 4: En esta etapa se narra con detalle la implementación en Matlab de dicho sistema, marcando las tareas requeridas, pautas de trabajo y problemas solventados. El diseño de este sistema consta de dos etapas claramente diferenciadas, como son la detección y la posterior localización, implementándose ambas por separado para después llegar a complementarse.

Capítulo 5: Una vez desarrollado el sistema, en este capítulo se analizarán los resultados obtenidos. Se trata de una evaluación exhaustiva del proyecto realizado, comprobándose que métodos ofrecen una mejor detección y localización del disparo teniendo en cuenta diversos entornos y estudiando sus ventajas e inconvenientes.

Capítulo 6: En esta sección se hará una recopilación global de lo más relevante del trabajo desarrollado y de los resultados obtenidos sacando las conclusiones más certeras del proyecto.

Capítulo 7: En este punto se abren diferentes líneas futuras para seguir trabajando sobre este aspecto, donde se aportan ideas de mejoras o de extensión.

Capítulo 8: Incluye la bibliografía de la información usada.

Anexo: En este apartado se realiza un manual de usuario sobre la plataforma Matlab para guiar al usuario final sobre el uso del sistema implementado y explicar su funcionamiento.

CAPÍTULO 2. OBJETIVOS

El objetivo principal de este proyecto es el diseño e implementación de un sistema de localización de fuentes sonoras de armas de fuego, basado en un array de micrófonos, aplicado a vehículos con la idea de minimizar los daños humanos. Concretamente, el sistema implementado deberá detectar un sonido procedente de un arma de fuego y ser capaz de proporcionar la dirección desde la cual proceden los disparos realizados con armas de fuego. Dicha localización deberá ser lo suficientemente fiable evitando detectar el mayor número de falsos positivos y falsos negativos. Este objetivo principal se divide en los siguientes objetivos secundarios necesarios para poder alcanzar el anterior y que se detallan a continuación:

- **Creación de las bases de datos pertinentes.** Se requiere de archivos de audio determinados a tratar por los métodos implementados para el diseño del sistema. Estas bases de datos contarán con señales de disparos, ruido y grabaciones específicas para la posterior localización de la dirección de la fuente sonora.

- **Identificación de un sonido de disparo de arma de fuego.** En una primera etapa, se aplicarán técnicas de procesado de audio que permitan determinar si un evento acústico puede ser considerado como un sonido originado por un arma de fuego a partir de las características temporales y espectrales que presenta dicho evento. Para ello, un conjunto de señales de armas de fuego serán utilizadas para análisis de características como para posibles enfoques de sistemas basados en entrenamiento.

- **Diseño del sistema de adquisición de señales multicanal.** Se diseñará el tipo de sistema de adquisición de señales multicanal basado en un array de micrófonos planar conectados a un interfaz de audio que permita la captura de dichas señales.

- **Localización de la dirección de la fuente sonora utilizando técnicas TDOA.** En esta etapa, se implementarán los algoritmos más relevantes de localización de fuentes sonoras basados en técnicas TDOA. Para ello, en un primer paso se estimarán los retardos entre las señales multicanal capturadas por los diferentes micrófonos del array. A continuación, se implementarán algoritmos que determinen la dirección de la fuente sonora a partir de dichos retardos. Entre los algoritmos a implementar se encontrarán los siguientes: GCC, GCC-PHAT, SRP-PHAT y SRP-PHAT modificado, tratándose este último de una modificación diseñada en este Trabajo Fin de Grado del método común SRP-PHAT.

- **Evaluación del sistema implementado.** Se realizará la evaluación del sistema implementado, observando si el sistema cumple con las premisas. Para ello, se utilizarán las métricas adecuadas y se atenderán aspectos tales como el tiempo de respuesta del sistema y fiabilidad de los resultados (número de aciertos, número de falsos positivos y número de falsos negativos).

CAPÍTULO 3. ESTADO DEL ARTE

3.1. Sonido de un arma de fuego. El disparo

Los sonidos originados por un arma de fuego son tan peculiares que pueden ser identificados rápidamente por un ser humano. Presentan características acústicas específicas útiles para poder diferenciarlos del resto de los sonidos.

La duración temporal del disparo es de milésimas de segundo. Es un ruido impulsivo del que poco a poco se irá conociendo más a lo largo de esta memoria, como factores de los que depende, fenómenos acústicos observados y su espectro, entre otros.

Básicamente, un sonido producido por un arma de fuego es un sonido de banda ancha, cuya energía se sitúa en altas frecuencias, concentrada en intervalos muy cortos de tiempo. Concretamente la mayor parte de energía está localizada en torno a 1 KHz, y sigue presentando bajo contenido espectral incluso alrededor de los 20 KHz [1].

Una parte de este proyecto se centra en la detección de un disparo en un ambiente con ruido, donde para ello es preciso conocer bien las características acústicas del disparo.

3.1.1. Caracterización del sonido de un disparo

Cuando un arma es disparada, distintos fenómenos acústicos [2] [3] son observados, distinguiéndose dos pulsos de energía distintos.

-<u>Muzzle blast:</u>

Este fenómeno se presenta como una señal acústica de una elevada energía originada desde el cañón (muzzle) del arma. Consiste en un frente de onda esférico que se propaga a la velocidad del sonido, donde el centro de la esfera indica la localización del cañón del arma. El nivel de presión acústica de pico asociado puede exceder de 150 dB en la proximidad del arma de fuego y aproximadamente 90 dB a unos 90 metros sin obstáculos [4]. La energía se encuentra localizada en la banda de 300 – 1000 Hz y la duración asociada suele ser menor de 3 ms, propagándose a la velocidad del sonido (c=343 m/s a 20ºC) [3] [4] [5].

Este estallido es el que se asocia típicamente con el sonido del disparo. El sonido de la explosión se emite en todas direcciones, pero la mayor parte de la energía se percibe en la dirección apuntada por el cañón del arma.

-Shockwave:

Segundo fenómeno acústico el cual se produce debido a que el arma comprime el aire con el que se topa creando un estampido sónico. La onda acústica supersónica se expande en forma de cono detrás de la bala. La apertura del cono depende del índice de la velocidad de la bala con respecto a la velocidad del sonido, donde la apertura del cono decrece con una velocidad de bala mayor. No obstante, algunos tipos de armas de fuego no impulsan balas a velocidades superiores a la velocidad del sonido por lo que no se produce ondas de este tipo, por lo que la detección debe ser realizada considerando únicamente el evento de estallido o *muzzleblast*.

El frente de onda no tiene una forma cónica perfecta, debido principalmente a dos razones: la trayectoria no es una línea recta, sino una curva parabólica debido a la acción de la gravedad y otros factores perturbadores como el viento, y además, la bala irá gradualmente desacelerándose por la fricción con el aire, perdiendo así velocidad [6].

Cuando la onda acústica '*shockwave*' golpea el micrófono, una característica señal en forma de N (Figura 2) puede ser observada en el dominio del tiempo. Su duración es de unos cientos de microsegundos (µs), dependiendo del calibre del arma y la distancia de ésta al micrófono y otros factores menos significantes.

Si los micrófonos están localizados cerca del arma de fuego, la onda directa del *muzzleblast* es la señal acústica primaria. Sin embargo, si el micrófono está localizado a una distancia considerable del arma de fuego, el sonido de la onda directa puede ser enmascarado y la señal recibida mostrará efectos debido a la propagación, reflexiones *multi-path* y reverberación [2]. En la Figura 3, se puede ver estos eventos, donde cada uno presenta diferente tiempo de duración, tiempo de subida (*rise time*) y tiempo de llegada (*arrival time*).

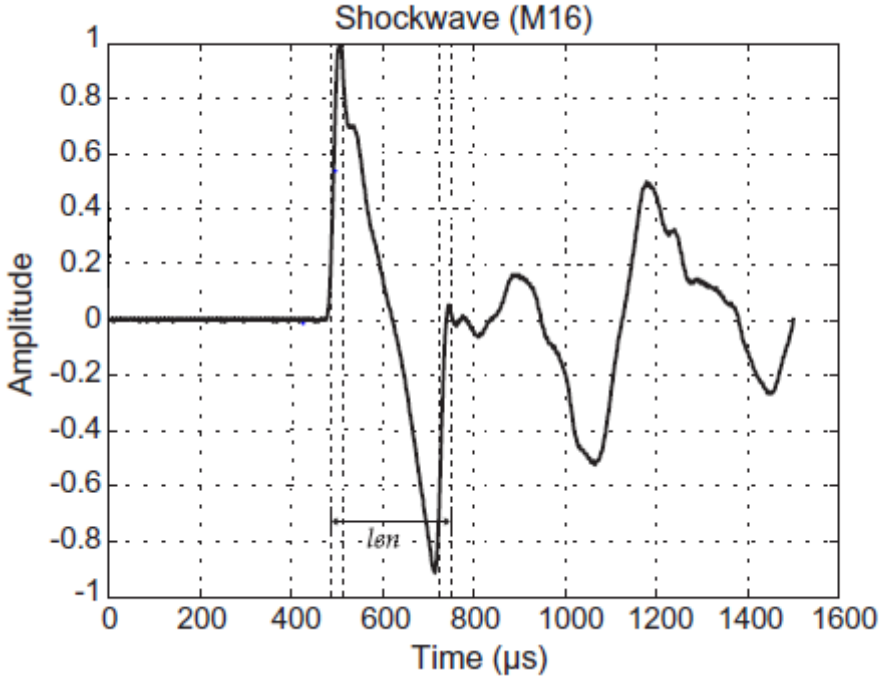


Figura 2. Evento acústico presente captado por un micrófono: 'shockwave' [6].

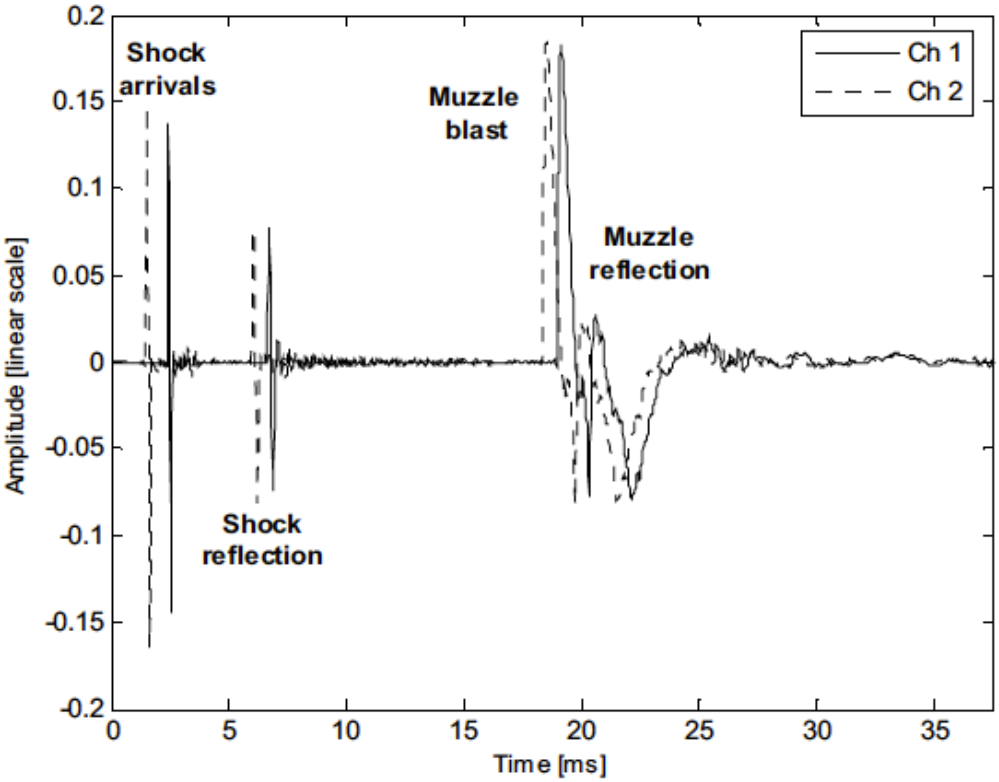


Figura 3. Eventos acústicos presentes en señales capturadas por dos micrófonos (30 cm de separación) de un disparo (rifle Winchester 308) a una distancia de aproximadamente 9 metros [2].

3.1.2. Tipos De Armas De Fuego

Conocida la caracterización acústica de un disparo, hay que saber que ésta depende de factores como el calibre de la bala, la longitud, anchura del cañón del arma o compuestos químicos de la pólvora usada. Las armas de fuego se pueden clasificar según diferentes criterios a tener en consideración [7]. Aquí se muestra una vista general de diferentes clasificaciones.

1. Según el tipo de material.

Se puede distinguir entre armas de hombro y armas de puño.

Las armas de fuego llamadas 'de hombro', son aquellas que para su empleo normal requieren ser apoyadas en el hombre del tirador y el uso de ambas manos. Se diferencian principalmente según tengan su cañón estriado o liso. Aquí estarían clasificadas escopetas, fusiles y carabinas.

Las armas de 'puño' o 'cortas' son aquellas que solo precisan de una mano para ser usadas, sin que intervenga otra parte del cuerpo. En esta subcategoría se encuentra principalmente pistolas, revólveres y pistolones.

2. Según el sistema de disparo.

Tiro a tiro: son las armas que carecen de almacén o cargador y obligan al tirador a repetir manualmente la acción completa de carga del arma en cada disparo; como por ejemplo en las escopetas 'de quebrar o de báscula' de uno o dos caños.

Repetición: son aquellas en las que el ciclo de carga y descarga de la recámara se efectúa mecánicamente por acción del tirador, estando acumulados los proyectiles en un almacén cargador; como por ejemplo los sistemas de cerrojo (un fusil Mauser); de palanca (la tradicional carabina Winchester).

Semiautomático: se trata de las armas en que es necesario oprimir el disparador (gatillo) para cada disparo y en el que el ciclo de carga y descarga se efectúa sin la intervención del tirador; como son por ejemplo la mayoría de las pistolas (Colt .45, Browning 9 mm, etc.).

Automático: son las que manteniendo oprimido el disparador, se produce más de un disparo en forma continua, como por ejemplo las ametralladoras.

3. Según su calibre.

El calibre de un arma de fuego es el diámetro del proyectil que se dispara. Se expresa habitualmente en milímetros (7,65 mm, 9 mm, 11,25 mm), en fracciones de pulgada (.38, .357, .44, .45) o en unidades absolutas (12, 16, 20, utilizado para escopetas y pistolones).

Otras formas de clasificación de armas de fuego es la siguiente: armas de fuego cortas, largas, deportivas, antiguas o históricas, etc. [8].

3.2. Detección del disparo

La señal del disparo es una señal impulsiva, caracterizada por una fase puntiaguda y propiedades no estacionarias [4] [9], como se puede apreciar en la Figura 4. Como se ha descrito en el apartado 3.1.1, la señal producida por el disparo de un arma de fuego tiene unas características peculiares en el tiempo, donde se aprecian dos fenómenos concretos, el muzzleblast y el shockwave, mientras que en frecuencia se trata de una señal de banda ancha donde la mayor parte de la energía se localiza en torno a 1 kHz. Dicho esto, se aprovecharán estas características para implementar algoritmos de detección y clasificación [4].

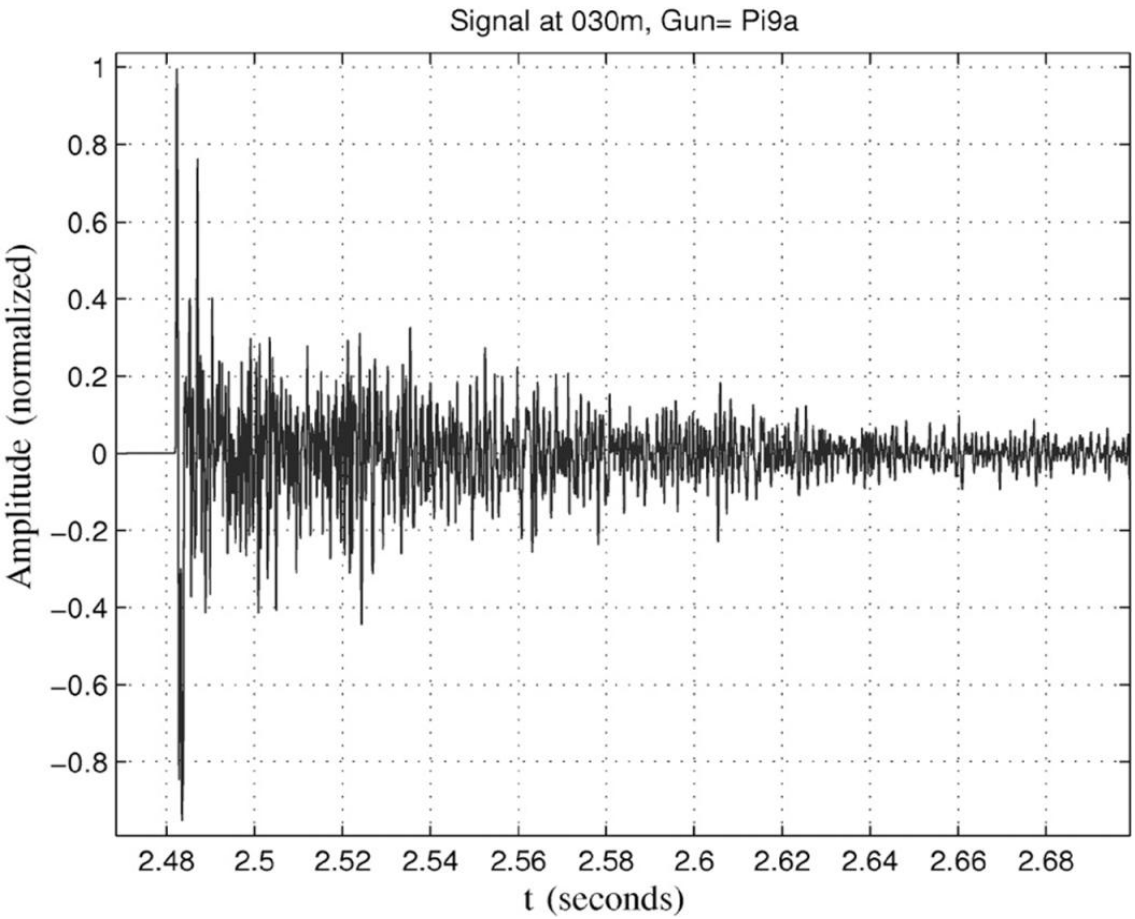


Figura 4. Típica señal de un disparo en tiempo. Pistola de 9mm a 30 metros del micrófono [4].

Los algoritmos para la detección y estimación de llegada (DoA) de un disparo deben ser muy rápidos computacionalmente [9], requieren una detección en tiempo real, de ahí la importancia de la rapidez en respuesta del algoritmo en cuestión. Igualmente, se debe tener en cuenta la eficiencia en términos de disipación de potencia [4], principalmente en términos de clasificación, ya que se suelen usar complejos algoritmos como STFT, transformada wavelet, modelo oculto de Markov (Hidden Markov Model HMM), mezclas gaussianas (GMM) y modelos de probabilidad que son caros en cuanto a potencia debido a sus necesidades computacionales.

A continuación se detallan las técnicas más relevantes usadas para detección y clasificación de disparos.

3.2.1. Métodos usados en detección de disparo

Se cuenta con amplia variedad de documentación que trata sobre diferentes algoritmos de detección de sonidos impulsivos, entre los que se encuentra la señal de un disparo de arma de fuego.

La técnica más sencilla en cuanto a detección de ruidos impulsivos es la *umbralización de energía* [4] [10] [11]. La Figura 5 muestra la arquitectura básica de detección, que consiste en una comparación entre un procesado previo de la señal y un umbral adaptativo, el cual es típicamente una media móvil de la propia señal pre-procesada escalada por un factor de ganancia C ajustable. Otra técnica muy usada dado a su alta eficiencia en la detección de ruido impulsivo es el *filtro de mediana* [4] [11], permitiendo seleccionar el pulso gracias a un filtrado de mediana de la señal de entrada y replicas de ésta pero con un retardo, igualmente la salida obtenida pasa por un umbral.

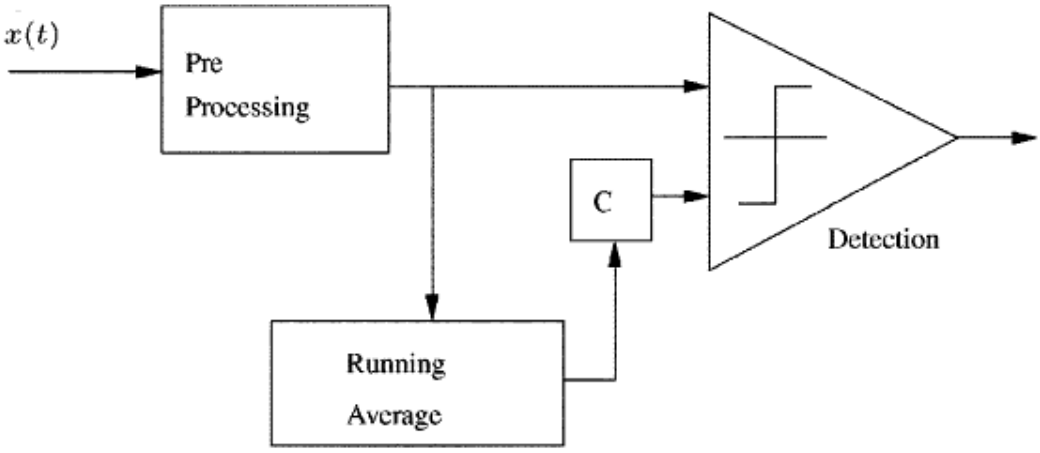


Figura 5. Estructura básica del algoritmo de detección basado en umbralización de energía.

Otros métodos, además de los anteriormente citados, de detección de disparo son los siguientes:

- *Correlación con patrones* [4]: En este caso se establecen unos patrones obtenidos mediante el promediado de señales específicas de de disparo y se almacena en un vector. Posteriormente se realiza la correlación de este vector y la señal de entrada al sistema. El resultado obtenido es llevado a un detector de umbral.
- *Wavelet* [4]: Uso de la transformada discreta wavelet (DWT) para formar un banco de filtros que determine la forma del disparo de manera que la señal de entrada atraviese este banco de filtros y el resultado será evaluado por un detector de umbral.

Además, existen técnicas muy interesantes que permiten la clasificación de sonidos impulsivos, reconociendo si este viene dado por una puerta, una explosión o un disparo, y que se deben mencionar.

- *Estimación bayesiana* [11]: Se trata de uno de los métodos de reconocimiento más simples. La norma general de esta técnica dice que las características de una determinada señal deben ser asociadas con la clase de sonido que posea la máxima probabilidad de acierto a posteriori. Permite la diferenciación de sonidos impulsivos comparando características.
- *Gaussian Mixtures Model (GMM)* [11]: Esta técnica procura mejorar la clasificación de sonidos impulsivos, ya que el reconocimiento en base a la máxima probabilidad de acierto de características, seguida en la estimación bayesiana, cae en el error

de que se considera estacionario el sonido impulsivo. Esto se corrige con la iteración de múltiples funciones de Gaussianas.

- *Hidden Markov Model (HMM)* [11]: Es la técnica estadística de reconocimiento de sonidos impulsivos que mejores resultados ofrece, basándose en en la consideración de que la señal se divide en varios estados y cada uno de ellos tiene sus características estadísticas propias, realizando así un análisis más exhaustivo de la señal.

3.2.2. Non-negative Matrix Factorization (NMF)

Nonnegative matrix factorization (NMF) es una popular técnica de reducción de dimensionamiento de una matriz V, la cual es factorizada en dos matrices W y H [12] [13] [14]. Estas tres matrices no presentan ningún valor negativo.

La matriz de partida V presenta una dimensión FxT, donde F indica los bins de frecuencia y T las frames resultantes. De modo que W y H serán matrices de dimensiones FxK y KxT respectivamente, donde K son las bases tomadas a elección según el caso concreto, para así cumplirse que:

$$V \approx WH \tag{1}$$

W es una matriz que constituye la base específica de la fuente, mientras que H se trata de una matriz de activación, que contiene el peso que indica las frames de activación de la matriz W.

Como se ha comenzado a explicar, esta técnica consiste en reducir la dimensión de la matriz V, de manera que FK + KT << FT, por lo tanto, la factorización es solo aproximada. La factorización (1) es buscada una vez se haya minimizado el problema

$$min_{W,H\geq 0}\, D(V|WH) \tag{2}$$

donde D(V|WH) es la función de coste definida por la siguiente ecuación:

$$D(V|WH) = \sum_{f=1}^{F}\sum_{t=1}^{T} d([V]_{fn}|[WH]_{fn}) \tag{3}$$

y donde d(x|y) es una función de coste escalar. Las más populares son la distancia Euclidea

$$d_{EUC}(x|y) = \frac{1}{2}(x-y)^2 \tag{4}$$

y la divergencia Kullback-Leibler, definida por

$$d_{KL}(x|y) = x\log\frac{x}{y} - x + y \tag{5}$$

Por último, mencionar la divergencia Itakura-Saito [12], definida del siguiente modo

$$d_{IS}(x|y) = \frac{x}{y} - log\frac{x}{y} - 1 \tag{6}$$

Definidas las ecuaciones anteriores (4) (5) (6), se debe señalar que dichas divergencias son casos particulares de β-Divergencia [15], donde la divergencia Itakura-Saito es un caso particular para $\beta=0$, la divergencia Kullback-Leibler para $\beta=1$, y la distancia Euclidea resulta de tomar $\beta=2$.

Para llegar al cálculo de las matrices no negativas W y H se realizan los gradientes de la función $D(V|WH)$ para W y H:

$$\nabla_{\mathrm{H}} \mathrm{D}_{\beta}(\mathrm{V}|\mathrm{WH}) = \mathrm{W}^{\mathrm{T}}\left((\mathrm{WH})^{\cdot[\beta-2]}.(\mathrm{WH}-\mathrm{V})\right) \tag{7}$$

$$\nabla_{\mathrm{W}} \mathrm{D}_{\beta}(\mathrm{V}|\mathrm{WH}) = \left((\mathrm{WH})^{\cdot[\beta-2]}.(\mathrm{WH}-\mathrm{V})\right)\mathrm{H}^{\mathrm{T}} \tag{8}$$

De manera que, una vez llegados a este punto se obtienen los factores W y H iterando produciéndose así una actualización de cada parámetro. Ésta actualización se consigue multiplicando el valor previo a la iteración del parámetro (W o H) con la parte negativa entre la parte positiva del gradiente calculado (7) u (8) de dicho parámetro. Obtenemos así, a modo general [12].

$$H \leftarrow H.\frac{W^T((\mathrm{WH})^{.[\beta-2]}.\mathrm{V})}{W^T(\mathrm{WH})^{.[\beta-1]}} \tag{9}$$

$$W \leftarrow W.\frac{((\mathrm{WH})^{.[\beta-2]}.\mathrm{V})H^T}{(\mathrm{WH})^{.[\beta-1]}H^T} \tag{10}$$

Señalar que, tanto la matriz de partida V como las resultantes estarán normalizadas.

Esta técnica se ha visto empleada en bastantes ámbitos, principalmente en el campo de procesado de señales musicales, donde se obtienen muy buenos resultados con divergencia IS [12], empleado para eliminar el ruido de fondo de una señal musical. Por otro lado, también se ha utilizado para limpiar señales de voz que presentan ruido [14].

La técnica de NMF supervisado y semi-supervisado consta de dos etapas, una de entrenamiento y otra de separación. La etapa de entrenamiento se encarga de tomar los patrones espectrales de la señal en cuestión, tomando como entrada grabaciones de dicha señal aislada, en el caso concreto de señal compuesta por voz y ruido se entrenaría el ruido y la voz haciendo uso de las ecuaciones (9) y (10), para el caso de NMF supervisado. En caso de NMF semi-supervisado únicamente se entrena una de las señales. Después se pasa a la segunda etapa, la de separación, donde usando la matriz de entrenamiento W, obtenida en la etapa previa, se consigue descomponer la señal, limpiándola del ruido, utilizando en este caso las ecuaciones anteriores pero utilizando la información almacenada en W de los patrones espectrales.

La señal de entrada vendrá dada por

$$v(t) = s(t) + n(t) \tag{11}$$

donde s(t) es la señal de voz y n(t) es el ruido. Es importante tener en cuenta que para comenzar a trabajar con esta técnica, los algoritmos de NMF se aplican al espectro de esta señal de entrada, así que la matriz V será

$$V = |STFT(v(t))| \tag{12}$$

Para tener más rigor y precisión en cuanto a la activación de bases que guardan las características del sonido se recurre a un nuevo concepto a tener en cuenta, el *SPARSENESS* [14]. El *sparse* ofrece soluciones donde solo unos pocos elementos de la matriz base (W) sean activados simultáneamente, lo que fuerza a que los elementos que contienen las características propias de la señal sean más específicos.

El *sparse* viene dado por un parámetro λ, que aparece en la matriz de activación, y toma valores comprendidos entre 0 y 1, donde el valor $\lambda=1$ es el más estricto en la activación de bases.

3.3. Localización de fuentes sonoras basada en TDOA (time delay of arrival)

La mayoría de los algoritmos existentes de localización acústica se basan en la estimación de la diferencia en el tiempo de llegada (TDOA, *Time Difference of Arrival*) de la fuente con respecto a cada par de micrófonos. Suponiendo que la señal es grabada por un array de 2 o más micrófonos, estos algoritmos llevan a cabo la localización en dos pasos: en primer lugar, se estima el TDOA de la fuente para cada posible par de micrófonos en el array, y en segundo lugar, a partir de los TDOA estimados (y conociendo la geometría del array) se determina la localización en el espacio resolviendo las ecuaciones de intersección resultantes. La ventaja de este tipo de algoritmos reside fundamentalmente en su eficiencia, lo que permite su ejecución en tiempo real. Por el contrario, aunque existen técnicas eficaces para estimar el TDOA, la precisión de estas se ve afectada por la reverberación, lo cual limita su aplicación en determinados entornos [16].

Existe otro tipo de algoritmos que determinan la localización de la fuente de forma directa, sin estimar explícitamente el TDOA entre pares de sensores. Generalmente, estos algoritmos dividen el espacio donde puede encontrarse la fuente en un conjunto discreto de posiciones (o direcciones) candidatas, y eligen aquella que maximiza un cierto criterio, normalmente basado en beamforming. Aunque estas técnicas son más robustas frente a la reverberación y las interferencias, presentan un coste computacional elevado, debido a que deben evaluarse todas las posibles posiciones. Un ejemplo de este tipo de técnicas se explicará en la sección 3.2.5.

3.3.1. Fundamentos TDOA

Para ilustrar la relación entre el TDOA y la localización de la fuente, supongamos un array compuesto por M micrófonos, donde cada micrófono ocupa una determinada posición en el espacio. Primero se considera el caso más sencillo, en el que se tienen dos micrófonos, separados una distancia *d* y a los que llegará un frente de onda de una dirección arbitraria, tal y como se indica en la Figura 6, por simplicidad, el frente de onda es plano. Esta suposición es perfectamente válida siempre que estemos en campo lejano. Suponiendo que la posición de la fuente es fija, la señal $x_m(t)$ recibida por un micrófono puede expresarse como:

$$x_m(t) = a_m s(t - \tau_m) + b_m(t) \tag{13}$$

donde s(t) es la señal de la fuente, a_m es un parámetro que indica la atenuación de la señal hasta el micrófono m, τ_m es el tiempo de propagación desde la fuente hasta el micrófono m y $b_m(t)$ es el ruido captado por el micrófono. Suponiendo que la separación entre los micrófonos es relativamente pequeña en comparación con la distancia a la fuente (es decir, que las atenuaciones se pueden considerar iguales, a_m = a, ∀m), resulta que todos los micrófonos reciben la misma señal fuente en diferentes instantes de tiempo, debido a que los micrófonos están situados en posiciones distintas. Se puede observar que este modelo de señal es extremadamente sencillo, pues asume que existe un único camino de propagación hasta cada micrófono (es decir, se ignora la reverberación) [17].

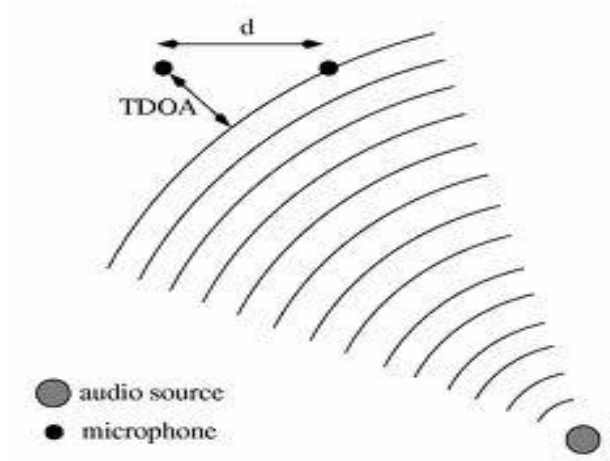


Figura 6. Representación de un modelo sencillo de localización de una fuente sonora mediante un array de dos micrófonos.

Aunque, en principio, los retardos de propagación τ_m son desconocidos, sí que resulta posible estimar el TDOA entre cada par de micrófonos, es decir, se puede determinar la diferencia de tiempo con que ambos micrófonos reciben la señal. Además, si se conoce la distancia *d* entre ambos micrófonos, es posible determinar la dirección de llegada (DOA, Direction of Arrival) del sonido, y por tanto, conocer en qué dirección está localizada la fuente. En la Figura 7, se ilustra un ejemplo de estimación del DOA para una fuente sonora siguiendo con el caso sencillo antes citado. Se puede observar como el frente de onda alcanza dos micrófonos separados una distancia *d* con un ángulo de incidencia θ. Tomando uno de los micrófonos como referencia, se observa que la señal necesita viajar una distancia adicional *d·cosθ* para alcanzar el segundo micrófono. Por lo tanto, el TDOA entre ambos sensores o tiempo de retardo es igual a

$$\tau = \frac{d \cos \theta}{c} \tag{14}$$

donde *c* es la velocidad del sonido en el aire (343 m/s, aproximadamente). En consecuencia, existe una relación directa entre el DOA y el TDOA, de tal manera que estimar el TDOA es esencialmente lo mismo que estimar la dirección de la fuente, $\theta = cos^{-1}(\frac{c\tau}{d})$.

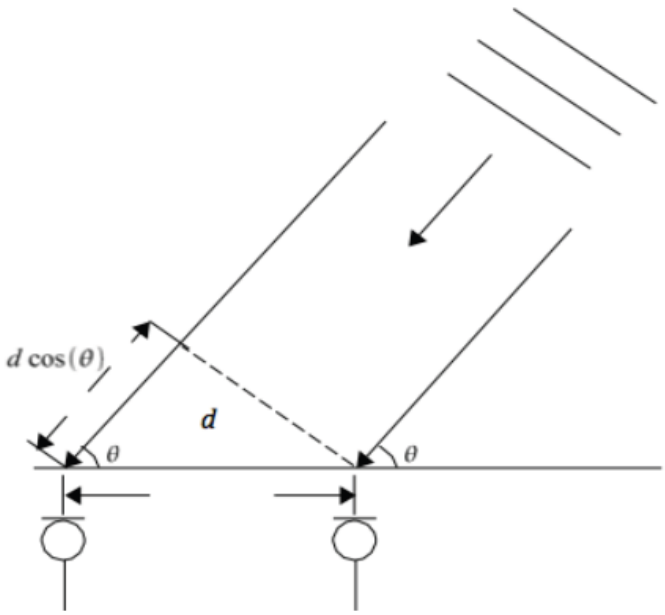


Figura 7. Estimación del DoA en el caso de dos micrófonos separados una distancia d.

Aunque, tal y como se observa en la Figura 7, resulta posible localizar la dirección de la fuente con solo dos micrófonos, existe la ambigüedad delante-detrás (un ángulo -θ y θ produce el mismo TDOA). En realidad, este problema es común a todos los arrays lineales. Para resolver esta ambigüedad, incrementando el número de micrófonos se consigue aumentar nuestra resolución espacial y por tanto, localizar mejor la dirección de la que proviene nuestra fuente sonora (usualmente se emplean arrays circulares de 4, 8, 16, 24… micrófonos). Los array uniformes circulares están siendo muy usados [18], en la Figura 8 se observa como con un array de 8 micrófonos, la ambigüedad en cuanto al ángulo de localización es inexistente

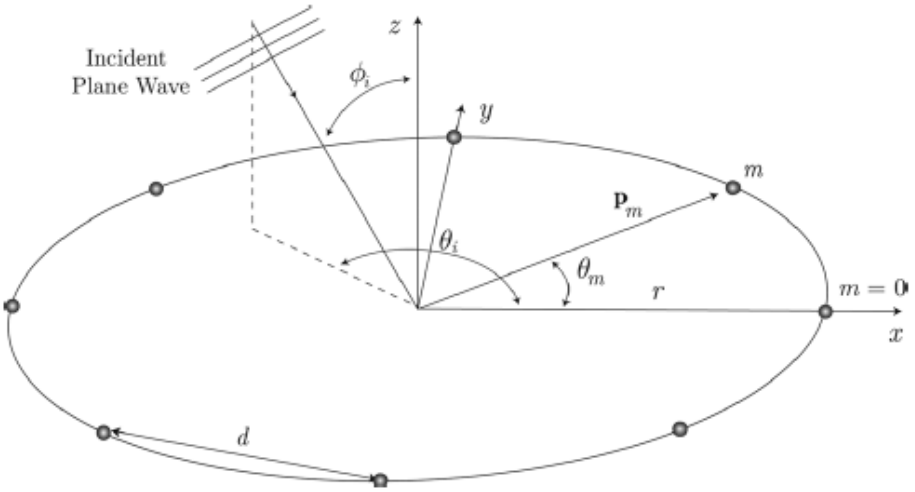


Figura 8. Geometría de un array circular uniforme con M=8 micrófonos. [18].

En este caso, hemos supuesto que la fuente se encuentra en una posición lejana, y en el mismo plano de los micrófonos. Sin embargo, la localización no es siempre tan sencilla.

En términos matemáticos, para un retardo τ fijo con dos micrófonos, y suponiendo que la fuente se encuentra en un plano determinado, el conjunto de puntos del plano que satisfacen un TDOA igual a τ viene dado por una hipérbola cuyos focos son las posiciones de los dos micrófonos [19]. Del mismo modo, si la fuente no está en un plano determinado, sino que puede encontrarse en cualquier posición del espacio tridimensional, el conjunto de soluciones viene dado por una hiperboloide, tal y como muestra la Figura 9. Cuando hay varios pares de micrófonos, los puntos (o regiones) de cruce de las diferentes hiperboloides constituyen las posibles localizaciones de la fuente, de modo que el problema consiste en determinar estos puntos de cruce y elegir la localización óptima. Existen diversos algoritmos eficientes que, a partir de los TDOA

estimados y de las posiciones de los micrófonos, son capaces de producir una solución a este problema. Suponiendo campo lejano, estas hiperboloides pueden aproximarse por conos (ver Figura 9), lo cual facilita el problema de localización y da lugar a algoritmos más simples. En tal caso, la distancia a la fuente se ignora y solo se estima el DOA (en [19] se referencian algunos de estos algoritmos).

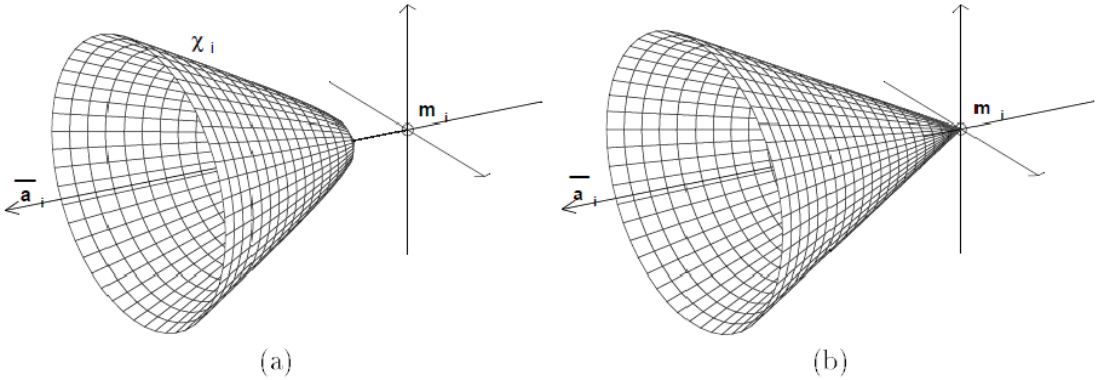


Figura 9. (a) Localizaciones posibles para un determinado TDOA entre dos micrófonos, (b) aproximación cónica.

A continuación, se hará una breve descripción de las técnicas más conocidas y comunes en la literatura para estimar el TDOA para un par de sensores.

3.3.2. CC (Correlación Cruzada)

El método más sencillo para la estimación del TDOA consiste en el cálculo de la correlación cruzada (CC) entre los dos canales de audio. Supongamos que $x_m(t)$ es una trama de señal captada por el micrófono m, donde m ∈ {1,2}. El TDOA estimado con esta técnica es aquel valor de retardo τ que maximiza la función de correlación cruzada entre las dos señales, la cual se define como $R_{x_1x_2}(\tau) = E\{x_1(t)x_2(t+\tau)\}$, donde E{.} es el operador esperanza matemática. En la práctica, el valor de la función $R_{x_1x_2}(\tau)$ es desconocido, por lo que suele estimarse por medio del siguiente cálculo [20]:

$$R_{x_1x_2}(\tau) = \begin{cases} \frac{1}{k}\sum_{t=0}^{k-\tau-1} x_1(t)x_2(t+\tau). \ \tau \geq 0 \\ \frac{1}{k}\sum_{t=-\tau}^{k-1} x_1(t)x_2(t+\tau). \ \tau < 0 \end{cases} \tag{15}$$

donde K es el número de muestras de la trama y donde τ se evalúa en el intervalo

$\tau \in [-\tau_{max}, \tau_{max}]$, donde τ_{max} es el retardo máximo esperado. Un método similar a este, también basado en el dominio del tiempo, consiste en el cálculo de la función diferencia [20] en lugar de la función en (15), obteniendo una precisión similar.

La principal desventaja de esta técnica radica en su debilidad ante la reverberación y ante la presencia de otras fuentes sonoras situadas en posiciones diferentes a la fuente deseada. Esto es debido a que, al tratarse de una técnica basada en dominio del tiempo, es incapaz de aprovechar la ortogonalidad que suelen presentar los sonidos en frecuencia.

3.3.3. GCC(Correlación Cruzada Generalizada)

La correlación cruzada generalizada (GCC, Generalized Cross-Correlation) es probablemente el método más conocido para la estimación del TDOA. El algoritmo fue propuesto inicialmente por Knapp y Carter en 1976 [21], y existen diversas variantes del mismo. El método puede verse como una versión mejorada del algoritmo CC, introduciendo un marco general en el dominio de la frecuencia para el cálculo de la correlación entre dos señales.

Al igual que el método anterior, el algoritmo se basa en el modelo de señal formulado en la ecuación (13). Para comprender mejor el algoritmo, resulta conveniente formular (13) en el dominio de la frecuencia para el caso de dos micrófonos. Sin pérdida de generalidad, podemos descartar los parámetros de atenuación y tomar como referencia el micrófono 1 para expresar el retardo hacia el micrófono 2. En ese caso, podemos escribir:

$$X_1(t,f) = S(t,f) + B_1(t,f)$$
$$X_2(t,f) = S(t,f)e^{-j2\pi f\tau_s} + B_2(t,f) \tag{16}$$

donde $S(t,f)$ y $X_m(t,f)$ denotan la STFT (Short Time Fourier Transform) de la señal fuente y la señal observada en el micrófono m, siendo t = 1,…T y *f* = 1,…F los índices de trama temporal y bin de frecuencia, respectivamente. En esta ecuación, τ_s es el TDOA de la fuente respecto a los dos micrófonos, el cual se debe estimar. Un τ_s positivo indica que la señal llega más tarde al micrófono 2, y viceversa.

La función GCC $R_{x_1x_2}(t,\tau)$ en una trama t para el valor de retardo τ se calcula como la transformada inversa de Fourier del espectro cruzado de la señal recibida $X_1(t,f)X_2^*(t,f)e^{-j2\pi f\tau_s}$, el cual es multiplicado por una función peso $W(t,f)$, de la siguiente forma [15]:

$$R_{x_1x_2}(t,\tau) = \Re\left(\sum_{f=1}^{F} W(t,f)X_1(t,f)X_2^*(t,f)e^{-j2\pi f\tau_s}\right) \tag{17}$$

donde $\Re(.)$ es el operador parte real. Obsérvese que, para cada bin tiempo-frecuencia el cálculo $W(t,f)X_1(t,f)X_2^*(t,f)e^{-j2\pi f\tau_s}$ produce un valor máximo cuando se evalúa $\tau = \tau_s$ siempre y cuando la señal observada se ajuste al modelo formulado en la ecuación (16).

El tipo de filtrado o función de peso utilizada en el método GCC tiene un impacto determinante en la robustez del algoritmo. Esta función de peso se introduce para contrarrestar los efectos de la reverberación y las interferencias, permitiendo que el algoritmo siga funcionando incluso cuando las condiciones se apartan mucho del modelo de señal descrito en (16). En la literatura se han propuesto diversas expresiones para la función $W(t,f)$ para conseguir dicho objetivo [21]. De entre todas las propuestas que se detallan en la Figura 10, la función que obtiene mejores resultados es la que se describe en la sección siguiente.

Processor Name	Weighting Function
Cross-correlation	1
ROTH	$1/G_{xy}(f)$
SCOT	$1/\sqrt{G_{xx}(f)\cdot G_{yy}(f)}$
PHAT	$1/\lvert G_{xy}(f)\rvert$
Cps-m	$1/\sqrt[m]{G_{xx}(f)\cdot G_{yy}(f)}$
HT (ML)	$\dfrac{\lvert\gamma_{xy}(f)\rvert^2}{\lvert G_{xy}(f)\rvert\cdot\left(1-\lvert\gamma_{xy}(f)\rvert^2\right)}$
Eckart	$\dfrac{\lvert G_{xy}(f)\rvert}{\left(G_{xx}(f)-\lvert G_{xy}(f)\rvert\right)\cdot\left(G_{yy}(f)-\lvert G_{xy}(f)\rvert\right)}$
HB	$\lvert G_{xy}(f)\rvert/G_{xx}(f)\cdot G_{yy}(f)$
Wiener	$\lvert\gamma_{xy}(f)\rvert^2$

Figura 10. Tabla de diferentes pesados encontrados en la literatura.

3.3.4. GCC-PHAT (Correlación Cruzada Generalizada con transformada de fase)

Este algoritmo, conocido como GCC-PHAT, surge de la combinación de GCC y la función peso para realizar transformada de fase (PHAT). Concretamente, la función peso viene dada por la siguiente expresión:

$$W_{PHAT}(t,f) = |X_1(t,f)X_2^*(t,f)|^{-1} \tag{18}$$

Esta función elimina la influencia del valor absoluto del espectro cruzado, haciendo que la función GCC dependa exclusivamente de la diferencia de fase entre los dos canales para cada bin de frecuencia.

A través de diversos experimentos, se ha demostrado en la literatura que GCC-PHAT es el algoritmo que produce mejores resultados en condiciones realistas, incluyendo entornos con cierta reverberación [22], siempre que el ruido ambiente no sea muy elevado. Igualmente, cuando existen varias fuentes de audio simultáneas, el algoritmo es capaz de detectar el TDOA de cada una de ellas, puesto que suelen aparecer como picos diferentes en la función GCC. En [23], se indican algunas modificaciones interesantes del algoritmo.

3.3.5. SRP-PHAT (Steered Response Power Con Transformada De Fase)

Además de localizar fuentes sonoras, un array de micrófonos puede actuar como beamformer. En otras palabras, las señales capturadas por los diferentes micrófonos pueden combinarse de cierta manera para seleccionar la señal procedente de una determinada dirección, atenuando el resto de sonidos captados. La técnica más sencilla de beamforming se conoce como retardo-suma, y básicamente consiste en retardar la señal recibida en cada micrófono de forma que la fuente deseada quede alineada en todos los canales. A continuación, todos los canales se suman, haciendo que la fuente deseada destaque sobre las demás. A la señal de salida producida por el beamformer se le denomina Steered Response.

Las técnicas de beamforming pueden utilizarse tanto para capturar una señal como para localizarla. Si la localización de la fuente es conocida, entonces el beamformer puede orientarse en esa dirección para recoger una versión mejorada de la fuente. Si la localización no es conocida, el beamformer puede utilizarse para escanear la región del espacio donde puede encontrase la fuente, de forma que la localización elegida es aquella para la cual la potencia de la señal de salida del beamformer (SRP, Steered Response Power) es máxima.

El algoritmo más conocido de localización mediante beamforming se denomina SPR-PHAT, el cual está estrechamente relacionado con la técnica GCC-PHAT. Supongamos un array compuesto por M micrófonos. En ese caso, la SRP producida por el beamformer cuando enfoca a la posición s (un vector de coordenadas espaciales, s = [x y z]) para una trama n de T muestras de define como:

$$P_n \equiv \int_{nT}^{(n+1)T} \left| \sum_{m=1}^{M} w_m x_m(t - \tau(s,m)) \right|^2 dt \tag{19}$$

donde w_m es un coeficiente de peso y $\tau(s,m)$ es el tiempo de propagación entre la posición s y el micrófono m. En [24] [25] se demuestra que el SRP de un beamformer se puede calcular sumando las funciones GCC (o GCC-PHAT) de todos los posibles pares de micrófonos (h,l) del array. De esta forma, el SRP se puede expresar como:

$$P_n(s) = \sum_{h=1}^{M} \sum_{l=h+1}^{M} R_{X_h X_l}(n, \tau_{hl}(s)) \tag{20}$$

donde τ_{hl} es el TDOA para el par de micrófonos (h,l) cuando la fuente se encuentra en la posición s.

El algoritmo SRP-PHAT consiste entonces en evaluar la función P_n s para un conjunto de posiciones candidatas, hasta encontrar la localización que proporciona el valor máximo. Aunque este procedimiento es extraordinariamente robusto, produciendo mejores resultados que las técnicas basadas en estimación del TDOA, la precisión del algoritmo se ve limitada por la resolución del espacio de candidatos. Además, cuando el número de candidatos es muy elevado, el coste computacional aumenta considerablemente, impidiendo la implementación del algoritmo en tiempo real [25].

Recientemente, diversas modificaciones del algoritmo se han propuesto para mejorar su eficiencia computacional, incluyendo la forma de implementarlo en GPUs [26].

3.3.6. Multiple Signal Classification

Otro tipo de técnicas que permiten estimar la localización de una fuente incluye a los métodos basados en subespacios. Originalmente, estos métodos fueron diseñados para detectar la procedencia de una fuente de banda estrecha en el campo de radar y

telecomunicaciones, siendo el algoritmo MUSIC (MUltiple SIgnal Classification) el más conocido de todos [27]. Esencialmente, estos métodos se basan en la descomposición de la matriz de covarianza espacio-temporal estimada a partir de las señales de los micrófonos. Esta matriz de covarianza puede descomponerse en una serie de autovectores ortogonales. Aquellos autovectores cuyos autovalores sean más altos constituyen la base de un subespacio conocido como subespacio de señal, mientras que el resto forman el subespacio de ruido. A partir de los autovectores de señal es posible conocer la localización de la fuente. Una ventaja de estos métodos es que permiten localizar varias fuentes de forma natural, aunque no son más precisos que las técnicas expuestas anteriormente para la localización de señales acústicas [23].

3.4. Plataformas y/o lenguajes de programación para la implementación de localización de fuentes sonoras en tiempo real

Como se ha explicado a lo largo del proyecto, aunque existen diferentes tipos de algoritmos de localización de fuentes sonoras TDOA, no todos permiten una implementación en tiempo real debido a su elevado coste computacional. A continuación, se presentan diferentes alternativas a considerar para implementar el sistema de localización propuesto en este proyecto en tiempo real.

3.4.1. GPU (Graphics Processing Unit)

GPU (Graphics processing unit) es una unidad de procesamiento gráfico utilizada para acelerar la ejecución de grandes volúmenes de operaciones de cálculo científico. Éste permite conseguir un rendimiento de aplicaciones muy rápido ya que descarga en la GPU las partes de la aplicación que requieren gran capacidad computacional, mientras que el resto del código sigue ejecutándose en la CPU. Así, la GPU consta de millares de núcleos más pequeños y eficientes diseñados para el rendimiento en paralelo. Las partes en serie del código se ejecutan en la CPU mientras que las paralelas se ejecutan en la GPU.

Hoy en día, la principal plataforma informática de GPU es NVIDIA´s CUDA (Compute Unified Device Architecture) [28]. Hay otra plataforma para GPU llamada Open CL (Open Computing Language), que se distribuye libremente, sin embargo CUDA presenta un mayor rendimiento. CUDA es un modelo de programación paralela y escalable que permite el uso directo con el lenguaje C/C++. Para un correcto manejo de CUDA, los

datos deben estar inicialmente paralelizados. Para ello, se crean funciones GPU, que se ejecutan en paralelo a través de un conjunto de hilos (estos hilos tienen una memoria local asociada). La programación establece una jerarquía para organizar estos hilos (ver Figura 6).

Particularizando a este proyecto, se podrían proponer diferentes implementaciones en GPU para implementar la localización de fuentes sonoras, paralelizando las diferentes ubicaciones posibles de la fuente sonora (p.e SRP-PHAT) lo cual no supondría un aumento significativo del coste computacional al incrementar el número de localizaciones candidatas como ocurriría implementando sobre CPU utilizando lenguaje C/C++. De esta forma, el algoritmo implementado en GPU es más eficiente y proporcionaría una mayor rapidez de ejecución.

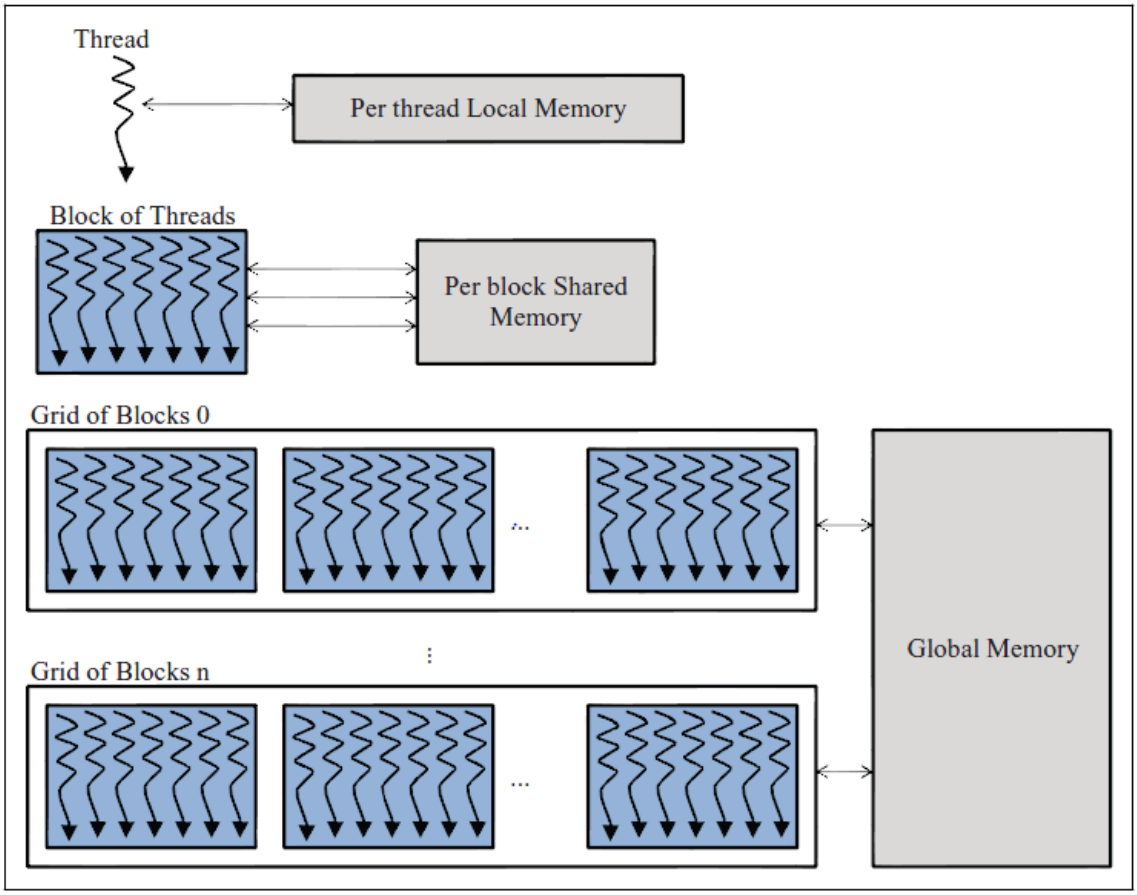


Figure 11. Esquema del funcionamiento de CUDA [29].

3.4.2. FPGA (Field-Programmable Gate Array)

Otra alternativa interesante se encuentra en FPGA (Field Programmable Gate Array). Una FPGA es un dispositivo semiconductor que contiene bloques de lógica cuya

conexión y funcionalidad puede ser configurada mediante un lenguaje especializado. Combinando los bloques lógicos e interconexiones de los FPGA con microprocesadores y periféricos, se obtiene como resultado tecnologías híbridas tales como Virtex-4 de Xilinx, capaz del procesado de señal [30]. Así, este dispositivo podría ser programado permitiendo procesar la señal de entrada en un corto periodo de tiempo, garantizando una muy buena respuesta (tiempo real).

Particularizando a nuestro proyecto, se podría disponer de tres módulos fundamentales: un módulo de muestreo, un módulo de control (DPRAM, Dual-Port Random Access Memory) y un módulo de localización, este último deberá realizar los cálculos requeridos para la localización de fuentes sonoras [31].

Comentar que la implementación con este sistema es ideal para aplicaciones en tiempo real o dispositivos móviles. Entre sus desventajas aparece su alto consumo de potencia y su coste de desarrollo.

3.5. Aplicaciones comerciales

El campo de trabajo de este proyecto está siendo muy estudiado y mejorado en la actualidad, contando con técnicas cada vez más robustas, precisas y rápidas, tratando siempre de buscar la optimización de estos métodos. Primeramente hay que indicar que ambas etapas desarrolladas, localización y detección de sonidos impulsivos como armas de fuego, cuentan con aplicaciones en la vida real conjuntamente y de manera independiente.

Para comenzar, señalar la importancia que está teniendo la localización de una fuente sonora en el ámbito de las comunicaciones, donde las técnicas citadas en la sección 3.2. se están usando en la implementación de sistemas en tiempo real de comunicaciones con manos libres, aplicado a conferencia [33] o a la videoconferencia automática [32], usando un array de micrófonos repartidos por una habitación como sensores, capaces de localizar la fuente sonora y enviar la información a una cámara. Además, una aplicación relativamente nueva es la de reconocimiento y localización de voz, implementado en el campo de la robótica [34]. Y no solo se encuentra en la literatura reconocimiento y localización de voz, ya que igualmente se encuentran técnicas de detección de diferentes tipos de sonidos, como el sonido impulsivo [10] (que como ya se sabe, el disparo es un tipo de sonido impulsivo). En la sección 3.1. se han podido ver técnicas propias de detección de ruido impulsivo, basándose en sus características espectrales y temporales, muy diferentes a las de sonidos no impulsivos. Además de estas aplicaciones de la

localización de una fuente sonora (SSL), en la amplia documentación se menciona otros campos de interés como son la seguridad y la vigilancia.

En cuanto a lo que concierne a la detección de un sonido ocasionado por un arma de fuego, se encuentra diversa información sobre aplicaciones principalmente en el ámbito militar, que ha sido una de las principales fuerzas de búsqueda de sistemas que se encarguen de localización, detección y clasificación del disparo, además de cálculo de trayectoria de un disparo [4][6]. En definitiva, señalar que los sistemas de localización de fuentes sonoras producidas por armas de fuego son particularmente interesantes en áreas relacionadas con la salud pública, vigilancia, cuerpos policiales o de seguridad y en el campo militar.

Se puede encontrar aplicaciones comerciales de localización de disparo concretas. A continuación se exponen algunos de ellos:

- Aplicaciones de tal envergadura como el sistema desplegado por la compañía de tecnología de defensa británica *QinetiQ* [42]. El Ejército de EE.UU. y el ejército británico usan sistemas de localización portátiles de esta multinacional. El sistema tiene dos componentes principales, cada uno tiene el tamaño de una baraja de cartas. Uno de ellos es el sensor con un pequeño conjunto de micrófonos montados en el hombro del soldado, que tiene la capacidad de cubrir 360°, mientras que el otro tiene una pequeña pantalla LCD monocroma que actúa como la interfaz de usuario. Siendo un sistema comercial militar, su especificación exacta no está disponible. La localización del disparo se basa en el conocimiento de las características propias del disparo, *muzzleblast* y *shockwave*, y en la separación entre pares de micrófonos para así poder determinar el TDOA y calcular la posición de la fuente sonora. El sistema responde en menos de un segundo ante un primer tirador.
- El sistema de detección de disparo *Boomerang* [43] fue desarrollado por *BBN* para advertir sobre amenazas de un francotirador. Hoy en día, más de 1.000 *Sistemas Boomerang* están operativos en todo el mundo. Inicialmente, se instaló el sistema en los vehículos, pero la mejora más reciente, es su implementación de manera portátil. El sistema proporciona una indicación inmediata de ataques de armas de fuego y localiza la posición del francotirador, lo que permite una respuesta rápida, informada y coordinada. El sistema puede ser instalado en un vehículo, en un puesto de control, o una escala superior para la seguridad del perímetro completo de una base de operaciones de avanzada. La capacidad para identificar las ubicaciones de tiro insurgentes elimina los despliegues y la exposición al fuego enemigo desconocido.

Estas dos versiones de los localizadores de de disparo citadas se examinaron en la AUSA (Association of the United States Army*).* Ambos se están abordando por los militares de EE.UU., en un intento de mejorar la eficacia de la respuesta y reducir la vulnerabilidad a los ataques por francotiradores y armas ligeras emboscadas.

- Otra de las aplicaciones comerciales que se puede encontrar es el sistema de nodos inteligentes en red inalámbrica. Los ingenieros del *Instituto de la Universidad de Vanderbilt de Software de Sistemas Integrados* (ISIS) han desarrollado un sistema que puede dar a los soldados información en tres dimensiones de la localización del tirador e identificar el calibre y el tipo de las armas con tan solo girar sus cascos de combate donde se ha instalada la red de sensores inalámbrica que forman los "nodos inteligentes" [44], el sistema va montado en el casco del soldado. *ISIS* ha desarrollado la tecnología con el apoyo de la Defensa Agencia de Proyectos de Investigación Avanzada (*DARPA*). Al igual que otros sistemas de localización de disparos desarrollado en los últimos años, el sistema de *ISIS* se basa en ondas de sonido que se produce cuando un rifle de alta potencia se dispara. Estas señales acústicas tienen características distintivas que permiten a los sistemas diferenciarlos de entre otros ruidos fuertes y rastrear de nuevo a su fuente. Los sistemas anteriormente citados, sin embargo, se basan en matrices de sensores centralizadas y estática, lo que límita la identificación de los tiradores en las localizaciones de línea de vista. El sistema *ISIS* combina la información de un número de nodos para triangular las posiciones del tirador y mejorar la exactitud de su proceso de identificación de ubicación.

CAPÍTULO 4. MATERIALES Y MÉTODOS

El sistema a implementar se divide en dos etapas principales bien diferenciadas, las cuales son implementadas sobre la plataforma Matlab. La primera etapa consiste en la detección del disparo, pretendiendo ésta identificar la fuente sonora producida por el arma de fuego a pesar del entorno de ruido en el que se halle dicho sonido. Vistas las múltiples técnicas con las que se cuenta en la actualidad y que funcionan bastante bien según lo citado en la documentación, se ha propuesto, sin embargo, a la detección usando un método no utilizado antes, a conocimiento del alumno, en este campo, se trata del método Non-Negative Matrix Factorization (NMF). Este método consiste en la factorización de la matriz no negativa del espectro de la señal de entrada llegando a dos matrices, también no negativas, de manera que, una contiene la información base de la matriz original y la otra sus activaciones, consiguiendo así, encontrar la activación producida por una señal de disparo.

Una vez se ha identificado un tramo de señal como portador de una fuente de sonido ocasionada por un disparo, se pasa a la segunda etapa del sistema a desarrollar. Esta etapa se centra en localización del arma causante del disparo, para así responder rápidamente ante un nuevo ataque en un entorno de implementación en tiempo real. De modo, que aplicando el sistema al caso en cuestión, que es el de implementar un sistema de localización de una fuente sonora producida por un arma de fuego adaptado a un vehículo para personal propenso a ataques terroristas, el interés principal es conocer la dirección desde la que se está produciendo el ataque, así que ésta etapa se centra en esto. Para alcanzar dicho objetivo, se empleará un array circular de micrófonos y como algoritmos de detección se recurre a los métodos GCC-Phat y SRP-Phat. Posteriormente, se diseña un algoritmo que tiene como base las características de SRP-Phat, pero con ciertas peculiaridades que puede hacerlo más robusto si cabe al ruido, se trata del SRP-Phat modificado, del que se hablará con detalle a lo largo de este capítulo.

A continuación se desarrolla con detalle el trabajo de implementación llevado a cabo para llegar a la realización de estas dos etapas del sistema. Como esquema general para comprender en qué van a consistir ambas etapas de este proyecto se puede observar la Figura 12 y 22 que aparecen en su respectiva sección de desarrollo.

4.1. Implementación de etapa de detección

En la Figura 12 se representa el esquema general en qué va a consistir el proceso de detección del disparo. Dentro de este apartado se va a contar las dos subetapas en las que se divide la detección utilizando el método de factorización NMF. Estas etapas consisten, primeramente en el entrenamiento del método para que sea capaz de identificar la señal de interés, para ello se hará uso de una base de datos formada por pistas únicamente de disparos de diferentes armas, atendiendo a cada caso visto en el Capítulo 3, sección 3.1.2., donde se detallan los tipos de armas que se pueden encontrar. Una vez se ha obtenido la matriz no negativa W de la etapa de entrenamiento, que almacena la información específica de la fuente (información de los disparos en este caso), ésta se usa en la segunda subetapa. La etapa de separación consiste en, dada una señal compuesta por ruido y un disparo, identificar el disparo, consiguiéndose gracias a la matriz no negativa H, que contiene el peso que indica las frames de activación de la señal de interés.

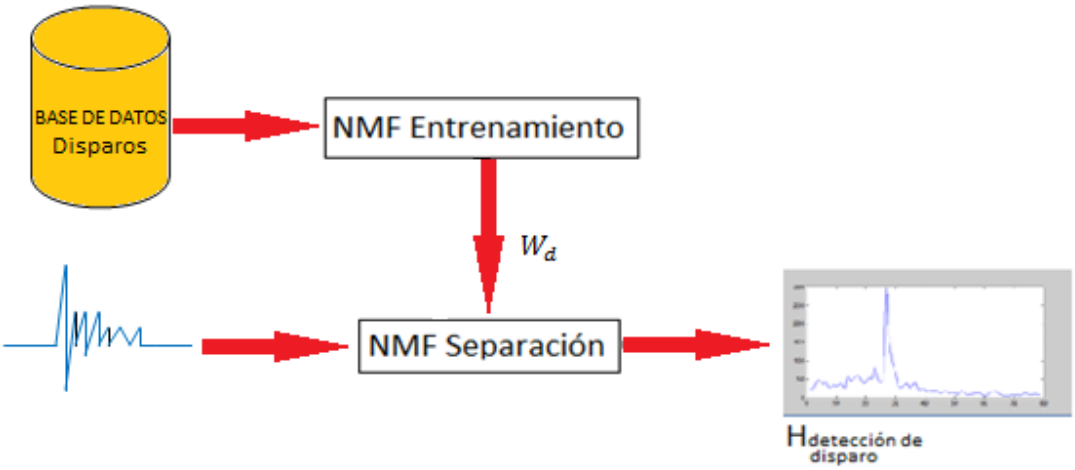


Figura 12. Esquema general de la etapa de detección del disparo.

Las señales de entrada que se han usado están a una frecuencia de 48 KHz. Se debe calcular en primer lugar su espectro (12), que como se sabe en base a los antecedentes es la entrada del algoritmo NMF. Para ello, la STFT ha sido computada con una ventana Hamming de 2048 muestras (42 ms), siguiendo la documentación [12] y teniendo en cuenta que la duración de la señal de disparo es menor que este valor, y con un 50% de overlap entre tramas.

Dado este esquema general, se pasa a contar con detalle los citados procesos, presentando las ecuaciones usadas en cada etapa del método NMF para detección de un disparo.

4.1.1. Etapa De Entrenamiento

El objetivo de la detección es el poder llegar a separar el disparo de todo lo demás, que es considerado ruido, de modo que para comenzar se realiza en esta primera etapa un estudio de las características de la señal de disparo, que como ya se sabe, se trata de una señal de banda ancha con la mayor parte de su energía localizada en torno a 1 KHz. La señal de entrada a dicho algoritmo consiste en un disparo limpio de ruido, ésta señal será factorizada de manera que quede separada en dos matrices tal y como indica la ecuación (1).

Para comenzar, se debe indicar que se cuenta con una base de datos de señales de disparo únicamente, que será la usada en esta etapa. Dicha base de datos cuenta con 150 señales de disparo. Las señales de entrada de la etapa de entrenamiento del NMF semi-supervisado serán cada uno de los disparos que forman parte de esta base de datos.

Primeramente, se comienza por realizar el valor absoluto de la STFT (Short-Time Fourier Transform) de la señal de entrada, $X = \left|STFT\left(x(t)\right)\right|$, obteniendo su espectro. Es necesario la normalización de la señal espectral, tal y como se indica en [12] y se muestra en la siguiente ecuación, para así manejar las señales sin consideración de bins en frecuencia ni frames temporales.

$$X_{norm} = \frac{X}{\sigma_x} \tag{21}$$

donde

$$\sigma_x = \left(\frac{\sum_{f=1}^{F}\sum_{t=1}^{T} X_{ft}^{\beta}}{FT}\right)^{1/\beta} \tag{22}$$

donde β hace referencia a β-Divergencia [15], con $\beta=0$ para la divergencia Itakura-Saito, con $\beta=1$ para la Kullback-Leibler y $\beta=2$ para la distancia Euclidea. En cuanto a los parámetros F y T, estos hacen referencia a los bins temporales y los frames respectivamente, indicando igualmente la dimensión de la matriz de entrada. La

normalización de X permite trabajar con el espectro de entrada sin consideraciones de bins y frames.

Posteriormente, se pasa a la factorización de la señal normalizada $X \approx WH$. Para llegar a estas matrices no negativas es necesario partir de la minimización de una función de coste determinada, esto se consigue calculando el gradiente de la función para W y H. Una vez calculado dicho gradiente, donde se puede ver su resultado en las ecuaciones (7) y (8) para un caso general de β- divergencia, se calculan las matrices actualizándolas de la siguiente manera general: $\theta \leftarrow \theta.[\nabla f(\theta)]_{-}/[\nabla f(\theta)]_{+}$ donde $\nabla f(\theta) = [\nabla f(\theta)]_{+} - [\nabla f(\theta)]_{-}$. Con esta notación lo que se pretende explicar es que dado el gradiente de la función de coste para cada uno de los parámetros, que estará compuesto por sumandos positivos y negativos, las matrices W y H resultan del cociente de la parte negativa del gradiente calculado para dicho parámetro entre la parte positiva.

Dicho esto, el algoritmo NMF comienza con la inicialización de W y H con valores aleatorios usando la función rand() de la librería de Matlab, para seguidamente iterar con los valores de W y H actualizando las matrices tal y como se ha explicado y se define en las ecuaciones (9) y (10), hasta alcanzar así el valor más próximo a X. El dimensionado de W y H viene dado por FxK y KxT, respectivamente, donde F corresponde a los bis de frecuencia, T a los frames temporales y K indica el número de bases a tratar. K es un parámetro que debe ser optimizado, observando con que valor de K se obtienen mejores resultados en cuando a detección.

Lo ideal sería que la señal de entrada al sistema (X) fuera igual a la señal de salida del sistema (X=Y=WH), pero esta condición no se dará. Lo que sí se puede llegar a conseguir es que sean muy parecidas, haciendo que la ecuación de divergencia del método usado, (4) para distancia Euclidea, (5) para Kullback-Leibler y (6) para Itakura-Saito, se minimice, haciendo que tienda a cero. En la Figura 13 se observa este hecho, donde se representa la curva de la divergencia de Kullback-Leibler para una señal de disparo concreta.

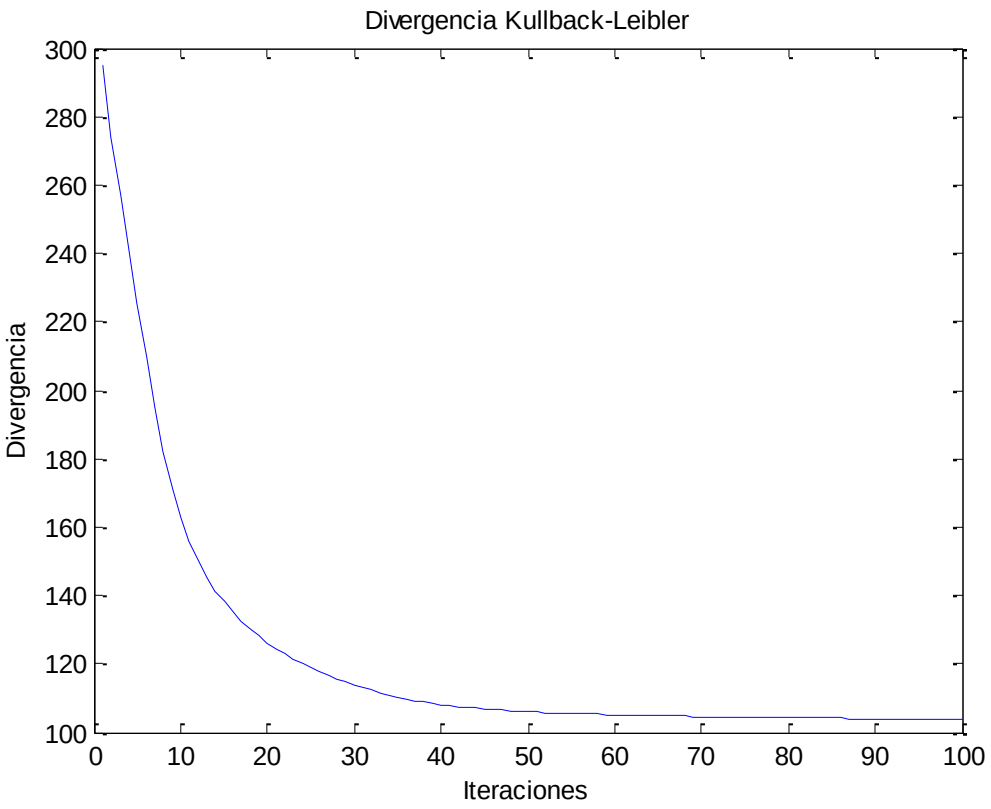


Figura13. Curva de divergencia Kullback-Leibler para una señal producida por una señal de un disparo del arma Vickers de calibre .303.

Para comenzar a desarrollar las ecuaciones de actualización de las matrices no negativas W y H, se tiene en cuenta que estas también deben estar normalizadas. Para normalizar estas matrices, se debe almacenar el peso de ambas matrices en otra de dimensión KxK, de manera que la señal reconstruida será

$$Y = WNH \tag{23}$$

N será la matriz de normalización. Debido a que W tiene una dimensión FxK, se debe normalizar respecto a columnas, quedando así la suma de la energía de las columnas de la matriz W en la unidad. Lo mismo ocurrirá con la matriz de activaciones H, pero en este caso se normaliza respecto a sus filas. Con esta normalización, la matriz N tendrá el peso de la matriz de información en sus columnas y el peso de la matriz de activación en sus filas, para así cumplir con lo establecido en (23). De esta manera, la actualización de dichas matrices en cada iteración viene dada por las matrices normalizadas W y H y la matriz que almacena el peso, la matriz N; esta última también se modifica en cada iteración y es inicializada como una matriz identidad de dimensión KxK.

$$\bar{W} = \frac{W}{\sqrt{\sum W^{\cdot(2)}}},\ \bar{H} = \frac{H}{\sqrt{\sum H^{\cdot(2)}}} \tag{24}$$

$$N \leftarrow \sqrt{\sum W^2} \cdot N, \quad N \leftarrow \sqrt{\sum H^2} \cdot N \tag{25}$$

La normalización es muy importante ya que se eliminan posibles indeterminaciones en las sucesivas iteraciones, dejando el valor de pesado de la función igual en cada iteración y así no acumular errores.

Puesto que se han tomado esta serie de consideracionesy atendiendo a lo explicado hasta ahora, a continuación se refleja cómo llegar a las ecuaciones de actualización de W y H. Se desarrollan las ecuaciones partiendo de cada tipo de divergencia discutida en el Capítulo 3 (4) (5) (6) en la sección 3.1.4. y tratando de minimizar estas divergencias (2). A partir de las ecuaciones citadas se obtienen las siguientes:

Atendiendo a Distancia Euclidea:

Se minimiza su función de costo (4), y se calcula el gradiente de esta para cada uno de los parámetros de factorización (W y H). En la ecuación (4) el parámetro '*y*' es sustituido por la salida esperada Y considerada en la ecuación (23), y el parámetro 'x' es sustituido por el espectro de la señal de entrada X. Así que se procede al cálculo del gradiente de la distancia Euclidea:

$$\frac{\partial D_{Euc}}{\partial W} = WN^2H^2 - XNH \tag{26}$$

$$\frac{\partial D_{Euc}}{\partial H} = HW^2N^2 - XWN \tag{27}$$

Teniendo en cuenta ahora que la actualización se consigue multiplicando el valor previo a la iteración del parámetro (W o H) por el cociente de la parte negativa entre la parte positiva del gradiente calculado en el paso anterior, llegamos a las siguientes ecuaciones considerando su comportamiento matricial, es decir, ajustando los términos de manera que encajen dimensionalmente hablando.

$$W_{FK} \leftarrow W_{FK} \circ \frac{X_{FT}H_{TK}^T N_{KK}}{W_{FK}H_{KT}H_{TK}^T N_{KK}^2} \tag{28}$$

$$H_{KT} \leftarrow H_{KT} \circ \frac{N_{KK} W_{KF}^T X_{FT}}{W_{KF}^T W_{FK} N_{KK}^2 H_{KT}} \tag{29}$$

Atendiendo a la Divergencia Kullback-Leibler, y siguiendo lo visto para Distancia Euclidea:

$$\frac{\partial D_{KL}}{\partial W} = \frac{WNH - X}{W} \tag{30}$$

$$\frac{\partial D_{KL}}{\partial H} = \frac{WNH - X}{H} \tag{31}$$

$$W_{FK} \leftarrow W_{FK} \circ \frac{X_{FT} \circ (W_{FK} H_{KT})^{\cdot(-1)} H_{TK}^T}{(W_{FK} H_{KT})^{\cdot(0)} H_{TK}^T N_{KK}} \tag{32}$$

$$H_{KT} \leftarrow H_{KT} \circ \frac{W_{KF}^T X_{FT} \circ (W_{FK} H_{KT})^{\cdot(-1)}}{N_{KK} W_{KF}^T (W_{FK} H_{KT})^{\cdot(0)}} \tag{33}$$

Y por último, atendiendo a la Divergencia Itakura-Saito:

$$\frac{\partial D_{IS}}{\partial W} = \frac{1}{W} - \frac{X}{W^2 NH} \tag{34}$$

$$\frac{\partial D_{IS}}{\partial H} = \frac{1}{H} - \frac{X}{H^2 NW} \tag{35}$$

$$W_{FK} \leftarrow W_{FK} \circ \frac{X_{FT} \circ (W_{FK} H_{KT})^{\cdot(-2)} H_{TK}^T}{(W_{FK} H_{KT})^{\cdot(-1)} H_{TK}^T N_{KK}} \tag{36}$$

$$H_{KT} \leftarrow H_{KT} \circ \frac{W_{KF}^T X_{FT} \circ (W_{FK} H_{KT})^{\cdot(-2)}}{N_{KK} W_{KF}^T (W_{FK} H_{KT})^{\cdot(-1)}} \tag{37}$$

El operador [°] hace referencia al producto de Hadamard, que consiste en el producto de dos matrices de la misma dimensión, donde cada elemento ij de la matriz solución es el resultado de la multiplicación de los elementos ij de las matrices origen. Igualmente ocurre con [$a^{\cdot n}$], cuyo resultado viene dado por a_{ij}^n. Los subíndices en cada matriz indica su dimensión y H^T y W^T hacen referencia a la transpuesta de las matrices H y W respectivamente.

La técnica NMF es capaz de reconstruir la señal espectral de entrada, en la Figura 15 se puede observar la reconstrucción espectral de la señal de disparo del arma M4A2 de calibre 5,6 para cada una de las divergencias vistas, cuyo espectro original aparece en la Figura 14.

El número de iteraciones empleado para llegar a $X \approx WH$ es de 1000, este al fin de cuentas será menor, ya que como se aprecia en la Figura 16, las divergencias convergen antes de llegar a las 1000 iteraciones.

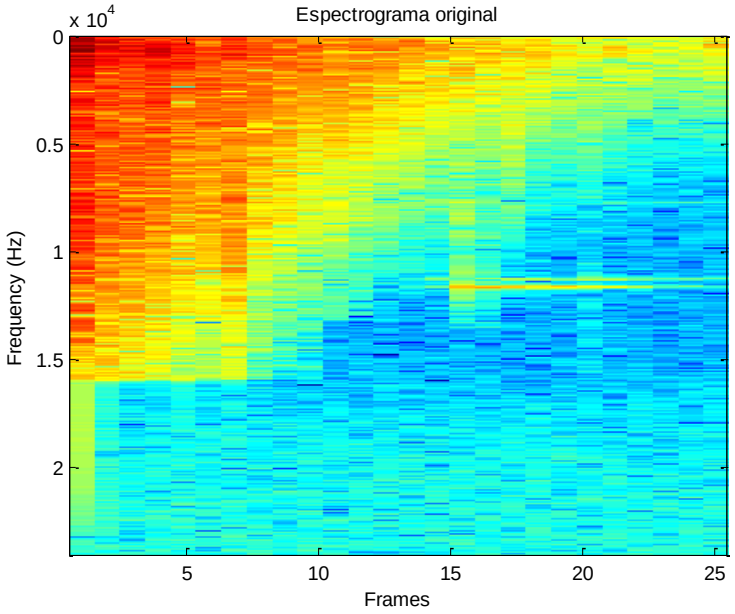
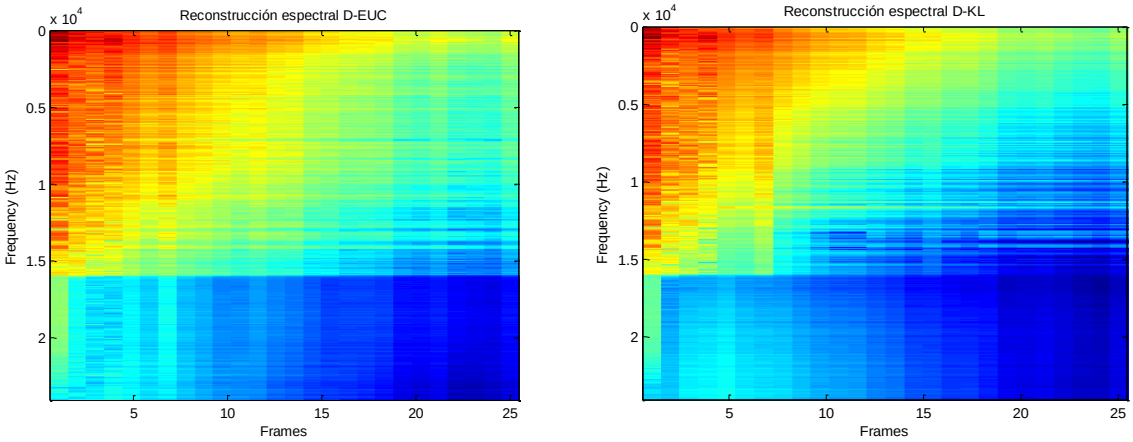
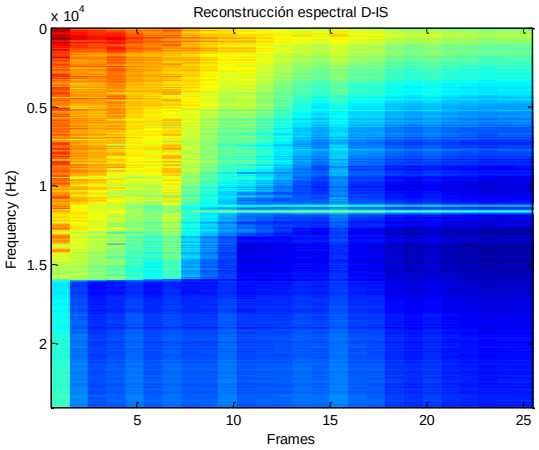


Figura 14. Espectro de la señal ocasionada por el disparo de una M4A2 de calibre 5,6.



a) b)



c)

Figura 15. A) Espectro de reconstrucción de la señal de disparo usando distancia Euclidea. B) Espectro de reconstrucción usando divergencia Kullback-Leibler. C) Espectro de reconstrucción de la señal de disparo usando divergencia Itakura-Saito.

En la Figura 16 donde se puede ver la convergencia de los distintos tipos de divergencia, se puede comprobar como 1000 iteraciones son demasiadas, de modo que se acabará optimizando este parámetro. Visto el comportamiento de la función de costo y el espectro de reconstrucción de la señal original, se comprueba que las tres responden bastante bien, pero la divergencia que minimiza la diferencia entre la señal original (X) y la reconstruida (Y) es Kullback-Liebler. Dicho esto, a partir de este punto se trabajará únicamente con la Divergencia Kullback-Leibler.

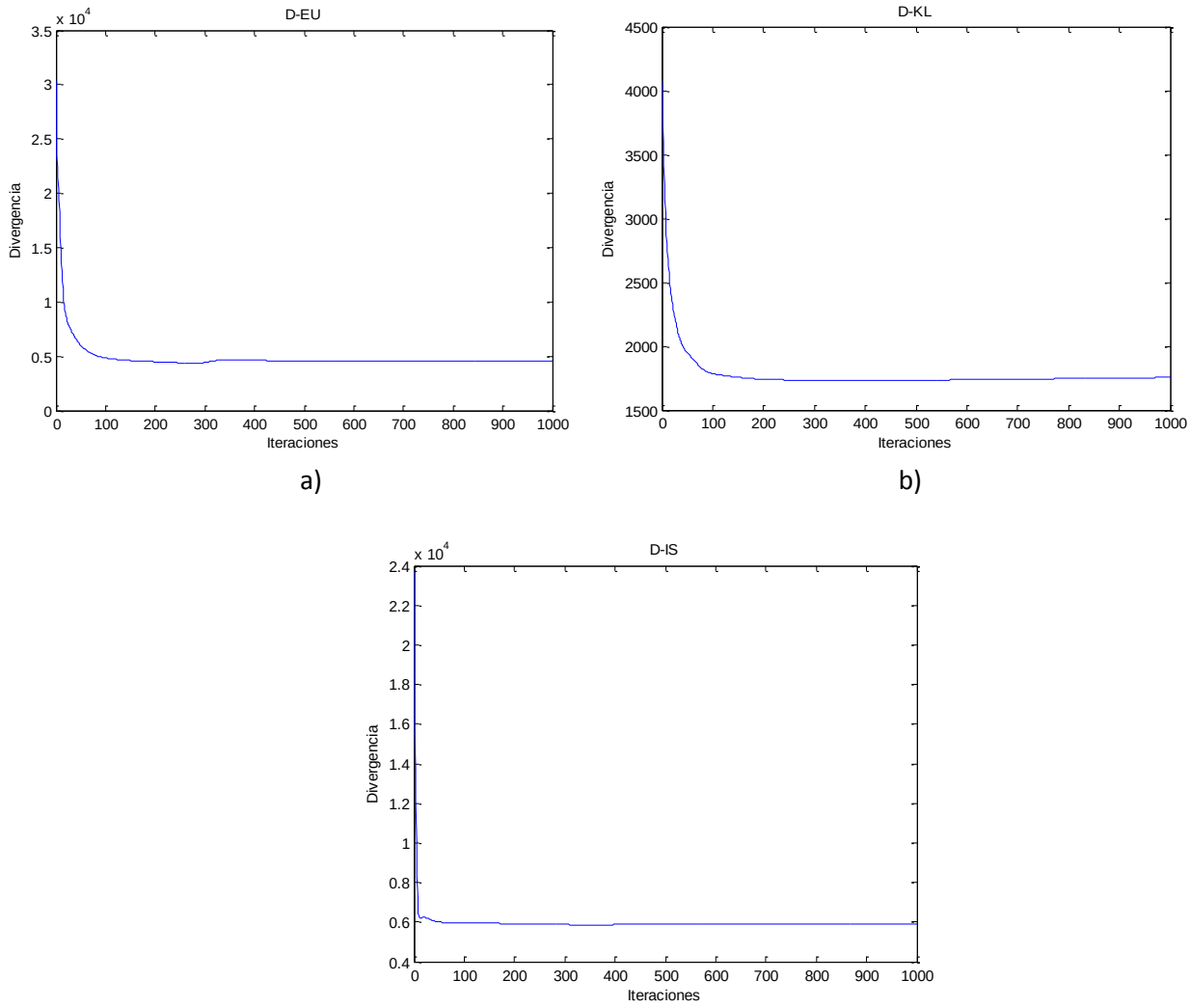


Figura 16. Gráficas de convergencia de las funciones de costo aplicada a una señal de disparo de una M42A de calibre 5,6: A) Distancia Euclidea. B) Divergencia Kulback-Leibler. C) Divergencia Itakura-Saito.

En la etapa de entrenamiento se realiza el proceso visto anteriormente para toda la base de datos de disparos de entrenamiento, aplicando las ecuaciones (32) y (33). De este modo, es necesario crear una matriz en la que se va a almacenar la información de los patrones espectrales de todos los disparos de la base de datos, de modo que estará formada por la matriz resultante W de cada señal de disparo. Ésta será usada en la etapa de separación, para así poder identificar una señal de disparo mezclada con ruido. Esta matriz que será reutilizada en la siguiente subetapa de detección se denominará a partir de aquí W_d. Como la base de datos está compuesta por 150 señales de disparo, la matriz W_d estará formada por 150 patrones espectrales y tendrá una dimensión de F x 150*K.

El número de iteraciones para actualizar los parámetros, junto con el amplio número de pistas a entrenar, hace que el algoritmo en Matlab tenga un coste computacional bastante alto para el objetivo buscado. De modo que, se implementa una función que hará que este coste computacional pueda verse reducido notablemente. La función usada consiste en reducir los bins del espectro de la señal de entrada, estableciendo un valor por el cual se van a dividir, de manera que la información en frecuencia se ve reducida, pero sin perder información útil del disparo, aligerando el cálculo de las matrices de factorización sin perjudicar en ningún momento al resultado final, ya que no es necesaria la reconstrucción de la señal de entrada temporal, solo se requiere la factorización de la señal del espectro de la entrada para llegar a las matrices de patrones espectrales y activación. Dicho esto, se debe tener en cuenta que esta reducción es importante pero no suficiente para su ejecución en tiempo real, este problema se solucionará en la etapa de implementación en tiempo real, donde debido a la capacidad de paralelización de la GPU se podrá conseguir. La compresión de los bins de frecuencia comentada seá igualmente usada en la subetapa de separación.

Como ya se ha mencionado, la matriz de información útil de disparo W_d, que será usada en la etapa posterior, se va a almacenar. Almacenando esta matriz no negativa no será necesario ejecutar este algoritmo más veces, ahorrando así el alto coste computacional que supone este proceso.

4.1.2. Etapa De Separación

Una vez se ha obtenido W_d en la etapa previa, que contiene la información de 150 disparos (material del que se compone la base de datos de disparos de entrenamiento), se pasa a la etapa de separación. El objetivo principal de esta etapa, es el poder identificar un disparo de una señal de entrada que estará formada por la señal deseada más ruido, como se indicaba en la ecuación (11).

Esta etapa consta de los mismos parámetros principales de entrada que la etapa de entrenamiento, como es un tamaño de ventana de 2048 muestras, uso de ventana Hamming y un overlap al 50% entre tramas.

Ya se especificó que, con los resultados obtenidos, se hacía más favorable realizar el proceso de detección siguiendo como partida la divergencia kullback-Leibler, de modo que, el desarrollo de las ecuaciones base de este algoritmo de etapa de separación partirán de la optimización de la función de costo mostrada en apartados anteriores de esta divergencia. La optimización consiste en el cálculo del gradiente (30) (31), pero en este caso, se debe tener presente que la señal de entrada vendrá dada por un disparo

más ruido, hecho que hace que las ecuaciones de actualización desarrolladas para la etapa de entrenamiento no sean las mismas que las que se necesitarán en esta etapa, ya que:

$$Y = W_d N_d H_d + W_n N_n H_n \tag{38}$$

W_n es la matriz no negativa que contendrá la información de ruido, H_n es la matriz de activación del ruido y N_n y N_d las de normalización de ruido y disparo respectivamente. Por otro lado se tiene H_d, matriz de activación del disparo, es la matriz que nos interesa conocer, ya que se trata del indicador de la presencia de disparo en la señal captada.

En esta etapa, la dimensión de las citadas matrices dependen de la dimensión de W_d e igual que antes, de la dimensión FxT de la señal de entrada. Las bases K tomadas en la etapa anterior serán las mismas, pero en este caso, la dimensión de la matriz H_d no vendrá dada por KxT, sino que se debe tener en cuenta que la matriz de información de disparos obtenida en la etapa de entrenamiento está formada por 150 señales de disparos aislados, así que H_d estará compuesta por K*150 filas y T columnas. Del mismo modo se verán afectadas el resto de matrices, siendo ahora la dimensión de las matrices de normalización 150*K x 150*K. Dicho esto, se denominará a partir de este punto el valor K'=K*150.

Se procede ahora al desglose de las ecuaciones hasta llegar a las actualizaciones de las matrices en las que se factoriza la señal de entrada. Como partida, recordar la función de costo de la divergencia Kullback-Leibler (5), e igual que en la etapa de entrenamiento, se optimizará dicha función y como se ha venido haciendo, para actualizar cada parámetro, se multiplica su valor, calculado en la iteración previa, por el cociente de la parte negativa entre la positiva de la derivada parcial calculada para cada parámetro. A continuación se refleja lo explicado:

$$D_{KL} = X\, log \frac{X}{W_d N_d H_d + W_n N_n H_n} - X + W_d N_d H_d + W_n N_n H_n \tag{39}$$

$$\frac{\partial D_{KL}}{\partial H_d} = W_d N_d - \frac{X W_d N_d}{W_d N_d H_d + W_n N_n H_n}, \quad \frac{\partial D_{KL}}{\partial W_n} = N_n H_n - \frac{X N_n H_n}{W_d N_d H_d + W_n N_n H_n},$$

$$\frac{\partial D_{KL}}{\partial H_n} = W_n N_n - \frac{X W_n N_n}{W_d N_d H_d + W_n N_n H_n} \tag{40}$$

Se debe tener en cuenta a la hora de establecer las ecuaciones de actualización que se realice siguiendo el correcto dimensionado de los parámetros.

$$H_d \leftarrow H_{d_{K'T}} \circ \frac{W_{d\,K'F}^T (X_{FT} \circ (W_{d_{FK'}} N_{d_{K'K'}} H_{d_{K'T}} + W_{n_{FK'}} N_{n_{K'K'}} H_{n_{K'T}})^{\cdot(-1)})}{W^T{}_{d_{K'F}} (W_{d_{FK'}} N_{d_{K'K'}} H_{d_{K'T}})^{\cdot(0)}} \tag{41}$$

$$W_n \leftarrow W_{n_{FK'}} \circ \frac{X_{FT} \circ (W_{d_{FK'}} N_{d_{K'K'}} H_{d_{K'T}} + W_{n_{FK'}} N_{n_{K'K'}} H_{n_{K'T}})^{\cdot(-1)} H_{n_{TK'}}^T}{(W_{n_{FK'}} N_{n_{K'K'}} H_{n_{K'T}})^{\cdot(0)} H_{n_{TK'}}^T} \tag{42}$$

$$H_n \leftarrow H_{n_{K'T}} \circ \frac{W_{n\,K'F}^T (X_{FT} \circ (W_{d_{FK'}} N_{d_{K'K'}} H_{d_{K'T}} + W_{n_{FK'}} N_{n_{K'K'}} H_{n_{K'T}})^{\cdot(-1)})}{W^T{}_{n_{K'F}} (W_{n_{FK'}} N_{n_{K'K'}} H_{n_{K'T}})^{\cdot(0)}} \tag{43}$$

Vistas las ecuaciones de actualización de los parámetros fundamentales, sólo queda señalar que las matrices N de normalización se calculan igual que en la etapa anterior, tal y como se muestra en (25) e igualmente, estos parámetros deben estar normalizados en sus iteraciones como se ve en (24).

Con lo que respecta al número de iteraciones, en esta etapa se ha establecido 70 iteraciones, ya que como se vio en la etapa de entrenamiento, la función de costo converge sobre este valor. De todas formas, en la Figura 18 se puede comprobar como con 70 iteraciones es más que suficiente para que la divergencia converja.

Para analizar con detalle lo que se consigue con el algoritmo de detección de un disparo realizado con el método NMF, se ha considerado un caso específico, donde se ha empleado una señal de entrada formada por el disparo de un arma semiautomática Cz 75 de 9mm mezclada con ruido de calle, con una relación señal a ruido de 5 dB. Como punto de partida del algoritmo se cuenta con el espectro de la señal de entrada normalizado (X_{norm}) y con la matriz W_d obtenida en la etapa de entrenamiento. A partir de estos datos se van a realizar las operaciones vistas en (41) (42) y (43) que parten de la divergencia Kullback-Leibler. Para comprobar si se detecta disparo, interesa analizar la

matriz no negativa H_d resultante. Ésta es la matriz de activación de las bases en las que acontece el disparo.

En la figura 17 se puede observar el resultado de la detección del disparo. En la figura 17.a) se observa el espectro de la señal de entrada, donde se distingue perfectamente las tramas donde se sitúa el disparo, en la figura 17.b) se muestra la reconstrucción espectral de la señal de entrada dada por $Y = W_d N_d H_d + W_n N_n H_n$, que como se puede comprobar son prácticamente iguales, de manera que la reconstrucción es buena usando la divergencia Kullback-Leibler. Por último, la parte de interés que va a informar si se detecta disparo es representada en las figuras 17.c) y 17.d), donde aparece la matriz de activación H_d y el sumatorio de ésta respectivamente. H_d activa las bases a causa de la detección del disparo, distinguiendose perfectamente donde se encuentra la señal de disparo, mientras que el ruido se ha conseguido separar.

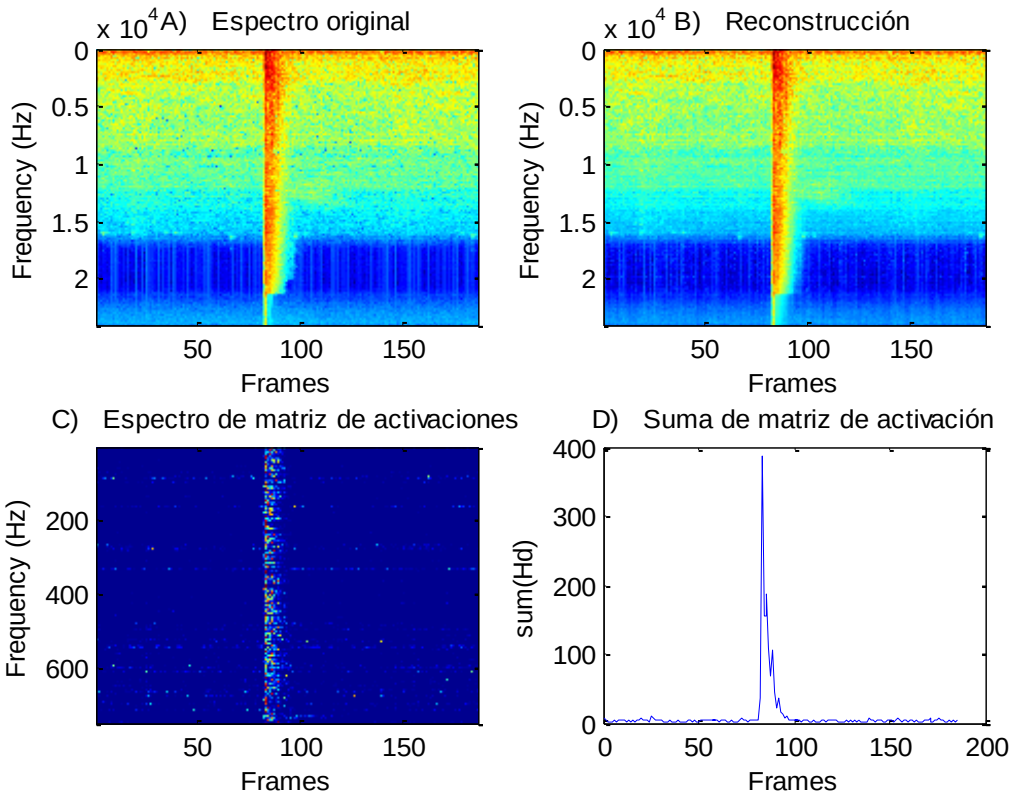


Figura 17. Señal de disparo de arma semiautomática Cz 75 de 9 mm mezclado con ruido de calle. A) Espectrograma de la señal original de entrada (X). B) Espectrograma de la señal reconstruida (Y). C) Espectrograma de H_d. D) Matriz de activación H_d.

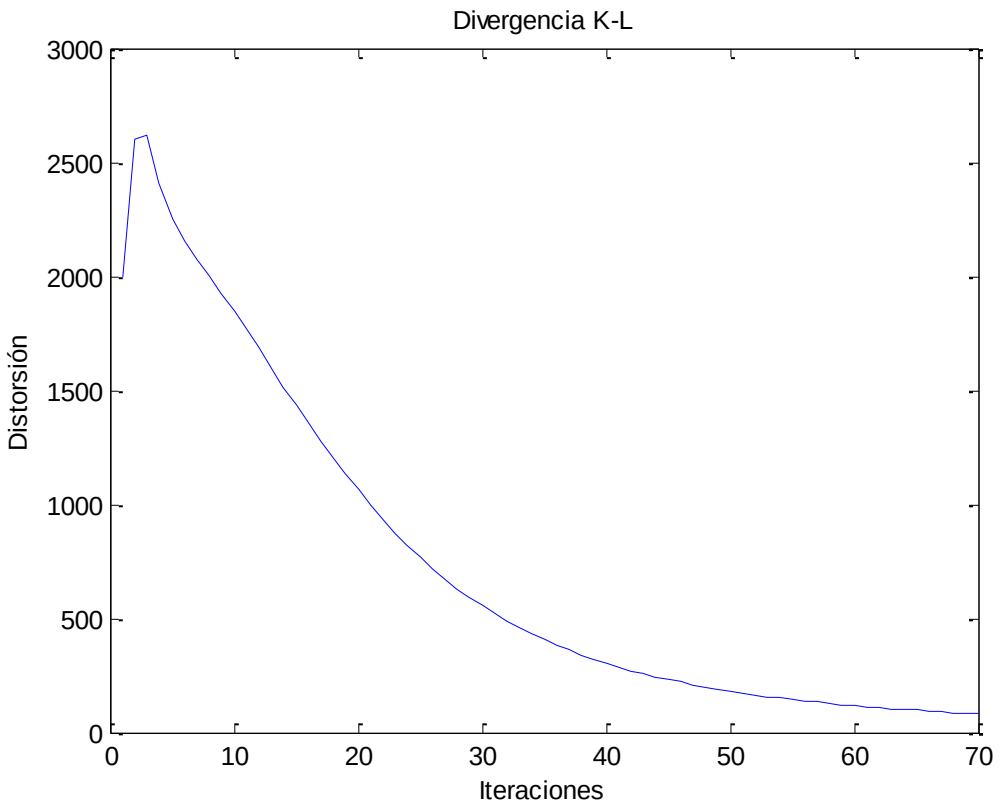


Figura 18. Convergencia de la Divergencia Kullback-Leibler para 70 iteraciones.

En la Figura 18 se puede ver como tiende a cero la función de costo en el caso citado, observándose la distorsión de la señal reconstruida con respecto a la original, donde se tiene como señal de entrada la mezcla del disparo de un arma semiautomática Cz 75 y ruido de calle.

El ploteado de la matriz H_d presenta demasiados máximos locales, que a la hora de encontrarse en una situación en la que la relación señal a ruido sea pequeña puede dar problemas en cuanto a la correcta identificación del disparo, para ello se propone una mejora. En la Figura 19 se representa la suma de la matriz de activación suavizada para el arma semiautomática Cz 75 de 9 mm. El suavizado se ha realizado con la función facilitada por Matlab 'smooth()', la cual realiza un promediado de diferentes subconjuntos de datos, minimizando con ello los máximos locales, aunque también el máximo absoluto (el disparo en cuestión) se verá rebajado por el suavizado. Con esto se consigue minimizar picos que puede causar el ruido, dejando así mucho más claro las bases activadas por el disparo.

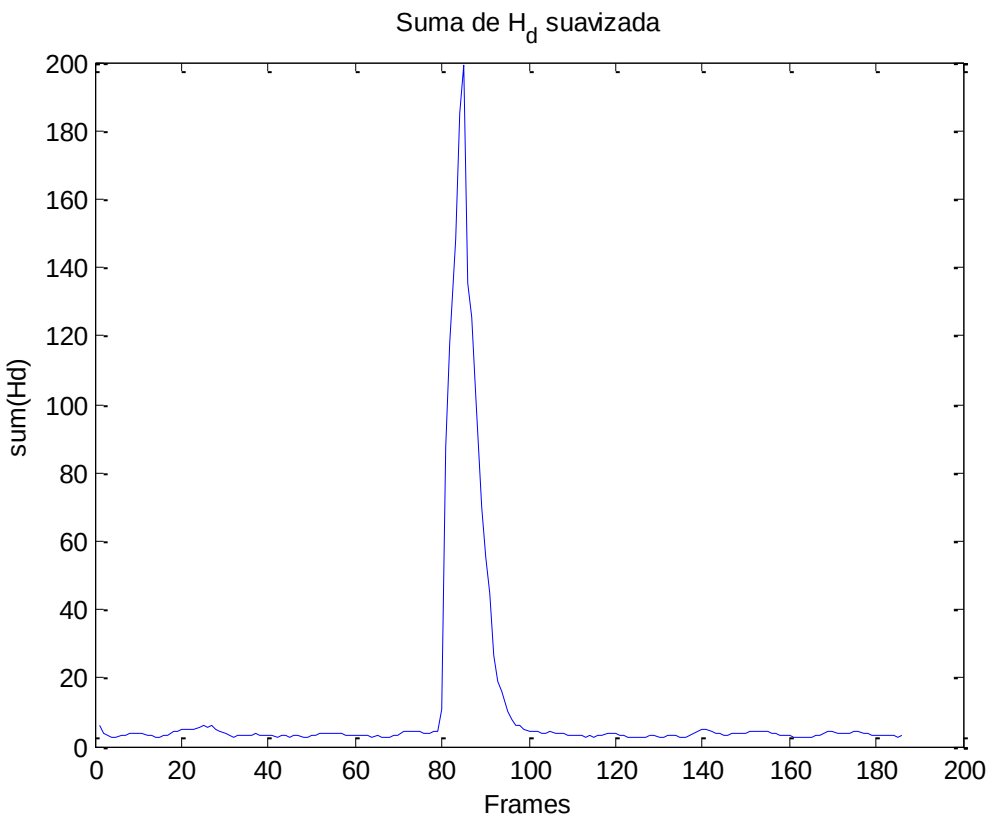


Figura 19. Suma de la matriz de activación con la aplicación de suavizado.

SPARSE

Otro punto importante a tener en cuenta en cuanto a la posible mejora del algoritmo de detección del sonido producido por un arma de fuego es el uso de *SPARSE*, basándose en la referencia [14]. El *SPARSE*, como ya se explicó en el Capítulo 3, pretende hacer que el algoritmo empleado devuelva soluciones más restrictivas en cuanto al número de bases activas, identificadas como disparo. Con esto se pretende conseguir, a priori, una mayor robustez frente al ruido, haciendo al algoritmo menos permisivo a la hora de identificar la señal deseada, en este caso el disparo. Esta técnica consiste en incluir un parámetro en la matriz de activaciones que consiga hacer más estricta la matriz H_d en cuanto a la identificación de la información que le llega, pudiendo ser disparo o ruido, limitando el número de bases activas a causa del disparo, incurriendo así, en principio, en menos errores de detección, causados por la no detección de un disparo o por la detección de un disparo sin haberlo.

Explicado la utilidad de *SPARSE* en este trabajo, a continuación se detalla en qué consiste exactamente el parámetro añadido en la matriz de activación. Este parámetro será empleado tanto en la etapa de entrenamiento como en la etapa de separación,

haciendo una posterior evaluación de los resultados obtenidos que se podrá ver en el Capítulo 5. Se trata pues de un parámetro λ, que toma valores comprendidos entre 0 y 1, donde el valor λ=1 es el más estricto en la activación de bases. En las siguientes ecuaciones se refleja dónde va a aparecer λ, tanto en la etapa de entrenamiento como en separación, al que se le dará diferentes valores entre 0 y 1 para así evaluar que es el que mejores resultados ofrece.

$$\lambda_d = \lambda \cdot F/K \tag{44}$$

$$H_{KT} \leftarrow H_{KT} \circ \frac{W_{KF}^T X_{FT} \circ (W_{FK} H_{KT})^{\cdot(-1)}}{N_{KK}\, W_{KF}^T (W_{FK} H_{KT})^{\cdot(0)} + \lambda_d} \tag{45}$$

$$H_d \leftarrow H_{d_{K'T}} \circ \frac{W_{d\ K'F}^T (X_{FT} \circ (W_{d_{FK'}} N_{d_{K'K'}} H_{d_{K'T}} + W_{n_{FK'}} N_{n_{K'K'}} H_{n_{K'T}})^{\cdot(-1)})}{W^T{}_{d_{K'F}} (W_{d_{FK'}} N_{d_{K'K'}} H_{d_{K'T}})^{\cdot(0)} + \lambda_d} \tag{46}$$

En la ecuación (44) se indica que se tiene que multiplicar el parámetro λ por el número de bins F y dividirlo por el número de bases K prefijadas, necesario para que tenga un peso similar al coste de la divergencia Kullback-Leibler y pueda tener influencia en ella, ya que sumar simplemente una unidad no surtirá ninguna mejora. Dicho esto, los valores que se han usado en esta memoria para λ son 0, 0.3, 0.5, 0.7 y 1.

En la Figura 20 se representa el espectro de las bases activas cuando se emplea un sparse de λ=0 para ambas etapas, mientras que en la Figura 21 se emplea un sparse de λ=1 también en ambas etapas. En ambos casos, donde la señal de entrada tiene una SNR=5 dB, se observa que los resultados obtenidos son muy buenos, ya que el disparo es claramente diferenciado activando así las bases correspondientes, aunque por otro lado se puede apreciar como el uso de sparse, al disminuir las bases activas, controla aún más el ruido, disminuyendo la activación de bases por su causa. En la evaluación que más adelante se hará del sistema, se realizará un análisis exhaustivo del empleo del sparse, sacando resultados estadísticos para comprobar si verdaderamente existe tal mejora.

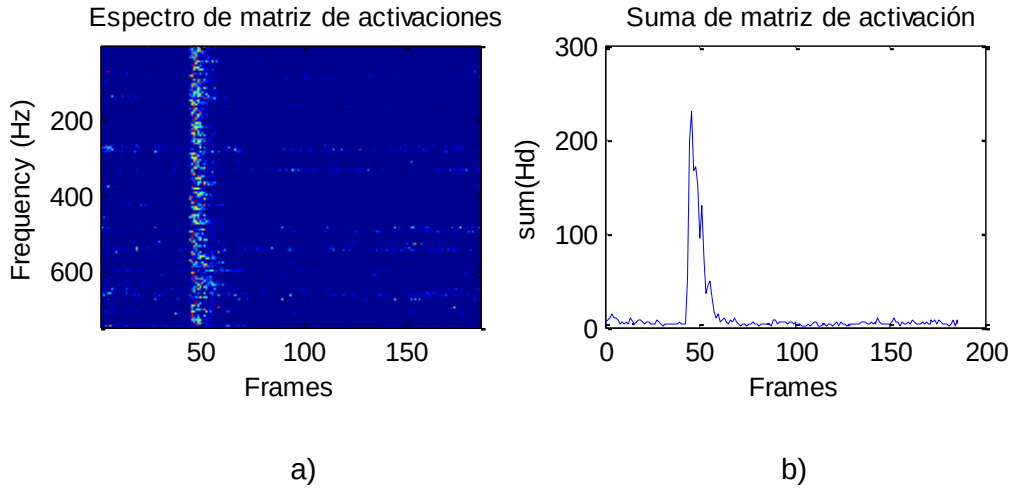


Figura 20. A) Matriz de activaciones y B) Sumatorio de matriz de activaciones para una señal producida por un arma semiautomática de tipo Beretta 92 Fs de 9mm. Sparse con $\lambda=0$ para ambas etapas.

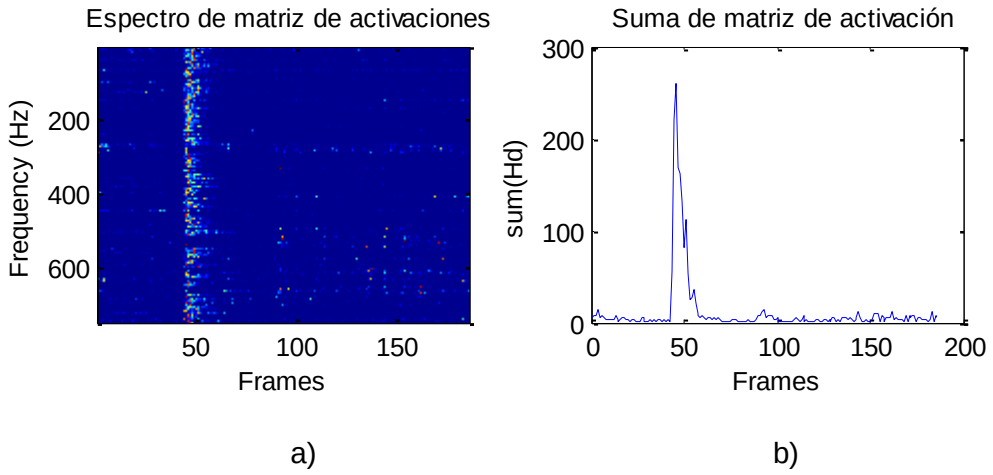


Figura 21. A) Matriz de activaciones y B) Sumatorio de matriz de activaciones para una señal producida por un arma semiautomática de tipo Beretta 92 Fs de 9mm. Sparse con $\lambda=1$ para ambas etapas.

Por último indicar, que en esta etapa, al igual que en la etapa de entrenamiento, justo después del cálculo de la STFT de la señal de entrada se recurrirá compresión de los bins de frecuencia que permite aligerar el coste computacional.

4.2. Implementación de etapa de localización

La etapa de localización es aquella que indicará la situación espacial con respecto a una referencia en la que se halla el arma de fuego que origina el disparo, se indica su

dirección pero no la posición en tres dimensiones. Esta etapa, por tanto se va a activar cuando se haya detectado que se ha producido el disparo, función de la etapa de detección. De esta manera, una vez se ha capturado la señal y se ha estimado que se trata de un disparo, entra en funcionamiento la segunda etapa que, mediante una serie de métodos que se van a describir a continuación, informará de la localización del arma expresada en grados, proceso que se describe en la Figura 22.

La captura de la señal se realiza con un array circular compuesto por seis micrófonos omnidireccionales los cuales irán conectados a un interfaz de audio que se encargará de recoger las pistas grabadas por estos seis micrófonos con ayuda de un software de captura de audio. El array circular utilizado tiene un diámetro de un metro y los micrófonos omnidireccionales se colocan formando un hexágono regular. Los cálculos en los que se van a basar la localización de la fuente sonora cuentan con que el disparo se va a producir siempre condiciones de campo lejano, para así poder considerar la onda sonora que llega a cada micrófono como una onda plana, para esto la distancia de la fuente a los micrófonos debe ser mayor que 10 veces la distancia entre cada par de micrófonos (la distancia máxima entre un par de micrófonos vendrá dada por los micros enfrentados y será 1 metro, el diámetro considerado).

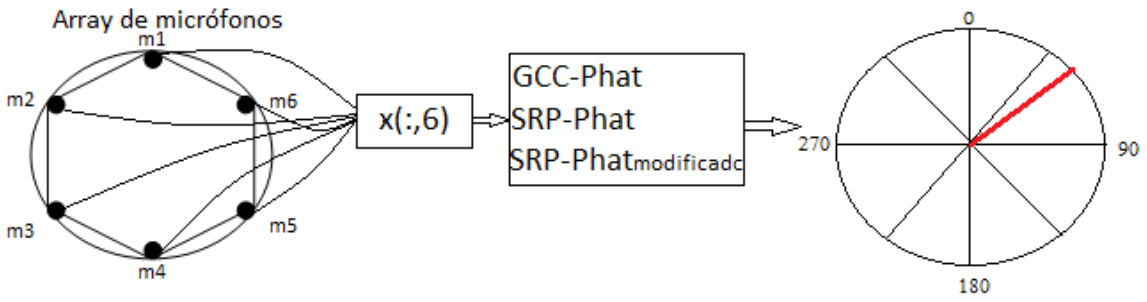


Figura 22. Esquema general de la etapa de localización de la fuente sonora originada por un arma de fuego.

El proceso de localización de una fuente sonora se realiza mediante algoritmos que estiman la diferencia de tiempos de llegada de la señal a cada par de micrófonos del array, compuesto por dos o más micrófonos. Esta técnica se denomina TDOA (Time Difference of Arrival). Por otro lado, se han ideado técnicas que no consisten en la estimación de TDOA entre pares de micrófonos, sino que la localización se realiza dividiendo el espacio en un conjunto discreto de posiciones candidatas en las que puede hallarse la fuente sonora y se elige aquella posición que maximiza un parámetro establecido, estas técnicas se denominan beamforming. Ambas ofrecen buenos

resultados, sin embargo ante reverberaciones y ruido la técnica del beamforming es más robusta pero más cara computacionalmente hablando.

La localización usando TDOA resulta sencilla. Suponiendo el caso en el que se cuente con dos micrófonos y que la distancia entre ellos sea mucho menor que la distancia a la fuente (condiciones de campo lejano), se puede comprobar cómo la señal que llega a cada micrófono es la misma pero llega en diferentes instantes de tiempo (13), debido a las diferentes posiciones espaciales que toman los micrófonos. Sabiendo esto se podrá determinar la diferencia de tiempo con que ambos micrófonos reciben la señal, variable que se denominará τ. Conocido el retardo entre micrófonos y la distancia que hay entre ellos, resulta fácil llegar al cálculo de la dirección de llegada DOA (Direction of Arrival) de la fuente sonora. Para comenzar, la distancia adicional que debe recorrer la señal para llegar al segundo micrófono puede ser conocida, en la Figura 7 ilustrada en el Capítulo 3 se aprecia un esquema que detalla la geometría en la que se basa el array de dos micrófonos, así que conocida la geometría del array y el TDOA, se obtiene la dirección de llegada de la fuente sonora producida por el disparo:

$$\theta = cos^{-1}(\frac{c\tau}{d}) \tag{47}$$

En la ecuación anterior τ hace referencia al retardo entre micrófonos, d a la distancia entre ellos, c a la velocidad del sonido y θ a la dirección de la fuente expresada en radianes. De modo que con esto queda clara la relación existente entre TDOA y DOA. La Figura 7 aclara por completo de donde se obtienen estos parámetros. Está sencilla técnica obtendrá muy buenos resultados siempre y cuando no se tenga reverberación, que no es considerada en ningún momento. Como es lógico, colocando un mayor número de micrófonos se obtendrá mejores resultados que con solamente dos, ya que en este caso se crea ambigüedad, debido a que un ángulo -θ y θ produce el mismo TDOA. En los métodos a desarrollar se va a emplear arrays circulares compuestos por 6 micrófonos, de manera que desaparece la ambigüedad comentada.

Esta técnica explicada es realizada en diversos métodos de localización existentes, como es el CC, el GCC o el método GCC-Phat, implementado en este Trabajo Fin de Grado, que es un caso concreto del GCC. Todos estos métodos pretenden realizar el cálculo del TDOA, que es desconocido. Para dicho cálculo, se va a realizar una evaluación en el intervalo $\tau \in [-\tau_{max}, \tau_{max}]$, donde el retardo máximo se calcula como se

indica en la ecuación (48). En estos métodos, el TDOA vendrá dado por el máximo de la correlación cruzada de las señales de los pares de micrófonos. La correlación cruzada es una medida de la similitud entre las dos señales, mediante la cual podemos encontrar características relevantes en una señal por medio de la comparación con la otra señal conocida, esta característica será el retardo entre micrófonos. En la ecuación (15) se refleja este proceso, realizado por el método CC, mientras que el método GCC supone una mejora al anterior, ya que se hace más robusto al trabajar en frecuencia, como se muestra en las ecuaciones (16) y (17). La implementación del método GCC-Phat será la que se desarrolle en la siguiente sección, tratándose éste de un caso concreto del GCC.

$$\tau_{max} = \frac{d}{c} fs \tag{48}$$

Por otro lado, como ya se ha comentado, la localización de la fuente sonora también se puede realizar siguiendo técnicas de beamforming. Se puede encontrar aplicaciones de esta técnica en radar, sonar, comunicación inalámbrica, sismología, acústica, biomedicina, etc. El beamforming consiste en un filtrado espacial, combinando las diferentes señales capturadas por el array de sensores de tal manera que se seleccione o estime una señal de interés procedente de una dirección determinada, atenuando el resto. En el caso de conocer la localización de la señal de interés, el beamformer se orienta en esa dirección, mejorando la señal de interés y atenuando el resto de la señal captada; pero el caso que se plantea es que no se conoce la localización de la fuente sonora de interés, de modo que aquí el beamformer se comportaría como localizador, escaneando la región del espacio donde podría encontrarse la fuente y eligiendo aquella que maximice un parámetro. De esta manera, el array de micrófonos se comportará como un beamformer. Dicha técnica es usada por el método de localización SRP-Phat, que más adelante se desarrollará por completo, que está estrechamente relacionado con el GCC-Phat, pero incorporando el concepto de beamforming, lo que lo hace aún más robusto frente a reverberaciones.

4.2.1. GCC-Phat

Es un método que tiene su base en el GCC (Generalized Cross-Correlation), donde en la ecuación (17) se ve su definición, y su función de peso viene dada por la transformada de fase (PHAT), $W_{PHAT}(t,f) = |X_1(t,f)X_2^*(t,f)|^{-1}$.

Este método estima el TDOA entre un par de micrófonos que forman parte del array calculando para ello la correlación cruzada entre dos señales en el dominio de la frecuencia, mientras que su método predecesor (CC) realiza esta correlación en tiempo. Así que dadas las señales de entrada para cada micrófono x_m con m ϵ {1,2}, que son iguales pero con instantes de llegada diferentes, se realizará la Transformada Discreta de Fourier (DFT) de las dos, se calculará la correlación entre ambas y se tendrá en cuenta el pesado estimado, Phat, por el método GCC-Phat. En las siguientes ecuaciones se define este proceso:

$$x_2(n) = x_1(n - \tau_s) \tag{49}$$

$$X_1(k) = \sum_{n=0}^{N-1} x_1(n) e^{-j2\pi\frac{kn}{N}},\ X_2(k) = \sum_{n=0}^{N-1} x_1(n) e^{-j2\pi\frac{kn}{N}(1+\tau_s)} \tag{50}$$

donde $x_1(n)$ y $x_2(n)$ son las señales de entrada de los micrófonos 1 y 2 y están formadas por la señal de interés (el disparo) y ruido, N es el número de muestras que se tomará para realizar la transformada, en este caso se toma 2048 muestras, ya que se agilizarán los cálculos ajustándose a la duración de un disparo y se obtendrán buenos resultados. En cuando a τ_s hace referencia al retardo de la señal sonora producida por el disparo que hay entre el micro uno y dos, tomando el primero como referencia.

$$GCC_{Phat} = Mask \cdot (X_1(k) \cdot X_2^*(k)) \tag{51}$$

$$R_{x_1x_2}(n,\tau) = \Re\left(IDFT\left(\frac{GCC_{Phat}}{W_{Phat}}\right)\right) = \Re\left(IDFT\left(\frac{GCC_{Phat}}{|X_1(k) \cdot X_2^*(k)|}\right)\right) =$$

$$= \Re\left(IDFT\left(\frac{Mask \cdot X_1(k)^2 e^{j2\pi k\tau_s}}{|X_1(k)^2|}\right)\right) = \Re\left(IDFT(Mask \cdot e^{j2\pi k\tau_s})\right) \tag{52}$$

Se obtiene de la correlación, como era de esperar, el retardo producido entre el micrófono uno y dos. Este proceso se debe realizar para cada par de micrófonos, más adelante se explicará cómo. En la ecuación (51) se observa que se ha introducido un nuevo concepto al que se ha denominado 'Mask'. Se trata de una máscara que se va a aplicar al cálculo de la correlación, dicha máscara consiste en centrar el cálculo del

retardo para aquellas frecuencias en las que la señal de interés a buscar se va a encontrar. La señal producida por un disparo, como sabemos, se trata de una señal de banda ancha, conteniendo información útil tanto en baja como en alta frecuencia, así que para el caso dado no se va a aplicar dicha máscara, para así tener información de todo el espectro de la señal. Distinto hubiera sido si se quisiera localizar la fuente sonora producida por voz humana, en ese caso si se hubiera recurrido al uso de dicha máscara, la cual tendría un límite inferior en los 80 Hz y un límite superior sobre los 5 KHz, para no suprimir en ningún momento información útil, creando así un filtro paso banda.

La DFT puede ser calculada de forma eficiente en la práctica utilizando el algoritmo de la transformada rápida de Fourier, FFT (Fast Fourier Tansform), método que ha tenido gran relevancia en la informática. Para la evaluación directa de la fórmula de la Transformada Discreta de Fourier se necesitan O(n^2) operaciones aritméticas, mientras que para la realización del algoritmo FFT se requieren O($n \log n$) operaciones. En Matlab es la función fft() la que implementa dicho algoritmo, introduciéndose como valores de entrada la señal temporal y el número de muestras que se toma para realizar la transformada, que va a coincidir con el tamaño de ventana considerado (2048 muestras). Análogamente, para la realización de la transformada inversa IDFT, existe otra función en Matlab que implementa la transformada inversa rápida de Fourier, esta es ifft().

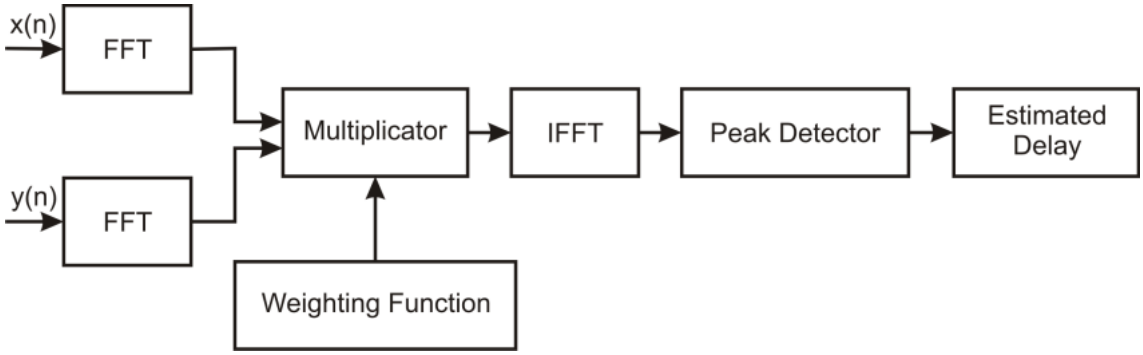


Figura 23. Esquema general del método GCC-Phat.

A modo de resumen, en la Figura 23 se detalla los pasos a realizar para obtener el retardo de interés, donde x(n) es la señal de entrada por uno de los canales e y(n) es la señal de entrada por otro micrófono.

La obtención del TDOA se realiza llevando a cabo una evaluación de retardos gracias a $R_{x_1x_2}(n,\tau)$ comprendida en el intervalo $\tau \in [-\tau_{max}, \tau_{max}]$, donde como se dijo en la ecuación (48), τ_{max} viene dado por la separación entre micrófonos y la frecuencia de muestreo de las señales de entrada. La correlación cruzada $R_{x_1x_2}(n,\tau)$ se hará máxima cuando dentro de dicha evaluación $\tau = \tau_s$, donde τ_s, es el retardo en uno de los

micrófonos de la señal de interés. Un τ_s positivo indica que la fuente sonora se localiza más próxima al micrófono primero y por tanto llega más tarde al segundo, mientras que si es negativo indica que se sitúa más cerca del segundo micrófono. De modo, que simplemente se debe buscar ese valor máximo en la matriz $R_{x_1x_2}$, donde se van a almacenar dichos retardos. Una vez calculado este retardo en muestras, se debe de establecer una relación que devuelva como solución una posición en grados, con el centro del array de micrófonos como referencia. Para ello se va a realizar una matriz de retardos, la cual va a ser la encargada de establecer esta relación. Esta matriz no solo va a ser usada para el algoritmo de GCC-Phat, sino también para el de SRP-Phat y el SRP-Phat modificado.

En primer lugar, antes de especificar en qué consiste la matriz que relaciona la correlación de cada par de micrófonos con la localización de la fuente con respecto al array como referencia, se debe explicar cómo se disponen los micrófonos en dicho array circular y las combinaciones que se tendrán de todos ellos. Al comienzo de este apartado 4.2. se especificó que el array estaría compuesto por 6 micrófonos, distribuidos de tal manera que formen un hexágono regular, de modo que la separación en grados que hay entre cada micrófono consecutivo es de 60°. Se debe pues, establecer todas las relaciones posibles entre los micrófonos para formar parejas, a las cuales se les realizará la correlación cruzada descrita en las ecuaciones anteriores (51) y (52) propia del GCC-Phat. De modo que si se cuenta con 6 micrófonos, se tiene $\frac{M(M-1)}{2} = 15$ combinaciones posibles de pares de micros, así que se deben realizar quince correlaciones para buscar el retardo máximo en el resultado de estas. Para situar correctamente la referencia con nuestro array es importante indicar que, el micrófono 1 corresponde a la situación de 0°, el micrófono 2 a 60° y así sucesivamente. Para el cálculo de cada $R_{x_ix_j}$ se debe tener en cuenta la distancia entre los pares de micrófonos para los que se va a realizar la correlación, y esta distancia se puede expresar, con ayuda de la trigonometría, de manera general como:

$$dist_{ij} = diam \cdot sen\left((i-j)\frac{\alpha}{2}\right) \tag{53}$$

donde $diam$ es el diámetro del array circular considerado, i y j hacen referencia a los micrófonos que forman la pareja a relacionar, $i, j \in \{1, 2, 3, 4, 5, 6\}$ y α es el ángulo de separación de un micro con su consecutivo, 60° en el caso dado.

Ahora bien, una vez se ha calculado $R_{x_i x_j}$, resultado del método seguido GCC-Phat, bastaría con tomar el valor máximo, el cual nos indicará la posición de la fuente, pero esto no es tan sencillo, ya que esta información se debe traducir a información útil para poderla interpretar rápidamente en el caso que un sujeto se encuentre dentro del vehículo, para ello se debe establecer una relación entre el retardo máximo devuelto por la correlación y la localización en grados respecto al array de micrófonos que irá situado sobre el vehículo. A partir de aquí se puede explicar en qué consiste la matriz de retardos máximos y mínimos que facilita la comprensión de la información devuelta del GCC-Phat. A continuación se desarrolla un apartado en el que se cuenta su elaboración, y se debe tener presente para los posteriores métodos que se van a definir en los siguientes apartados dentro de este capítulo.

Matriz de retardos (Tau)

Esta función desarrollada en Matlab, denominada tau(), devuelve como resultado una matriz con información a usar por el GCC-Phat o cualquier otro método y así localizar gradualmente, de la manera adecuada, donde se sitúa la fuente sonora. Para ello se ha usado como referencia el artículo [36], donde igualmente hacen uso de un array circular.

En la ecuación (14) se calculaba el retardo de la señal conocida la distancia entre micrófonos y la localización de la fuente (en grados o radianes) respecto a la línea que une ambos micrófonos (Figura 7), pues bien, el objetivo es cambiar esta referencia dado que se cuenta con un array circular. Para ello, como se ve en [36], se calcula el retardo de una fuente sonora a un micrófono tomando el centro del array como referencia:

$$\tau_{s,m} = \frac{r}{c} \cos\left(\theta_s - \theta_m\right) \tag{54}$$

En esta ecuación se tiene que θ_m es el ángulo que forma el micrófono con el centro del array, de modo que según el micro que se elija, θ_m tendrá los siguientes valores:

$$\theta_m = \begin{cases} \theta_1 = 0° & \theta_4 = 180° \\ \theta_2 = 60° & \theta_5 = 240° \\ \theta_3 = 120° & \theta_6 = 300° \end{cases} \tag{55}$$

Continuando con la definición de la ecuación (54), se tiene que θ_s es el ángulo que forma la fuente sonora medido desde el centro del array. Por último indicar que r es el radio del array y c es la velocidad del sonido (se tomará 343 m/s, valor definido para condiciones normales de 20°C). En la Figura 24 se puede ver gráficamente lo descrito en estas ecuaciones. Considerando que el retardo se calcula para cada pareja de micrófonos según el GCC-Phat, en la ecuación (54) se debe de tener en cuenta este hecho, tomando así dos micrófonos en lugar de uno para el cálculo del retardo entre ellos con referencia al centro del array, resultando lo siguiente:

$$\tau_{s,ij} = \frac{r}{c}\cos\left[\cos\left(\theta_s - \theta_j\right) - \cos(\theta_s - \theta_i)\right] \tag{56}$$

donde se está considerando el retardo entre los micrófonos i, j para una determinada dirección de la fuente sonora (véase la Figura 24).

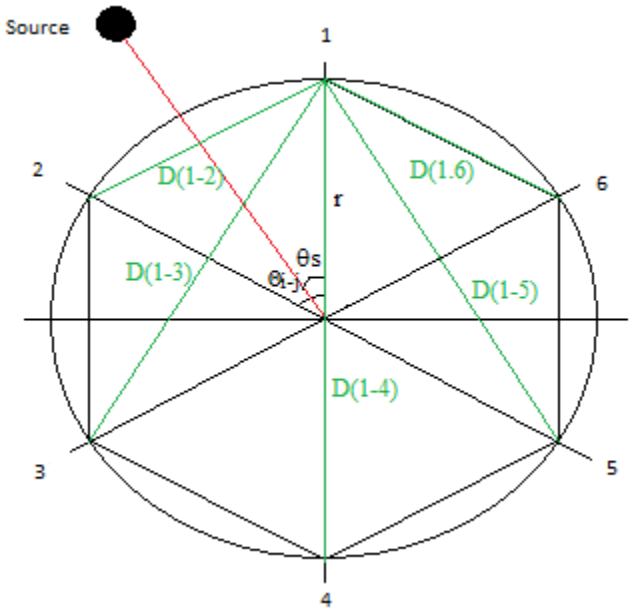


Figura 24. Representación geométrica del array de micrófonos.

En esta última imagen se puede ver, además del ángulo que forma la fuente sonora con respecto al array, donde el micrófono uno indica la posición de 0°, también las distancias entre el micrófono uno y los demás (D_{1-2}, D_{1-3}, D_{1-4} ...). Tal y como ya se mencionó habrá 15 combinaciones posibles de pares de micrófonos, y se tomarán las

distancias entre ellos correspondientes para el cálculo del GCC-Phat. Estas combinaciones son: 1-2, 1-3, 1-4, 1-5, 1-6, 2-3, 2-4, 2-5, 2-6, 3-4, 3-5, 3-6, 4-5, 4-6, 5-6.

El ángulo de la fuente sonora θ_s es desconocido. La matriz de retardos tau() implementada va a utilizar θ_s como parámetro de iteración, donde este pasará a adquirir un concepto de posible posición de localización.

Una vez se han comprendido estos conceptos geométricos, se realiza una evaluación de todas las posibles posiciones candidatas, θ_s, de la fuente sonora producida por el disparo, que van desde los 0° a los 359° y para todas las combinaciones de pares de micrófonos posibles, obteniéndose así la matriz denominada 'retardos' la cual almacena los $\tau_{s,ij}$ para cada caso. Será útil que la información de estos retardos se dé en muestras, para ello simplemente se multiplica $\tau_{s,ij}$ por la frecuencia de muestreo de las señales de entrada.

En resumen, la matriz de retardos tau(), consiste en una matriz que da información sobre los retardos en muestras que existen entre cada par de micrófonos para direcciones candidatas de la fuente en todo el espacio.

Ahora es el momento de ver donde se maximiza la correlación cruzada que determina el retardo para cada par de micros. Para ello, se busca el máximo de la matriz $R_{x_i x_j}$ y se identifica el par de micrófonos que devuelve esa GCC máxima. Una vez hecho esto, se debe identificar la posición en grados de la fuente usando la información anterior y la función tau(), así que se hace una evaluación de la correlación $R_{x_i x_j}$ calculada para ese par de micros, para ello se toma de $R_{x_i x_j}$ la muestra que indique $\tau_{s,ij}$ (retardo de cada posición candidata, almacenado en la matriz llamada 'retardos') para cada una de las posiciones de la fuente candidatas. De manera que, tomando por ejemplo la posición candidata de la fuente 0°, se selecciona en la matriz de $R_{x_i x_j}$ de la pareja de micros de retardo máximo el valor que indique la posición de retardo dada por $\tau_{0°,ij}$. Este valor seleccionado será comparado con la posición candidata de 1°, 2°, 3° y así sucesivamente con todos los demás valores que se tomarán para cada una de las posiciones candidatas, de modo que se va a encontrar entre todos ellos un máximo, y este máximo indica cual de todas esas posiciones candidatas es la dirección de la que proviene el disparo.

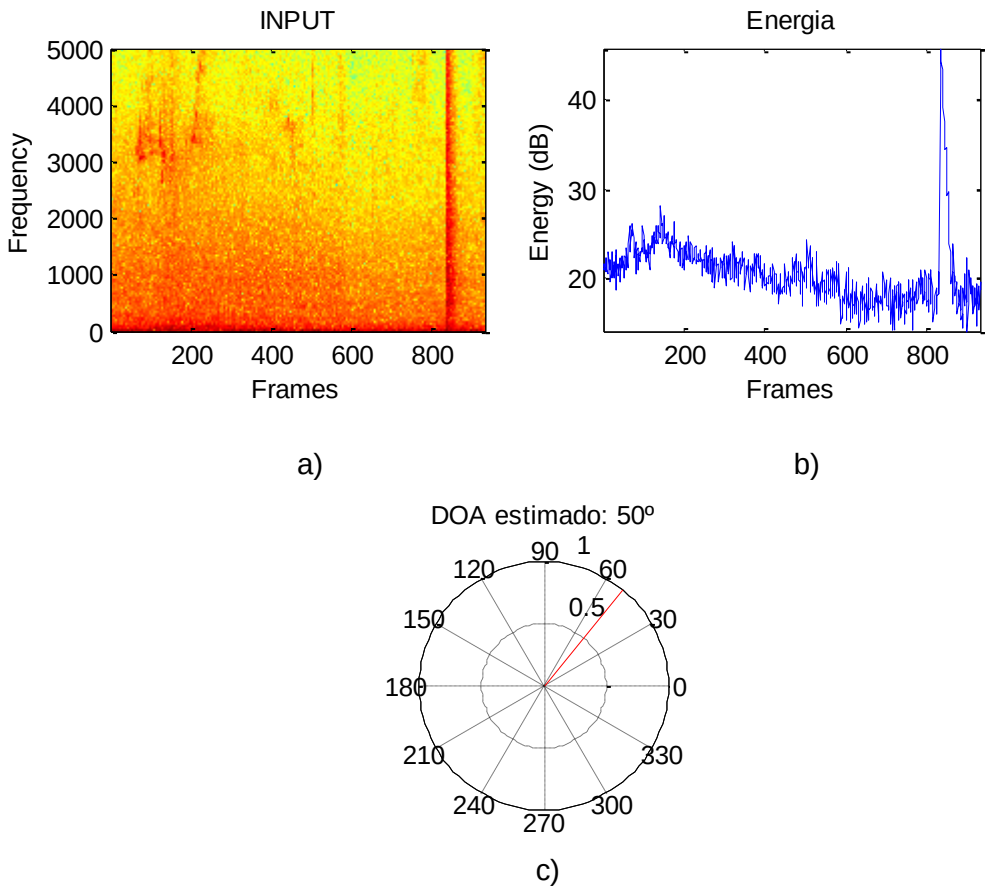


Figura 25. Localización de un disparo usando el método GCC-Phat. A) Espectrograma de la señal de entrada. B) Energía de la señal de entrada. C) Representación polar de la localización de la fuente sonora

En la figura 25 se aprecia como el espectro de la señal de entrada, tomada únicamente de un micrófono, presenta las características propias de una señal impulsiva como se vio en la anterior etapa de detección del disparo. En este ejemplo, se ha usado como entrada las señales capturadas por 6 micrófonos cuando la fuente sonora producida por un arma de fuego se situaba a 50° con respecto al centro del array y al micrófono 1 (que es considerado 0°). Como se puede ver, el método GCC-Phat ha obtenido la dirección de llegada (DOA) de la fuente sonora en 50°, sin desviarse de su posición real. Sin embargo, los resultados de este método no son tan buenos como el caso dado, ya que para ciertos ángulos se desvía totalmente de su posición real. En el Capítulo 5 se realizará un análisis de los resultados, y se podrá tener más detalles sobre el error de este método.

Teniendo en cuenta los antecedentes narrados en el Capítulo 2, se sabe previamente que el método GCC no obtiene muy buenos resultados frente a reverberaciones o baja

relación SNR. Esto será evaluado igualmente más adelante, comparando la robustez de este método con los métodos que posteriormente se van a describir.

4.2.2. SRP-Phat

El método SRP-Phat se basa en la técnica vista de beamforming, pero no solo eso, sino que también parte del anterior método estudiado, el GCC-Phat.

Ya es bien sabido que un array de micrófonos puede comportarse como beamformer, de modo que las señales capturadas por diferentes micrófonos puedan combinarse de cierta manera para seleccionar la señal procedente de una determinada dirección. Con este método se pretende realizar algo parecido, concretamente usa una técnica de beamforming denominada delay-sum-beamforming, conocida como retardo-suma, que se encarga, como su nombre indica, de retardar la señal recibida en cada micrófono de forma que la fuente de interés quede alineada en todos los canales y posteriormente sumarlos todos, obteniendo que la fuente sonora de interés destaque sobre el resto. Cuando la fuente es desconocida, un beamformer puede ser utilizado para escanear sobre la región del espació predefinida. La respuesta del beamformer es conocida como respuesta encaminada (Steered Response). La localización elegida después de realizar el escaneo, es aquella para la que la potencia de la señal de salida del beamformer es máxima, proviniendo de aquí el nombre completo del método, Steered Response Power (SRP).

El algoritmo SRP-Phat, que es el que se va a implementar, es el más común y está relacionado con el explicado anteriormente GCC-Phat. Se ha demostrado que se puede llegar al cálculo del algoritmo SRP de un beamformer sumando las funciones GCC para todos los posibles micrófonos del array. Esta es la técnica que se usará para la implementación del SRP-Phat, por eso, primeramente se debe realizar la implementación del GCC-Phat, tal y como se explicó en la sección anterior, de modo que una vez se han obtenido las funciones GCC-Phat para cada par de micrófonos, se realiza el siguiente paso, siguiendo con la notación anterior:

$$P_s = \sum_{i=1}^{M} \sum_{j=i+1}^{M} R_{x_i x_j}(n, \tau) \tag{57}$$

donde P_s se hará máximo para el TDOA $\tau = \tau_s$, ocasionado por la fuente sonora.

El algoritmo, por tanto, consiste en evaluar la ecuación (57) para un conjunto de posiciones candidatas, hasta encontrar aquella localización que maximice dicha ecuación. Para llegar a esta evaluación de las posiciones candidatas de la fuente se recurre a una función implementada también para GCC-Phat, que trataba de extrapolar la información obtenida por la función GCC-Phat, de manera que estableciendo la referencia en el centro del array se pueda interpretar correctamente la posición del arma de fuego. Esta función es tau(), explicada en el apartado anterior.

La forma de usar la función tau() para este algoritmo es algo diferente. La matriz 'retardos' devuelta por esta función es la misma que en el caso anterior, pero ahora esta matriz no será usada únicamente para aquellos pares de micrófonos para los que $R_{x_i x_j}$ sea máximo. El proceso a seguir será el siguiente: primero se calcula $R_{x_i x_j}$ para cada pareja de micrófonos (es decir, se calcula el GCC-Phat), se realiza una evaluación de ésta dividiéndola en tramas, después se realiza el sumatorio de los valores seleccionados según la posición de retardo $\tau_{s.ij}$ almacenado en la matriz 'retardos' para todas las combinaciones de micrófonos dadas para una posición candidata concreta, es decir, la matriz tau() almacena los retardos de cada pareja de micrófonos para una posición concreta θ_s y estos valores son los que se eligen en la matriz de GCC para ser sumados. Este valor de suma P, visto en la ecuación (57), se guarda en una variable; acto seguido se realiza la misma operación para la siguiente posición candidata, y así hasta tener todos los sumatorios (57) para todas las posiciones candidatas; el paso final consiste en seleccionar el máximo obtenido P_s, indicando éste la dirección en grados en la que se halla la fuente sonora.

En la Figura 26 se muestra la dirección de llegada de la fuente sonora para cada trama. Se puede ver en el espectro de la señal de entrada en qué instante sucede el disparo y como para las tramas en las que ocurre el disparo, se obtiene como resultado que se localiza a 10°, coincidiendo con la localización real del la fuente sonora producida por el disparo.

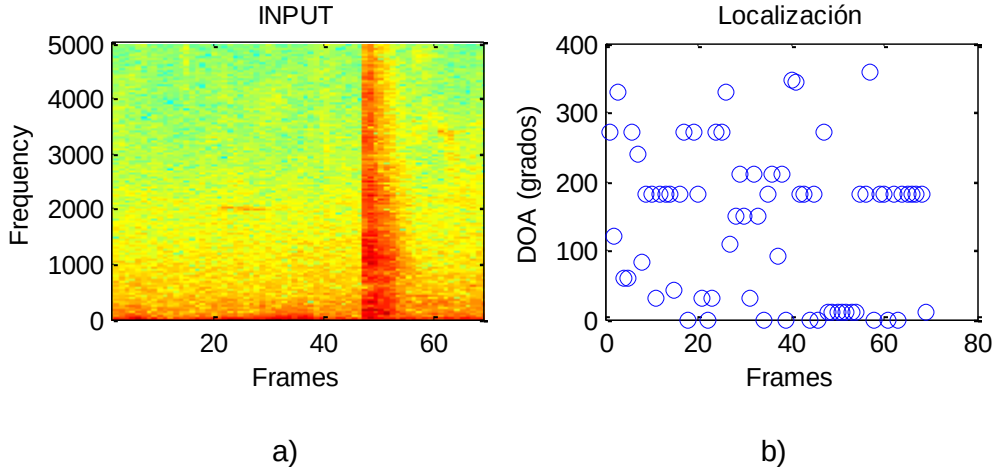


Figura 26. A) Espectro de la señal capturada por unos de los micrófonos y B) Localización de la fuente sonora evaluando en cada trama.

En las tramas de la captura en las que no hay disparo, solo señal considerada ruido, también se indica de donde proviene la fuente sonora que produce este ruido, pero esto no es una información útil, de modo que se puede encontrar la manera de eliminar esta información irrelevante. Esto se puede conseguir siguiendo lo descrito en [23]. El proceso a seguir es sencillo, consiste en evitar esta información irrelevante de la localización de ruido, cuando lo único que interesa es la localización de la señal de disparo, que es la que más potencia aporta. Para ello, calculadas las correlaciones para cada par de micrófonos, se van a simplificar buscando el máximo de cada fila, obteniendo así una sola trama de información de retardos para cada par de micros, de manera que, realizando el mismo proceso anteriormente descrito para el cálculo del SRP-Phat, se va a obtener como resultado un único ángulo, siendo éste el del disparo. Lo que ocurra en ambientes de muy baja relación señal a ruido se analizará en la etapa de evaluación de resultados. En la Figura 27 se puede ver el resultado obtenido para la misma señal de entrada capturada por los seis micrófonos que la vista en la Figura 26, pero en este caso se ha tenido en cuenta el obviar la información que no es útil.

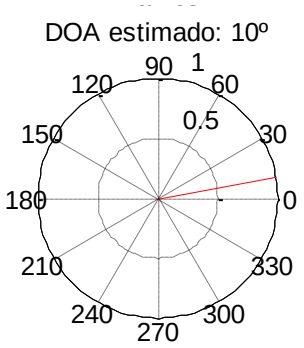


Figura 27. DOA estimado por el método SRP-Phat para una fuente sonora situada a 10° con respecto al centro del array de micrófonos, evitando información irrelevante.

Se puede comprobar a priori, que comparando el algoritmo GCC-Phat y el SRP-Phat para una misma señal capturada por los seis micrófonos, donde la localización real de la fuente sonora es 10°, el método SRP-Phat ha sido más preciso (ver Figura 25 y 27). Al realizar los sumatorios de las correlaciones cruzadas está siendo mucho más robusto, ya que no se basa solamente en la determinación del máximo de la correlación. Más adelante se observará cómo de robusto es en entornos con baja relación señal a ruido.

4.2.3. SRP-Phat modificado

Siguiendo el método explicado en la sección anterior, el SRP-Phat, se ha tenido en consideración ciertas modificaciones sobre este método con el fin de mejorar los resultados obtenidos, haciéndolo más robusto si cabe e intentando que en condiciones de relación señal a ruido muy bajas funcione lo mejor posible. A priori no se puede saber si esta modificación ayudará a conseguir los objetivos citados, esto será evaluado más adelante.

La premisa para la creación de esta modificación parte de la idea de que a la hora de localizar de dónde proviene el disparo que capta los sensores colocados en el coche, el principal y más inmediato interés es conocer la zona de la que proviene, es decir, poder situar la fuente sonora que activa los sensores en un determinado sector. De modo que, conocer la localización exacta, gradualmente hablando, del francotirador provocará la misma reacción en la victima de intentar protegerse que si el algoritmo informa del sector en el que se localiza el francotirador, entendiendo como sector una franja estrecha de ángulos de posible localización. Para concretar más, se podría considerar un sector de 10°, de manera que si el algoritmo de localización detecta que la fuente sonora proviene

de 86°, la interpretación correcta sería que el francotirador está situado entre 81° y 91° ($\pm\frac{10°}{2}$ de la localización indicada por el algoritmo).

Como se ha mencionado, siguiendo con el ejemplo anterior, la reacción del personal de seguridad del vehículo atacado será, en este caso, la de protegerse en la parte izquierda del vehículo, ya que según la colocación del array de micrófonos sobre el vehículo, los 90° harían referencia a la parte derecha (la del conductor). Así, el personal de seguridad se escondería en la parte más alejada para protegerse de la localización de la fuente sonora indicada por al algoritmo, que, ya sea 86°, 87° o 90°, el personal se refugiaría en la misma posición (parte izquierda del vehículo).

Así que, teniendo presente esta idea de sectorización del espacio donde se puede ubicar la fuente de localización, se pasa a introducir este concepto en la implementación del método SRP-Phat. Como es lógico, este concepto se va a considerar en la matriz de retardos de posiciones candidatas, la función tau(), implementada para los métodos explicados anteriormente. A continuación se ve cómo se modifica la ecuación (56) para considerar el tamaño de sector deseado dentro de los retardos en muestras que se darían para diferentes posiciones candidatas:

$$\tau_{s,ij,inicial} = \frac{r}{c}\cos\left[\cos\left(\theta_s - \frac{\theta_{sector}}{2} - \theta_j\right) - \cos\left(\theta_s - \frac{\theta_{sector}}{2} - \theta_i\right)\right] \cdot f_s \qquad (58)$$

$$\tau_{s,ij,final} = \frac{r}{c}\cos\left[\cos\left(\theta_s + \frac{\theta_{sector}}{2} - \theta_j\right) - \cos\left(\theta_s + \frac{\theta_{sector}}{2} - \theta_i\right)\right] \cdot f_s \qquad (59)$$

Los ángulos θ_s, θ_{sector}, θ_i, θ_j son establecidos según el array circular de micrófonos, donde θ_{sector} es el ángulo que se toma como sectorización para este método de SRP-Phat modificado. Señalar que el parámetro f_s hace referencia a la frecuencia de muestreo de la señal de entrada al algoritmo. Como se ve en (58) y (59), ahora se va a tener un retardo inicial y otro final para el sector en cuestión, ya que no se considera una sola posición candidata, sino un conjunto de posiciones en torno a la posición candidata. En la Figura 28 y 29 se muestra una captura de Matlab de la matriz de 'retardos', tanto para SRP_Phat como para el método SRP-Phat modificado respectivamente. En ellas se puede observar que como diferencia se tiene que en la matriz 'retardos' para SRP_Phat modificado se va a tomar un rango de muestras comprendidas entre $\tau_{s,ij,inicial}$ y $\tau_{s,ij,final}$, de manera que en las matrices GCC para cada par de micrófonos se van a

seleccionar más muestras para cada posición candidata, lo que se supone que va a mejorar la robustez del original SRP-Phat.

1	2	0	60.0000	4	-31	-31
1	3	0	120.0000	4	-100	-100
1	4	0	180	4	-140	-140
1	5	0	240.0000	4	-109	-109
1	6	0	300.0000	4	-39	-39
2	3	60.0000	120.0000	4	-70	-70
2	4	60.0000	180	4	-109	-109
2	5	60.0000	240.0000	4	-78	-78
2	6	60.0000	300.0000	4	-8	-8
3	4	120.0000	180	4	-39	-39
3	5	120.0000	240.0000	4	-8	-8
3	6	120.0000	300.0000	4	61	61
4	5	180	240.0000	4	31	31
4	6	180	300.0000	4	100	100
5	6	240.0000	300.0000	4	70	70

Figura 28. Recorte de la matriz de retardos que devuelve la función tau() usando el método GCC-Paht original. Columna1: Micrófono i. Columna 2: Micrófono j. Columna 3: θ_i. Columna 4: θ_j. Columna 5: Ángulo candidato. Columna6: Retardo inicial. Columna 7: Retardo final.

1	2	0	60.0000	4	-41	-19
1	3	0	120.0000	4	-111	-87
1	4	0	180	4	-139	-136
1	5	0	240.0000	4	-98	-116
1	6	0	300.0000	4	-28	-49
2	3	60.0000	120.0000	4	-70	-68
2	4	60.0000	180	4	-98	-116
2	5	60.0000	240.0000	4	-57	-97
2	6	60.0000	300.0000	4	13	-29
3	4	120.0000	180	4	-28	-49
3	5	120.0000	240.0000	4	13	-29
3	6	120.0000	300.0000	4	82	39
4	5	180	240.0000	4	41	19
4	6	180	300.0000	4	111	87
5	6	240.0000	300.0000	4	70	68

Figura 29. Recorte de la matriz de retardos que devuelve la función tau() usando el método SRP-Phat modificado con un sector de 20°. Columna1: Micrófono i. Columna 2: Micrófono j. Columna 3: θ_i. Columna 4: θ_j. Columna 5: Ángulo candidato. Columna6: Retardo inicial. Columna 7: Retardo final.

En las Figuras 28 y 29, se aprecia diferencia entre ellas en las dos últimas columnas de valores, las cuales hacen referencia a $\tau_{s,ij,inicial}$ y $\tau_{s,ij,final}$, como se puede comprobar

en la Figura 28, que se trata de la matriz de retardos del método original SRP-Phat, no existe diferencia alguna en las dos últimas columnas, debido, como es lógico a que no se ha tomado un sector en torno a la posición candidata, sino, únicamente el retardo en muestras que produce esta posición candidata. Sin embargo, en la imagen 29 sí que se aprecia que el retardo inicial y final son diferentes, ya que se está evaluando entre ± 10°. Cabe señalar que, la primera columna indica el micrófono i, la segunda el micrófono j, la tercera el ángulo que forma el micrófono i, la cuarta el ángulo que forma el micro j y la quinta columna nos informa qué posición candidata se está barajando, en este caso el de 4°. Como se puede apreciar en ambas imágenes, se cuenta con quince combinaciones de micrófonos para cada posición candidata y cada una cuenta con un valor en muestras de retardo inicial y final determinado.

En esta modificación del método original, el cálculo del GCC-Phat para cada par de micrófonos será el mismo, cumpliendo con la ecuación (52) y el cálculo del máximo retardo sigue la ecuación (57), propia del SRP-Phat original, sin embargo, ahora la búsqueda de este máximo entre las posiciones candidatas está sujeta a un rango de retardos que van desde $\tau_{s,ij,inicial}$ a $\tau_{s,ij,final}$. De modo que en la matriz $R_{x_i x_j}$ se va a seleccionar aquellos retardos ϵ $[\tau_{s,ij,inicial}\ ,\ \tau_{s,ij,final}]$ para cada θ_s y realizando (57) se encontrará aquella posición en la que se sitúa la fuente sonora producida por el arma de fuego.

En la Figura 30 se aprecia en la parte izquierda el espectro de la señal que capta un micrófono, la energía en decibelios de dicha señal, donde se puede distinguir perfectamente el disparo. En la parte de la derecha se puede ver donde el algoritmo sitúa la fuente sonora producida por el disparo, en la imagen superior no se observa claramente, ya que se está recogiendo toda la información de la señal capturada, mientras que en la inferior, se elimina toda aquella información irrelevante, tal y como cuenta [23]. El resultado obtenido es que la fuente sonora se localiza en 9°, mientras que la medida real era 10°. Esto se debe a que el sector tomado es de 20°, de modo que la medida real está comprendida en ± 10° de la dada.

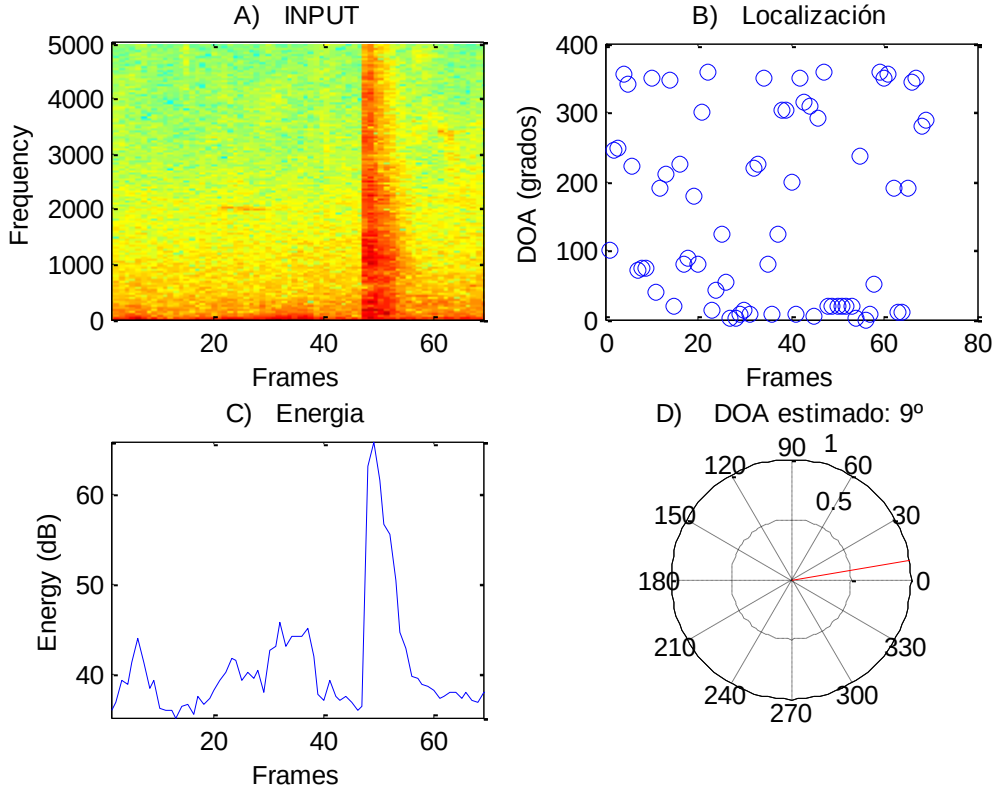


Figura 30. Implementación del algoritmo SRP-Phat modificado con un sector de 20° para una localización de la fuente sonora de 10°. A) Espectrograma de entrada. B) Localización trama a trama. C) Energía de la señal de entrada. D) Ángulo estimado de disparo.

La grabación analizada en el algoritmo presenta una muy buena relación señal a ruido, por lo que ha dado muy buenos resultados tanto en GCC-Phat y aún más exactos con el método SRP-Phat. Sin embargo, en este último se ha desviado más, por lo que se debería estudiar si realmente este método presenta buenos resultados en entornos más ruidosos o no. Se podrá establecer diferentes tamaños de sectorización, donde con menor sectorización los resultados obtenidos se deben de parecer más a los resultados del SRP-Phat original. En la Figura 31 se puede ver como con un sector de 2°, el resultado ha sido exacto, igual que ocurría con el método sin sectorización.

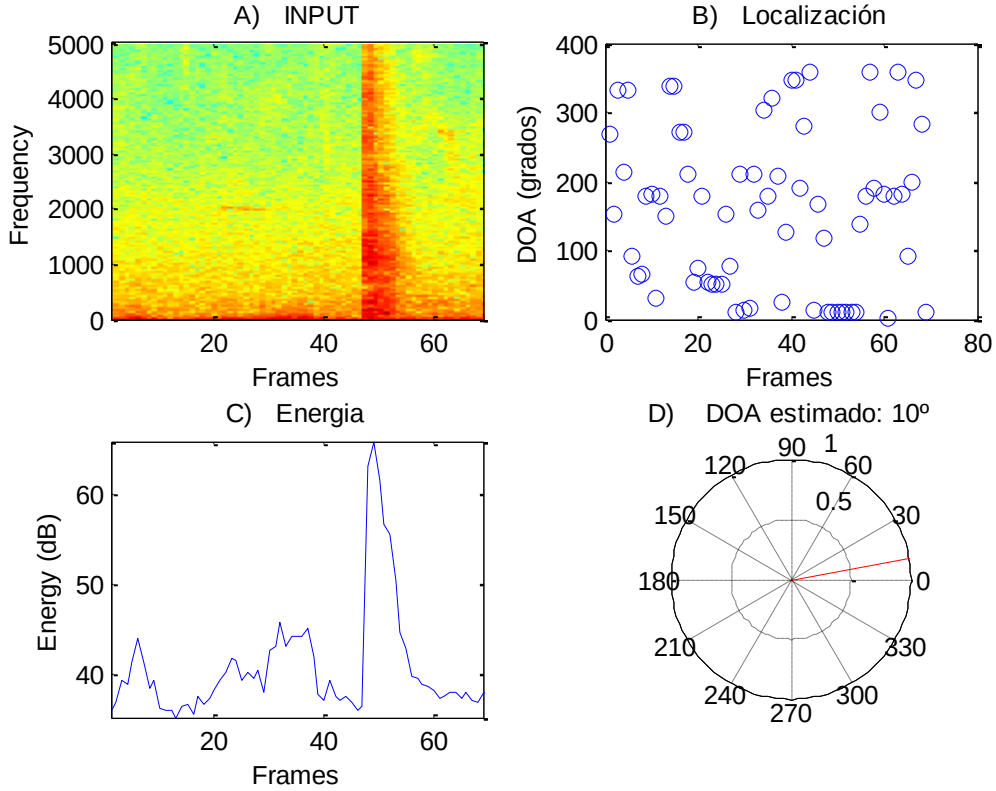


Figura 31. Implementación del algoritmo SRP-Phat modificado con un sector de 2° para una localización de la fuente sonora de 10°. A) Espectrograma de entrada. B) Localización trama a trama. C) Energía de la señal de entrada. D) Ángulo estimado de disparo.

Está claro que para poder examinar con detalle el comportamiento de todos los métodos de localización determinados, incluido el SRP-Phat modificado con diferentes sectores, se debe considerar distintas relaciones señal a ruido, para comprobar cómo funciona cada uno con señales de diferente SNR. Esta evaluación exhaustiva se realizará en el Capítulo 5.

CAPÍTULO 5. RESULTADOS Y DISCUSIÓN

En el Capítulo anterior se ha podido comprender en qué consiste el proyecto a desarrollar, donde se ha explicado con detalle tanto la implementación de la etapa de detección como la de localización. Se ha usado el entorno de Matlab para dicha implementación y para la evaluación de cada uno de los algoritmos. Para estas evaluaciones se ha tenido en cuenta la comparativa entre cada uno de los métodos desarrollados en las distintas etapas y, a la vez, se ha estudiado su comportamiento ante diferentes relaciones señal a ruido, para así comprobar qué tan robustos son frente al ruido.

En primer lugar ha sido necesario crear diferentes bases de datos que serán utilizadas por los algoritmos. En la sección 5.1 se detalla las acciones llevadas a cabo para su formación. Ha sido necesarias tanto bases de datos para la etapa de detección, como una base de datos para la etapa de localización. Partiendo de las diferentes bases de datos como entrada a los algoritmos implementados, se obtendrán una serie de resultados que se van a analizar a lo largo de este Capítulo.

5.1. Bases de datos

Una tarea elemental de este Trabajo Fin de Grado ha sido la creación del material necesario de partida con el que poder trabajar. Ante la dificultad de encontrar bases de datos de disparos que se ajusten a los intereses de este proyecto, es decir, que presenten toda la información necesaria para poder trabajar sobre ellas en la implementación de los algoritmos, se ha tenido la necesidad de crear una biblioteca propia de grabaciones útiles tanto para la etapa de localización como para la de detección. Esta etapa ha supuesto un gran esfuerzo, ya que no se cuenta con bases de datos de disparos de fácil alcance y en el desarrollo de diversos proyectos de este ámbito, cada cual crea sus propios archivos. Igualmente complicado ha sido conseguir una base de datos para localización con seis micrófonos, por lo que se suma una tarea más a este Trabajo Fin de Grado.

A continuación se detalla las tareas llevadas a cabo para cada una de las recolectas de grabaciones que forman las bases de datos.

5.1.1. Detección

Como entrada a los algoritmos NMF para detección de disparo son necesarias señales que contengan, como es lógico, el disparo. Además, precisamente en la etapa del NMF de separación se hará necesaria una base de datos de ruidos ambiente, para mezclar dichas señales con las pistas de disparos. En la explicación de la implementación de la detección de un disparo ya se fue haciendo referencia al uso de estas pistas como entrada a los algoritmos. Para esta etapa ha sido necesaria la creación de varias bases de datos, las cuales se van a ir describiendo y mencionando igualmente la utilidad que tienen en esta etapa.

Cabe destacar que las señales de disparo empleadas en la etapa de entrenamiento de la técnica NMF son diferentes a las señales de disparo usadas en la etapa de separación, en la parte de testeo. De esta manera, se comprueba que el método funciona siguiendo las características propias del disparo, y no únicamente es válido para las señales de disparo entrenadas.

A continuación, en la Tabla 1 se traza un esquema general de las bases de datos que conciernen a esta etapa, para posteriormente ser explicadas una a una con detalle, donde se establece la nomenclatura de las diferentes bases de datos y se indica a que librería hace referencia.

Nomenclatura de la Base de Datos	**Especificación de la Base de Datos**	**Señal de la Base de Datos**
BD.1	Disparos de Entrenamiento	Únicamente disparos
BD.2	Disparos de testeo	Únicamente disparos
BD.3	Disparos obtenidos de [39]	Únicamente disparos
BR.1	Grabaciones de ruido	Ruido de calle
BR.2	Ruido obtenido de [38]	Ruido de tráfico y colegio
BI.1	Pistas de audio impulsivas	Claxon, gritos, cristales…
BI.2	Señales impulsivas obtenidas de [39]	Claxon, gritos, aplausos, cristales…

Tabla 1: Esquema general de Bases de Datos para la etapa de detección.

- BD.1: Disparos de Entrenamiento

En la etapa de entrenamiento del NMF se necesitan una serie de señales de entrada que contengan las características propias de la señal de interés para recoger su información en la matriz no negativa W_d, la cual será usada posteriormente para

identificar el disparo. Dicho esto, se debe disponer de una serie de señales de disparo, de modo que se debe proceder a la creación de una base de datos de disparos limpios de ruido y lo más variada posible. Dada la inmensa cantidad de tipos de armas de fuego, se hace necesaria la recopilación de grabaciones de disparos de la máxima variación de armas posibles.

Para ello se ha recurrido a una página web que cuenta con una amplia base de datos de fuentes sonoras, entre ellas de armas de fuego [37]. Esta web dispone de un amplio abanico de disparos de diferentes armas de fuego, por lo que ha resultado muy útil. El único problema ha surgido de la imposibilidad de acceder a la libre descarga de estas grabaciones, de manera que se ha recurrido a la captura de cada archivo mediante el micrófono interno del PC y se ha modificado dichos archivos con ayuda de Audacity. Esta modificación únicamente se basa en recortar cada pista ajustándose a la señal del disparo, eliminando así partes de la grabación inútiles para el entrenamiento del disparo, de manera que nuestra señal de disparo vendrá claramente caracterizada por su muzzleblast y shockwave.

La base de datos de entrenamiento vendrá dada por disparos de diferentes tipos de armas de fuego, como son rifles de asalto, pistolas, armas semiautomáticas y automáticas, cañones, armas antiguas, disparo con silenciador, entre otras. Dentro de esta clasificación general se pueden ver diferentes modelos de armas, como por ejemplo, dentro de las armas automáticas se puede encontrar disparos de M4A2, Fn P90, MP40, Springfield, etc, con su calibre correspondiente. Además, se cuenta con grabaciones de diferentes disparos producidos por la misma arma, ya que nunca se escucharán dos disparos iguales, debido al viento, temperatura, entre otros factores.

La base de datos para disparos de la etapa de entrenamiento está formada por 150 pistas .wav a una frecuencia de muestreo de 48KHz y cuantificación de 16 bits, para que Matlab pueda trabajar fácilmente con ellas, leyéndolas con su función wavread().

Las 150 señales de disparo serán la entrada del algoritmo NMF de entrenamiento, de modo que a la salida se obtendrá la matriz W_d, con una dimensión $F\ x\ K * 150$. Así que en la etapa de separación el número K' de bases no será el mismo que en la etapa de entrenamiento, sino que viene determinado por el número de señales que hayan servido de entrenamiento a este algoritmo, así que $K' = K * 150$.

- BD.2: Disparos de testeo

La etapa del algoritmo NMF de separación consiste en identificar la señal de interés, no es necesaria la obtención de las señales de manera separada realmente, sino

identificar que en dicha señal se ha producido un disparo en el instante correcto. Para ello se requiere una señal de entrada donde aparezca disparo y ruido. Así que, en esta etapa se van a precisar dos bases de datos que van a funcionar en el algoritmo de manera conjunta, una base de datos de disparos y otra de ruido ambiente. En este apartado se va a describir las grabaciones de disparos para separación.

Igual que para los disparos de entrenamiento, para recopilar estas señales se ha recurrido a [37] y se ha realizado un recorte sobre dichas pistas para seleccionar únicamente el disparo, de tal manera que el shockwave, que presenta temporalmente esa característica forma de 'N' producida al chocar el sonido con el micrófono, esté comprendido en el primer 0.1 segundo de la pista, lo cual es fácil ya que, como se vio en el Capítulo 2, éste tiene una duración de unos 200-300 µs. El saber que el máximo del disparo se sitúa en el primer 0.1 segundo será muy útil, como posteriormente se va a ver.

Estos disparos de separación serán diferentes a los disparos de la base de datos de entrenamiento, pudiéndose tratar de la misma arma, pero de disparos diferentes. Con ello se comprobará que el algoritmo detecta el disparo basándose en sus características impulsivas y no importa que no haya sido entrenado justamente ese disparo, ya que la información útil coincide.

Esta base de datos estará formada por 65 pistas de audio .wav a una frecuencia de muestreo de 48000 Hz y cuantificación de 16 bits. Las señales de disparo guardadas deberán ser mezcladas con pistas de ruido, por lo que igualmente debemos crear una base de datos de estas.

- BD.3: Gunshot_clips

Esta base de datos de disparos será usada en la etapa de separación mezclándolos con ruido para la comparación del método implementado con otros métodos. Gunshot_clips es una librería facilitada por el autor de [39]. La base de datos está compuesta por 332 señales de disparo a una frecuencia de muestreo de 44.1KHz y una cuantificación de 8 bits.

Posteriormente se examinará con detalle el artículo [39], explicando el método de detección usado y en base a él, poder comparar los resultados de dicho método con los resultados obtenidos por parte de los algoritmos implementados en este proyecto usando esta base de datos.

- BR.1: Grabaciones de Ruido

El ruido ambiente que se añada a la señal de disparo debe ser aquel que se dé en espacios abiertos y que se adapten al caso dado, un sistema de seguridad de localización de disparo aplicado a vehículos. Por lo tanto, no estarán contemplados ruidos que se puedan dar en salas, sino ruido de calle.

Se han barajado diferentes ambientes y horarios para realizar una serie de grabaciones, ya que el ruido va a depender tanto del momento del día como de la localización del vehículo. Las grabaciones de las que se compone la base de datos consisten en ruido de tráfico, ruido producido a la salida del colegio y ruido de calle en horas de poco y mucho tránsito.

Esta base de datos está formada por aproximadamente 23 minutos de ruido de calle, capturados con el micrófono externo de un PC y con ayuda del software de captura Audacity. Igual que en los casos anteriores, las pistas de audio han sido guardadas como archivos .wav a una frecuencia de muestreo de 48KHz y una cuantificación de 16 bits.

- BR.2: Segunda base de datos de Ruido

Se cuenta con una segunda base de datos de pistas de ruido, en este caso se han obtenido de [38], donde se hace referencia a pistas de ruido en exterior entre otros escenarios. Estas son señales biaurales, lo cual se debe tener en cuenta en su uso en el código Matlab; sólo se requiere un canal. Las pistas igualmente presentan una frecuencia de muestreo de 48 KHz.

Tanto para esta base de datos como para la anterior de ruido, la manera de mezclar las señales con los disparos es la misma. En primer lugar se selecciona un intervalo de una de las pistas de ruido; de manera aleatoria se elige un punto (posición) de la señal de ruido donde se va a introducir el disparo, procurando siempre que no queden fuera muestras del disparo, sino Matlab lanzará un error; seguidamente la señal de disparo se iguala en muestras añadiendo ceros a la sección de señal de ruido tomada, procurando que el disparo quede en la posición fijada; por último se suman ambas señales. De esta manera queda la señal de entrada a la etapa de separación del método NMF. Se considerarán diferentes relaciones señal a ruido, así que antes de mezclar ambas señales se debe atenuar o amplificar el ruido como se vio en la ecuación (63), para que resulte la SNR deseada.

Igualmente las señales generadas con diferente SNR se guardarán como archivos de audio .wav para su uso en sucesivas implementaciones o evaluaciones, consiguiendo así numerosas librerías de audio.

- BI.1: Pistas de audio de señales impulsivas

La señal producida por el disparo de un arma de fuego no es la única señal impulsiva que se puede encontrar. Cabe analizar cómo se comporta el método descrito frente a otras señales de tipo impulsivo que no sean disparos y que pueden tener características similares como, gritos, cristales al romperse o puertas al cerrarse. De modo que se realiza una base de datos con diferentes ruidos impulsivos, obtenidos de [40] [41]. Estas grabaciones se han podido obtener de manera gratuita mediante descarga.

Dicha base de datos está compuesta por 50 pistas de audio .wav, entre las que se encuentra: 12 pistas de gritos, 6 de pinchazos de globo, 11 pistas de rotura de cristal, 9 de claxon, 10 de portazos, 1 de sirena y 1 de taladradora.

Esta base de datos estaba compuesta por pistas de diferente frecuencia de muestreo, por lo que ha sido necesario adaptarla a las demás bases de datos, donde las demás usan una frecuencia de 48 KHz (excepto Gunshot_clips). De modo que se ha recurrido a la interpolación, que consiste en realizar un muestreo obteniendo así nuevos valores entre muestras. Se ha conseguido con la función de Matlab resample(), que realiza una interpolación lineal.

Estas señales impulsivas serán mezcladas con ruido de la misma forma que si se tratara de disparo y serán evaluadas de la misma manera, con la diferencia que sería preferible recibir malos resultados del detector si se introducen sonidos impulsivos distintos que para el que ha sido diseñado.

- BI.2: Outsider_clips

Esta base de datos también ha sido obtenida de [39], pero en este caso no se trata de señales de disparo, sino de señales de otras fuentes impulsivas como gritos, portazos o aplausos. Será usada para comparar cómo se comporta la técnica NMF frente a otros sistemas de detección de disparos cuando se enfrenta a diversos sonidos impulsivos.

Examinadas a fondo todas las librerías de audio que se han recopilado, se debe detallar también que éstas se han visto utilizadas en la elaboración de bases de datos específicas para: formación de señales de entrada de disparo más ruido necesarias en la etapa de separación y formación de pistas de diferente SNR para su posterior evaluación.

5.1.2. Localización

Para la etapa de localización es preciso disponer de señales de entrada en las que el ángulo de localización real en el cual se sitúa fuente sonora sea conocido, para así

determinar cómo de bien funciona el algoritmo implementado. Obtener una base de datos de este tipo a través de artículos o diferentes librerías es una tarea muy complicada y más aún para un caso tan concreto. Debido a esto, ha sido necesario realizar la base de datos de localización completamente, simulando el entorno para el cual se está desarrollando este proyecto.

Se han realizado una serie de grabaciones de señales sonoras producidas por un arma de fuego, en este caso, concretamente la señal vendrá producida por un arma de un francotirador situado a una distancia del vehículo, donde se encuentran los sensores de captura, un mínimo de diez veces el tamaño del array para ser considerado campo lejano. Se ha creado para esta simulación un array sobre una tabla de madera tal y como se aprecia en la Figura 32, cuya longitud de diámetro es de un metro, por lo tanto, la fuente sonora ha de estar un mínimo de 10 metros de distancia, ya que un metro es la distancia máxima que habrá entre los pares de micrófonos.

Figura 32. Tablero para la simulación de un array circular de 1 metro de diámetro formado por seis micrófonos.

Las grabaciones se han realizado en un entorno controlado, donde en un primer momento no se ha añadido ninguna señal de ruido de fondo. Las pruebas de grabación han tenido lugar en el patio de la Escuela Politécnica Superior de Linares, donde se ha procedido a la colocación de un vehículo cuyo motor se mantenía apagado, de modo que se tiene así una captura de señal en la que no hay apenas ruido de fondo. Esto será el primer contacto con señales con las que poder trabajar en la localización del disparo. Posteriormente se ha repetido para algunas de las localizaciones la grabación del

disparo, pero con el motor del coche encendido y acelerando, produciéndose así un incremento del ruido.

Se han capturado señales de disparos cada 20 grados, considerando únicamente media circunferencia. De modo que se cuenta con señales para las siguientes localizaciones de la fuente sonora: 10°, 30°, 50°, 70°, 90°, 110°, 130°, 150° y 170°. Para la localización de la fuente en estas posiciones ha sido necesaria la estimación de la posición en la que debe ir el altavoz que produce la señal sonora del disparo, para ello se ha trasladado la medida de los grados hacia la fuente sonora con la ayuda de un puntero láser, debiendo colocar el altavoz justo donde indique el láser, el cual irá señalando los diferentes grados tomando como referencia el centro del array. El micrófono número uno del array irá colocado en la posición delantera del vehículo, indicando éste la posición de 0°, mientras el micrófono dos corresponde con la posición de 60°, el tercero con 120° y así sucesivamente. En la Figura 33 se puede ver como la fuente sonora (altavoz que genera la señal de disparo) se localiza a 10° y se encuentra en campo lejano.

En cuanto a la distancia desde la fuente sonora y el vehículo, no es necesario conocerla, ya que nuestro objetivo es el conocer el sector en el que se encuentra el arma de fuego. Lo único que se necesita garantizar es que la fuente sonora se encuentra en campo lejano.

Figura 33. Vista general del entorno de simulación.

Para la realización de estas grabaciones se utilizará una serie de equipos que se van a detallar a continuación. En la Figura 33 se pueden ver diferentes equipos, cada uno realizando su función.

- **Equipos de captura para la creación de la base de datos de localización**

- Micrófonos: El sistema implementado consta de seis micrófonos, por lo que se tendrá 6 señales para cada grabación de diferente localización. Sobre estas señales se calcularán los retardos para llegar al cálculo final de la localización de la fuente sonora.

  Los micrófonos usados son de la marca SHUERE y modelo *Microflex, serie MX 183,* son del tipo electrect, que es una variante del micrófono de condensador que utiliza un electrodo que al estar polarizado no necesita alimentación, las placas estarán cargadas permanentemente desde su fabricación. Se requerirá alimentación phantom para proporcionar energía al preamplificador al que el micro está unido por un cable.

  Se trata de un micrófono de corbata diseñado para captar voces de oradores, grabaciones o monitorización. Presenta un rango dinámico de 96 dB (1 KΩ de carga a 1 KHz) y una respuesta en frecuencia amplia, que va de los 50 a los 17000 Hz. Presenta diferentes cabezales intercambiables que permite elegir el patrón polar, en nuestro caso se va a usar de tipo omnidireccional (de 360°), así que tiene una sensibilidad constante para todos los sonidos independientemente de la dirección desde donde lleguen.

  Estos micrófonos tienen como complemento un paravientos de espuma, que no se ha utilizado y unas pinzas de corbata, las cuales han sido útiles para instalar los micros en su posición sobre el array. Es interesante mencionar que presentan un filtro de RF junto con el preamplificador, el cual optimiza la inmunidad a interferencias RF.

  En la hoja de características del micrófono aparecen algunos datos más, como SPL (117 dB), rechazo en modo común (45 dB) o impedancia de salida (150Ω) entre otros.

- Roland Octa-Capture: Interfaz de audio que consta de 10 entradas y 10 salidas que permite una producción de audio de alto nivel de manera portátil, ya que presenta conexión USB 2.0. En la Figura 34 aparece una imagen de este dispositivo. Presenta ocho entradas para preamplificador de micrófono, con una alimentación phantom independiente, inversión de fase y compresión digital para cada canal. Es un hardware de fácil uso que además permite ajustar los niveles de

entrada simplemente pulsando el botón de AUTO-SENS y ofrece gran estabilidad. Puede ser usado en múltiple plataformas, en este proyecto se usará con Windows 7, de modo que es necesario instalar los drivers correspondientes en el PC, que se facilitan en un CD de instalación.

Figura 34. Tarjeta multicanal Octa-Capture de Roland.

Los micrófonos electrect que incorporan un preamplificador irán conectados a las ocho entradas de preamplificador y será necesario activar la alimentación phantom de 6 de los canales donde irán conectados los micrófonos. Octo-Capture permite trabajar a una frecuencia de muestreo de 44.1/48/96/192 KHz, en este proyecto se va a trabajar a una frecuencia de 48 KHz, siendo necesario ajustar esta opción en el hardware.

Este interfaz de audio incluye un software de procesado de audio: *SONAR X1 LE*, pero debido a diferentes problemas se ha optado por el uso de *ADOBE AUDITION*.

- Ordenador portátil: El interfaz de audio irá conectado al PC mediante USB y las señales capturadas por los micrófonos se almacenarán aquí como archivos .wav. Se ha usado como software de grabación *AUDITON,* siguiendo los siguientes pasos para grabar: se crea un nuevo proyecto pinchando en Archivo > Nuevo > Sesión Multipista; en Edición >Preferencias >Hardware de Audio, se elige como clase de dispositivo ASIO, apareciendo así el hardware conectado Octa-Captue, como muestra la Figura 35; sobre el proyecto que se ha creado aparecerán las pistas correspondientes a los micrófonos conectados a la mesa de audio; posteriormente se debe asignar cada entrada a su correspondiente pista de Audition; se marca el botón R sobre las pistas (Ver Figura 36) para que todas estén activas en la grabación y se procede a dicha acción.

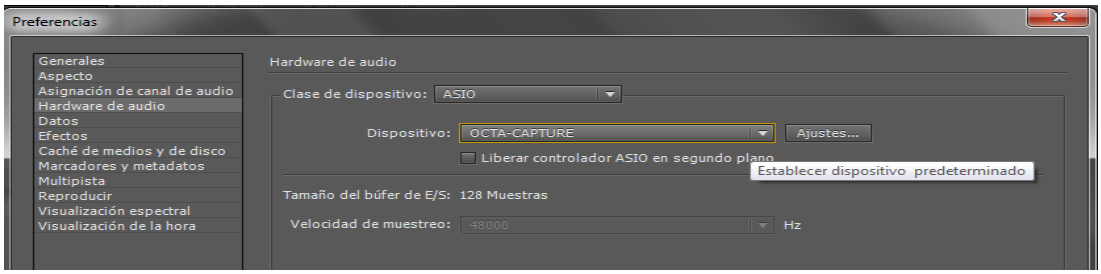


Figura 35. Edición de preferencias en la plataforma AUDITION.

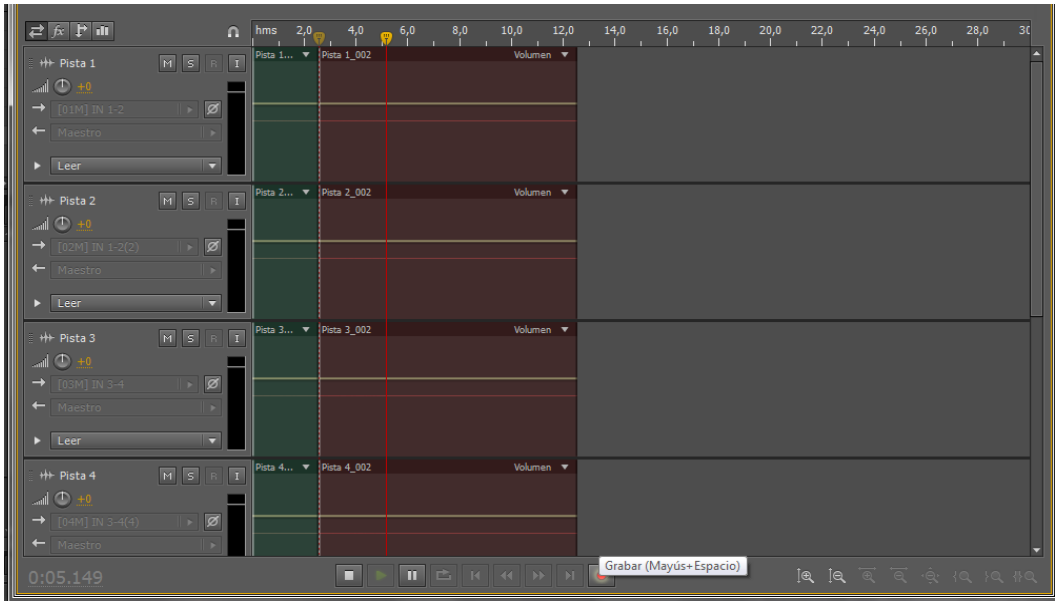


Figura 36. Asignación de canales a pistas de audio en AUDITION.

- **Equipos de reproducción de la fuente sonora producida por un arma de fuego**

Se colocará una serie de equipos a una distancia considerable de la parte de captura para simular el disparo. Estos equipos están compuestos por una mesa de audio *M-AUDIO* que conecta el PC con los elementos de reproducción, siendo estos un altavoz y un subwoofer. En la Figura 37 aparecen los equipos mencionados necesarios para la simulación de disparo.

Figura 37. Equipos encargados de la reproducción de la fuente sonora producida por un arma de fuego.

Estos equipos se van a ir desplazando para realizar simulaciones de disparo en diferentes ángulos.

A continuación se va a explicar brevemente las características de estos dispositivos:

- M-AUDIO Profire 610: interfaz de audio que se conecta al PC convirtiéndolo en un potente estudio de grabación portátil. Cuenta con seis entradas y 10 salidas A/D, pero de estas seis entradas solo cuenta con dos para micrófono con phantom, por lo que no es útil en este proyecto para realizar la captura de señales, pero sí para la reproducción, permitiendo conectar el subwoofer a este interfaz. Presenta 8 salidas balanceadas, de las cueles se requiere solamente una, ya que se necesitará un único altavoz. En la Figura 38 se puede observar este interfaz de auido.

- Subwoofer GENELEC 7050B PM: es un perfecto complemento para el altavoz utilizado, con una amplia respuesta en graves hasta los 25 Hz. Presenta cinco

canales de entrada principales y un canal con selección de reproducción con un ancho de banda de 85 a 120 Hz. Este equipo presenta un nivel máximo de presión sonora (SLP) de 100dB. A él irá conectado el altavoz que se detalla a continuación por medio de un conector XLR, al igual que al interfaz de audio.

- 8020B Bi Amplifier: Su respuesta en campo libre va de los 60 Hz a ls 20 KHz, por lo que es perfecto para el cometido y mejor aún si se complementa con el subwoofer especificado. Presenta valores muy bajos de distorsión y buena directividad. En la Figura 39 aparecen estos dos últimos aparatos utilizados para formar la base de datos.

Figura 38. Interfaz de audio M-Audio para la reproducción del disparo.

Figura 39. Equipos utilizados en la reproducción del disparo. Altavoz a la derecha y subwoofer a la izquierda.

5.2. Setup

5.2.1. Consideraciones en la etapa de detección

Las señales utilizadas presentan una frecuencia de muestreo de 48 KHz, y están formadas por ruido y disparo. Es necesario, para las sucesivas evaluaciones, partir de una relación señal a ruido determinada. Debido a la imposibilidad de realizar múltiples grabaciones con diferentes SNR cada una, ya que es una tarea excesivamente pesada, se han creado pistas de diferente relación señal a ruido mediante Matlab usando la base de datos BD.2 para la señal de disparo mezclada con BR.1 o BR.2, tratándose éstas de señales de ruido. Se han formado 50 pistas de audio de 4 segundos con diferente relación señal a ruido, eligiendo las tramas correspondientes de ruido de manera aleatoria. Las SNR consideradas han sido -15, -10, -5, 0 y 5 dB.

Para establecer la SNR deseada, se pone la señal de disparo de la BD.2 y el ruido a la misma potencia (SNR=0) y posteriormente se le aumenta o disminuye la potencia al ruido, ya que es preferible no modificar la potencia del disparo para que nunca se llegue a producir saturación si aumentamos su potencia. Para conseguir todo esto, se establecen las siguientes ecuaciones, donde se define como se calcula la potencia de la señal de disparo y ruido, seguidamente se calcula la SNR entre ambas señales y por último el ruido original se ve atenuado o amplificado según la relación señal a ruido que se desee imponer.

$$P_{disparo} = \frac{Disparo^{\cdot(2)}}{length(Disparo)} \tag{60}$$

$$P_{ruido} = \frac{Ruido^{\cdot(2)}}{length(Ruido)} \tag{61}$$

$$SNR_{real} = 10\log_{10}\frac{P_{disparo}}{P_{ruido}} \tag{62}$$

$$Ruido_{SNR} = Ruido \cdot 10^{\left(\frac{SNR_{real}}{20} - \frac{SNR}{20}\right)} \tag{63}$$

En la ecuación (63), el parámetro SNR hace referencia a la relación señal a ruido que se va a imponer, de modo que el ruido se verá amplificado o atenuado, mientras que la potencia del disparo no va a variar, de manera que si se vuelve a calcular la relación

señal a ruido entre el disparo y el ruido modificado se obtendría la SNR deseada ($10\log_{10}\frac{P_{disparo}}{P_{ruido_{SNR}}} = SNR$). Dicho esto, solo falta sumar ambas señales para así obtener una pista nueva conocida su SNR. Estas nuevas librerías de archivos de audio van a ser empleadas como señales de entrada para el testeo del algoritmo.

Otro asunto a considerar son los diferentes parámetros a tener en cuenta, además de la SNR ya mencionada. El número de iteraciones usado para la factorización en las matrices W y H es de 70 en la etapa de separación, más que suficiente como se pudo ver. Igualmente se define el tamaño de ventana, que será de 2048 muestras, para así descomponer la señal en intervalos de 0.427 segundos, mayor que la duración del disparo, por lo que es posible reducirlo a 1024 (esto se considerará en la futura fase de implementación en tiempo real). Otro parámetro de interés es el número de bases K a considerar, se va a tomar un valor de K=5 en primer lugar, ya que las referencias han llevado a él. Finalmente este parámetro también será evaluado para otros valores, comprobando los resultados obtenidos. Los valores que se analizarán serán K= 2, 3, 5, 7.

Para proceder a realizar esta evaluación es necesario determinar un umbral que indique que se está ante una señal de disparo. Para ello se ha procedido a dicho cálculo de umbral, analizando la matriz de activaciones H_d. Dada una serie de señales de entrada que presentan disparo y ruido de fondo, se comprueba que la matriz no negativa de activaciones detecta perfectamente el disparo identificando las frames de activación correctas en las que se sitúa la señal de disparo. De modo que si esto se cumple, se calcula una media de los máximos de la suma de la matriz H_d (ocasionados sin lugar a dudas por la fuente de disparo). El umbral medio obtenido es 115.6905. Dado que dichos máximos presentan valores muy cambiantes, se decide estimar otro umbral más fiable.

Este nuevo umbral consiste en seleccionar las muestras de la matriz de suma de H_d que son activaciones del disparo (este intervalo de muestras conforma los valores máximos de la matriz), estas muestras se suman entre sí y se dividen entre la suma del resto de valores de la matriz, que son el ruido, expresando el resultado en decibelios; tomando así una especie de relación señal a ruido para todas las señales destinadas al cálculo de umbral. De modo que, se hace una media de todos los valores en decibelios calculados para cada señal destinada a umbralización (habrá un total de 30 señales para esta tarea) e igualmente se realiza la desviación típica, para no ser demasiado restrictivos. Los valores obtenidos son una relación señal a ruido de 10.1705 dB y una desviación típica de 2.5054.

Se procederá a evaluar 50 señales para recopilar los resultados que se obtienen y para ello se ha elegido la segunda opción de umbralización. De modo que si la relación disparo a ruido de una señal de entrada es mayor que el umbral calculado restándole la desviación típica, esta señal es identificada como disparo, en caso contrario se trataría de un falso negativo. En el caso de que la señal superara este umbral pero el disparo no se encuentre en dicho punto, se trataría de un falso positivo.

Por último queda establecer que ocurre con un concepta que se ha explicado en el Capítulo 4, el *SPARSE*. Además de comprobar cómo funciona la técnica NMF en entornos ruidosos, se compararán los resultados obtenidos para el uso de *SPARSE,* tomando λ=0, 0.3, 0.5, 0.7 y 1, ya que λ está comprendido entre 0 y 1. Estos valores van a ser usados en las diferentes etapas de detección, en la etapa de entrenamiento y en la etapa de separación.

Estos análisis serán realizados usando los archivos almacenados en las bases de datos presentes en la sección anterior. Para la etapa de entrenamiento se va a usar BD.1, obteniendo de resultado la matriz W_d considerando los diferentes valores del *SPARSE.* En la etapa de separación se usan 50 señales elegidas de manera aleatoria de la base de datos BD.2 mezcladas con tramos de ruido BR.1 o BR.2. Por lo que se va a realizar evaluaciones para ambas bibliotecas de ruido.

5.2.2. Consideraciones en la etapa de localización

En esta etapa se va a utilizar un tamaño de ventana de hamming de 2048 muestras, un noverlap al 50% y un tamaño de fft de 2048 muestras. Las señales de entrada están grabadas a una frecuencia de muestreo de 48 KHz, de modo que estos datos coinciden con los citados en la etapa anterior.

En este caso se ha formado pistas de audio para una SNR de -15, -10, -5, 0, 5, 10, y 15 dB, analizando así entornos de mejor relación señal a ruido. La base de datos de localización está formada por las grabaciones realizadas en el patio del Edificio B de la Escuela Politécnica de Linares, y estas son las que se van a procesar para variar su SNR. Para ello, en primer lugar se debe estimar cual es la relación señal a ruido existente en las grabaciones, así que se busca el máximo absoluto de la grabación, que es el disparo sin duda, y se selecciona un rango de muestras a su alrededor de tamaño algo mayor a la duración de un disparo (se ha tomado 0.4 segundos), para que lo seleccione completamente, de modo que a esta trama se le calcula su potencia, siendo esta la potencia del disparo (60). Igualmente se calcula la potencia del ruido (61), donde este viene dado por una de las pistas que forma una segunda base de datos de grabaciones

de ruido de tráfico, obtenida de [38]. De modo que así se puede establecer la relación SNR de la grabación realizada, como indica la ecuación (62). Una vez conocida ésta, se pasa a establecer la relación señal a ruido que se considere oportuna, modificando la potencia del ruido como se indica en (63). El ruido creado será sumado a cada uno de los micrófonos, sin perturbar la localización del disparo captado. En la posterior sección 5.4 se hará una evaluación sobre cómo se comportan los algoritmos de localización dadas pistas con diferentes SNR, dónde cabe esperar que para relaciones señal a ruido muy pequeñas, en las que la potencia del ruido supere a la del disparo, los algoritmos se comportarán peor que para aquellas pistas donde el ruido mezclado sea muy leve.

Para finalizar, señalar que se va a realizar una evaluación para aquellos ángulos de la fuente sonora para los que se formó la base de datos: 10°, 30°, 50°, 70°, 90°, 110°, 130°, 150° y 170°. En el caso concreto del uso dl método SRP-Phat modificado, se va a utilizar una sectorización de 20°, 10°, 6°, 4° y 2°.

5.3. Métricas

Lo siguiente será la realización de un análisis exhaustivo de los métodos implementados de localización y detección de un disparo. A continuación se concreta el análisis específico para cada una de estas etapas.

5.3.1. Métricas de detección

Para comenzar, la primera evaluación tenida en cuenta de la detección de una fuente sonora producida por un arma de fuego utilizando la técnica NMF, es realizar un estudio de 50 señales que contienen disparo, comprobar para cuántas de estas señales se detecta disparo (True Positive, TP), para cuántas no se detecta disparo (False Negative, FN), y en cuantas de las señales se detecta disparo sin tratarse de esto (False Positive, FP). Este estudio sigue como referencia el artículo [4].

Una vez se obtienen TP, FP y FN, se realizan una serie de métricas que tienen bastante interés:

- Exhaustividad (Recall). La métrica Exhaustividad se define como el cociente entre el número de frecuencias correctas (TP) y la suma del número de frecuencias correctas más el número de falsos negativos (FN). Es un indicador del porcentaje de eventos que dejan de ser detectados por el estimador.

$$Reall = \frac{TP}{TP+FN} \tag{64}$$

- Exactitud (Precision). La métrica Exactitud se define como el cociente entre el número de frecuencias correctas (TP) y la suma del número de frecuencias correctas más el número de falsos positivos (FP). Es un indicador del porcentaje de eventos espúreos generados por el estimador.

$$Precision = \frac{TP}{TP+FP} \tag{65}$$

- Precisión (Accuracy) *Acc*. Esta medida *Acc* es una métrica orientada a tramas. Establece la relación entre el número de eventos correctamente detectados (TP) y la suma de TP, el número de eventos detectados erróneamente (FP) y el número de eventos no detectados.

$$Acc = \frac{TP}{FP+FN+TP} \tag{66}$$

De igual manera, se va a estudiar el comportamiento de la técnica NMF implementada utilizando señales impulsivas, tales como portazos, gritos, cristales rotos, etc, en lugar de señales producidas por un arma de fuego. Para ello se cuenta con la base de datos BI.1, que será mezclada con la segunda base de datos de ruido BR.2. En este caso, igual que tratándose de disparo, se obtendrán resultados de TP, FP y FN, donde los True Positive y los False Positive identifican al ruido impulsivo como una señal de disparo.

Por último, otra de las métricas a tener en cuenta en evaluación es aquella que sigue [39], donde se va a comparar la técnica NMF con las técnicas evaluadas en este paper. En [39] se propone un sistema de detección de disparo con alta probabilidad de detectar disparo, robusto al ruido y de baja complejidad computacional. Los algoritmos que se utilizan son la máxima correlación cruzada, MFCC (Mel Frequency Cepstral Coefficients), que suelen emplearse en reconocimiento del habla, SVMs (Soport Vector Machine), que se utiliza para entrenar un modelo para posteriormente poder predecirlo y coeficientes LPC (Linear Predictive Coding). La evaluación llevada a cabo usará como base de datos

BD.3 y BI.2, consistiendo en lo siguiente: se va a estudiar la probabilidad de detectar un disparo como disparo (True positive Rate, TPR) y la probabilidad de detectar otro sonido impulsivo extraño como disparo (False Positive Rate, FPR). Para obtener dichos resultados para la técnica NMF implementada en este Trabajo Fin de Grado se va a considerar una relación SNR= 20 dB, ya que según [39] es la relación que reduce las falsas alarmas de disparo, es decir, de detectar disparo sin haberlo y además es una relación común usada en sistemas de detección de disparo. Como señal de ruido se usará ruido blanco gaussiano de media 0. Esta comparación se va a realizar con el *SPARSE* en ambas etapas que mejor se comporte.

Sobre esta comparación, también se tiene muy en cuenta cómo se comporta el sistema frente a sonidos impulsivos diferentes a un disparo, por ello [39] presenta una mejora en este campo, usando RBF kerner (Radial Basis Function Kernel) complementando a SVM en lugar d Kerner lineal. En definitiva, lo que este paper pretende explicar es cómo se ha conseguido minimizar los falsos positivos originados por ruidos impulsivos usando una función base radial de Kerner. A priori, el sistema NMF diseñado debe comportarse muy bien ante sonidos producidos por armas de fuego, pero no se ha diseñado ninguna mejora que impida que otros sonidos impulsivos puedan ser detectados como disparo. En la siguiente sección se va a comprobar esta conjetura.

5.3.2. Métricas de Localización

El objetivo de la evaluación de la etapa de localización es el de comparar los métodos implementados explicados en el Capítulo 4. Se realizará un estudio del GCC-Phat, SRP-Phat y SRP-Phat modificado para analizar la robustez que presenta cada uno frente al ruido y su exactitud a la hora de localizar la fuente sonora.

En primer lugar, se requiere el uso de la base de datos formada por grabaciones en el patio de la Escuela Politécnica de Linares para la fuente sonora situada en diferentes ángulos con respecto al array de micrófonos. Esta base de datos será ampliada realizando archivos de audio con diferente relación SNR, para así evaluar la robustez de cada método frente a baja SNR.

Las métricas estadísticas usadas para realizar estas comparativas serán:

- Error absoluto: Se trata de la diferencia entre el valor real de la magnitud a medir y el valor obtenido en una medida.

$$E_{abs} = |X - X_{real}| \tag{67}$$

La magnitud a medir es la localización de la fuente, de modo que el valor real será el ángulo de posición de la fuente en el momento de la captura. En el momento de evaluar el comportamiento de cada método para una determinada relación señal a ruido, se hará una media del error absoluto obtenido para cada ángulo de posición distinto de la fuente para dicha SNR.

- Raíz de error cuadrático medio: Se trata de una medida que representa la desviación de la diferencia entre el valor predicho por un estimador y el valor real observado. Se ha decidido tomar esta métrica siguiendo el ejemplo de [25].

$$RMSE = \sqrt{\frac{\sum(X - X_{real})^2}{N}} \tag{68}$$

donde N es el número total de medidas a evaluar.

- Desviación estándar: Es una medida de dispersión que indica la desviación que presentan los datos en su distribución respecto de la media aritmética de dicha distribución, con objeto de tener una visión más acorde con la realidad. Matlab presenta una función para el cálculo de la desviación típica, esta es std().

Estas métricas darán una visión global del comportamiento de los métodos de localización.

En el apartado 5.4.2. se contemplarán los resultados obtenidos, analizando las ventajas e inconvenientes de cada uno de los métodos, además de seleccionar el método idóneo para la futura implementación en tiempo real.

5.4. Análisis de resultados

Una vez establecidos los diferentes campos a considerar para evaluar ambas etapas, visto en la sección anterior, se procede a reflejar los resultados obtenidos y comentarlos con el mayor rigor posible. Se va a plasmar una serie de tablas y gráficos a partir de las cuales se analizará que tan robustos son los sistemas implementados y cuáles son las ventajas de unos con respecto a otros.

Igual que ocurre en los capítulos anteriores, esta sección se va a dividir en las etapas en las que el proyecto se puede fraccionar, para así determinar cuál es la configuración de parámetros que mejor funciona en según qué situaciones en cada caso.

En la sección 5.3.1 y 5.3.2 se plasman los resultados obtenidos siguiendo las métricas ya explicadas a considerar.

5.4.1. Resultados de etapa de detección

En las Figuras 40, 41, 42, 43 y 44 se representa un diagrama de barras de los resultados para TP, FP y FN para las 50 señales con ruido procedente de la base de datos de grabaciones procesadas con distinta SNR. El objetivo es comparar los resultados obtenidos para una λ diferente en el uso de *SPARSE* y sacar conclusiones sobre con cuál se comporta mejor el sistema de detección

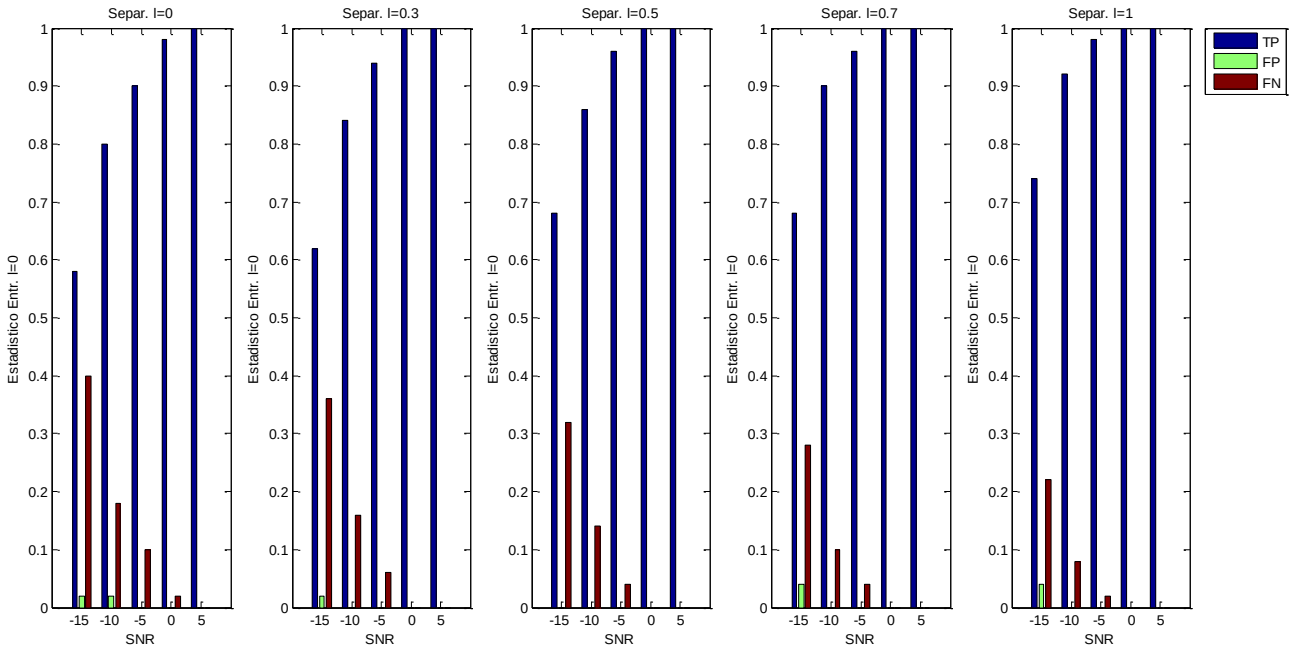


Figura 40. Diagrama de barras para $\lambda=0$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos de ruido BR.1.

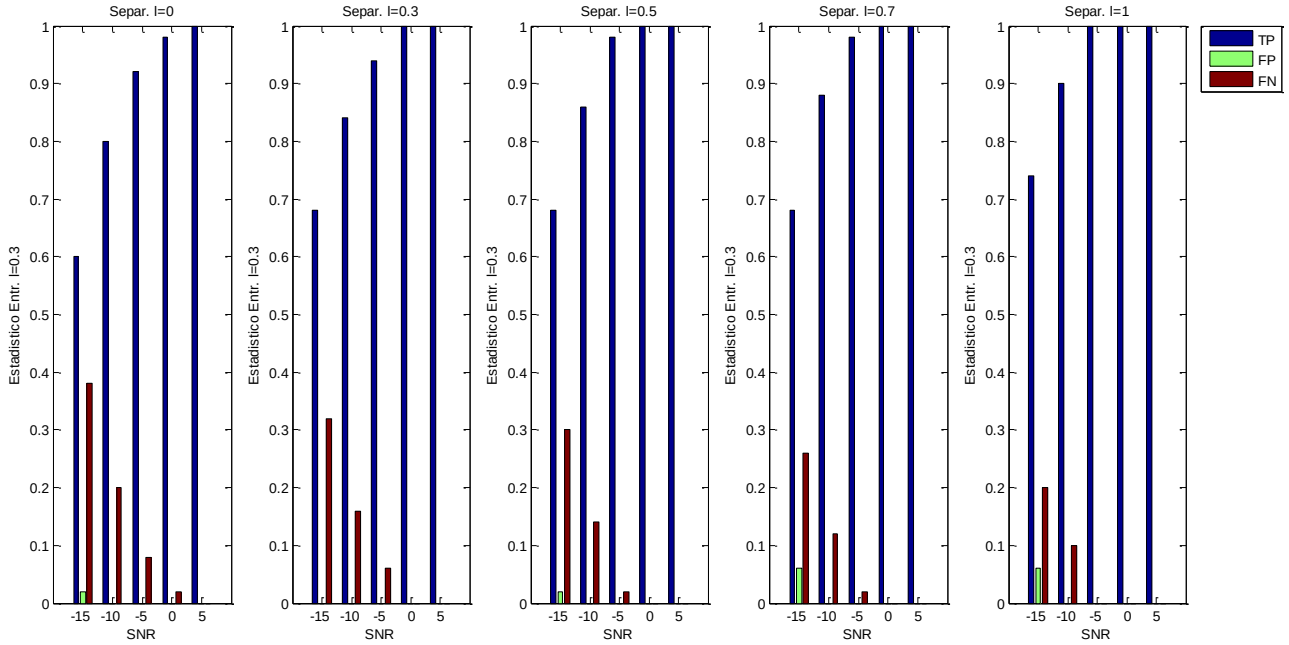


Figura 41. Diagrama de barras para $\lambda=0.3$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos de BR.1.

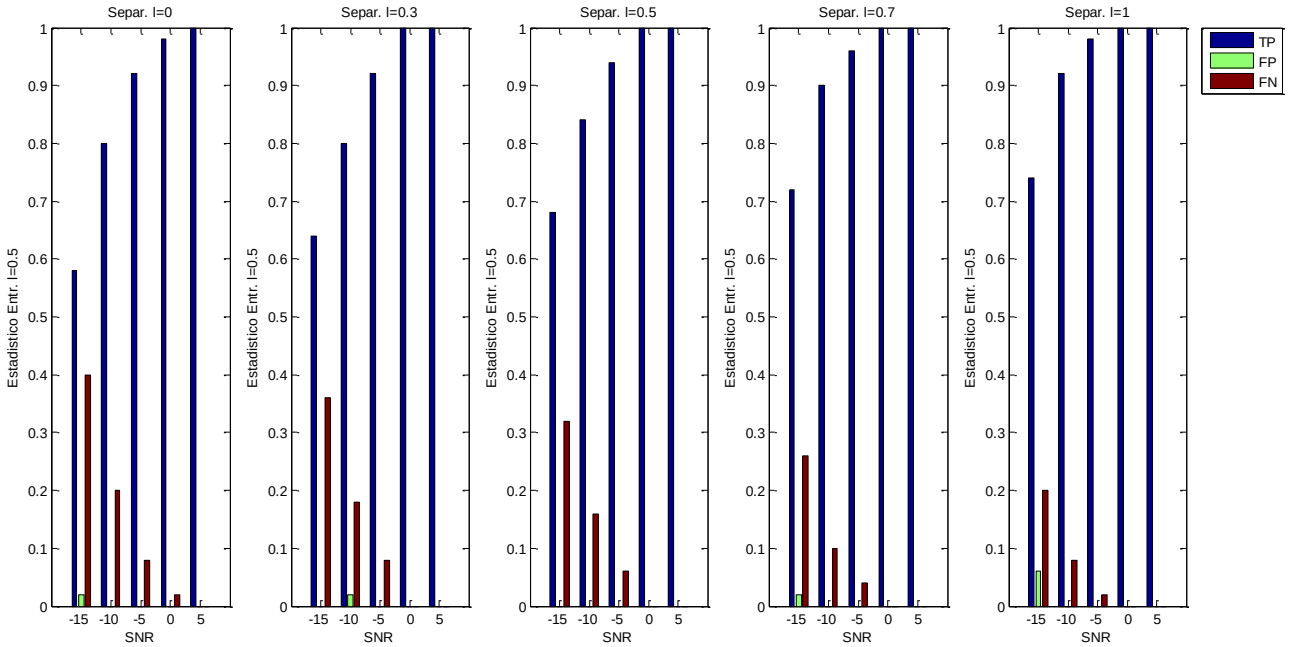


Figura 42. Diagrama de barras para $\lambda=0.5$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos de ruido BR.1.

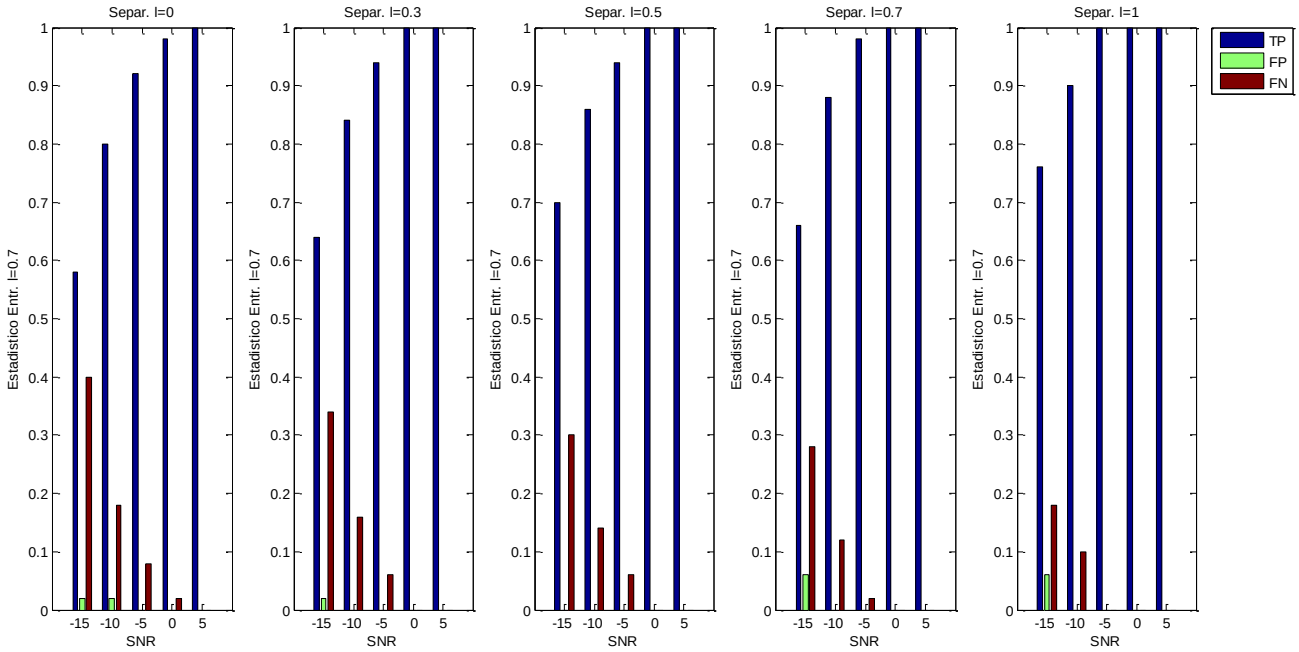


Figura 43. Diagrama de barras para λ=0.7 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.1.

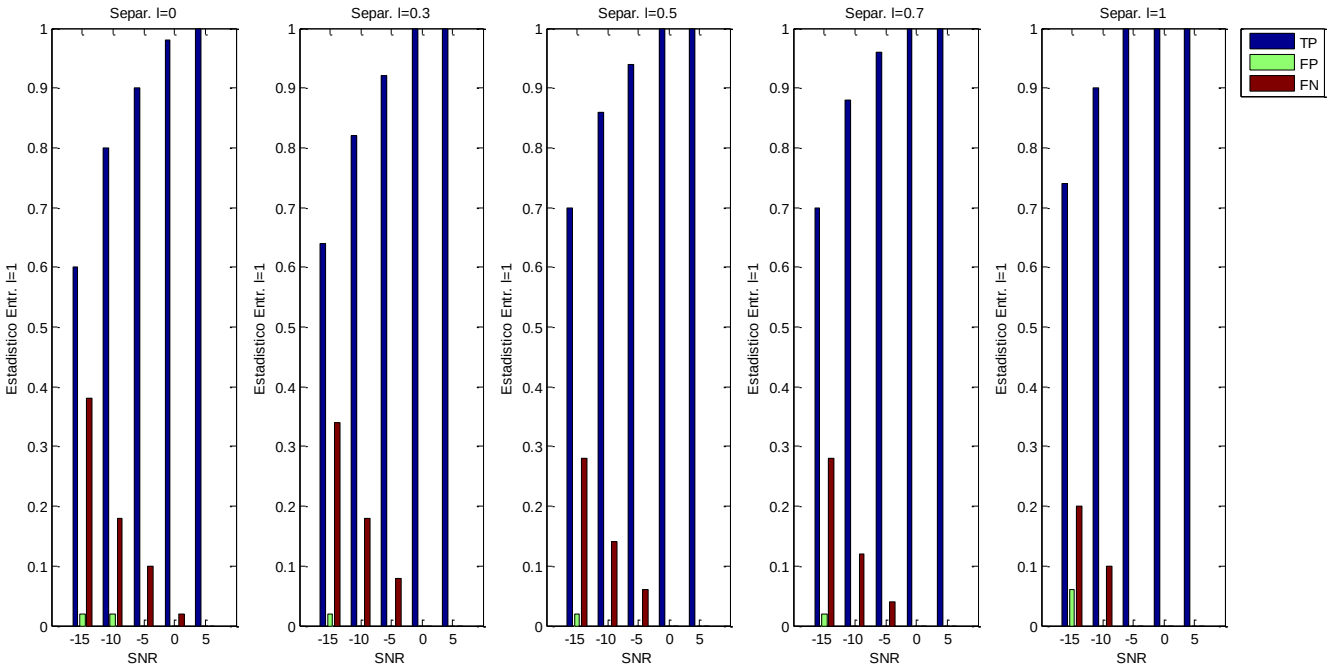


Figura 44. Diagrama de barras para λ=1 en la etapa de entrenamiento y λ= [0, 0.3, 0.5, 0.7, 1] en la etapa de separación. Base de datos de ruido BR.1.

A simple vista queda claro que cuanto mejor es la SNR mejor serán los resultados obtenidos por este sistema. En el caso de SNR=5 dB no se produce mejora al aumentar λ, ya que el 100% de las señales han sido identificadas como disparo en todos los casos y por tanto no hay mejora posible. Cuando observamos relaciones señal a ruido peores, si que se identifica variación en la proporción de señales que han sido identificadas como señales sin disparo. En vista general, a mayor λ en la etapa de separación menor proporción de FN. En cuanto a la etapa de entrenamiento, el uso del *SPARSE* no modifica llamativamente los resultados, pero si se aprecia una ligera mejora cuando aumenta. Un hecho a destacar es que aparecen muy pocos falsos positivos, algo que a priori es bueno, ya que indica que el método de detección y el umbral usado identifican correctamente el disparo en la gran mayoría de los casos. Es en -10 y -15 dB dónde se comienza a ver la aparición de falsos positivos, el aumento de ruido enmascara al disparo y detecta sonidos extraños como disparo. Hay que analizar también el hecho de que al aumentar λ, se ve disminuido el porcentaje de falsos negativos pero aumentan los falsos positivos, pero no de manera comparable, ya que viendo casos extremos con $\lambda=0$ y $\lambda=1$ en la etapa de separación, los falsos positivos disminuyen del 40% al 20% en SNR=-15 dB aproximadamente y los falsos positivos aumentan del 2% al 5% aproximadamente.

Se va a comparar los resultados obtenidos al mezclar disparo y ruido de la base de datos de grabaciones, con la suma de disparo y ruido de la base de datos BR.2. Éstos vendrán dados los las Figuras 45, 46, 47, 48 y 49.

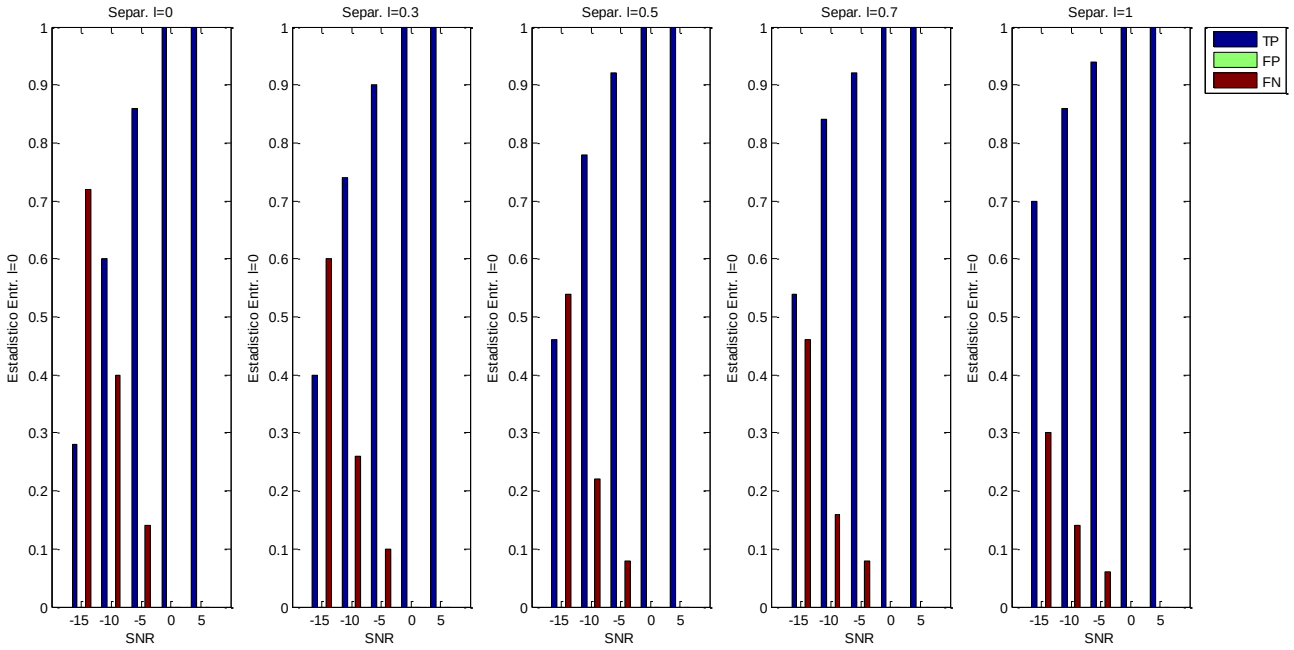


Figura 45. Diagrama de barras para $\lambda=0$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos de ruido BR.2.

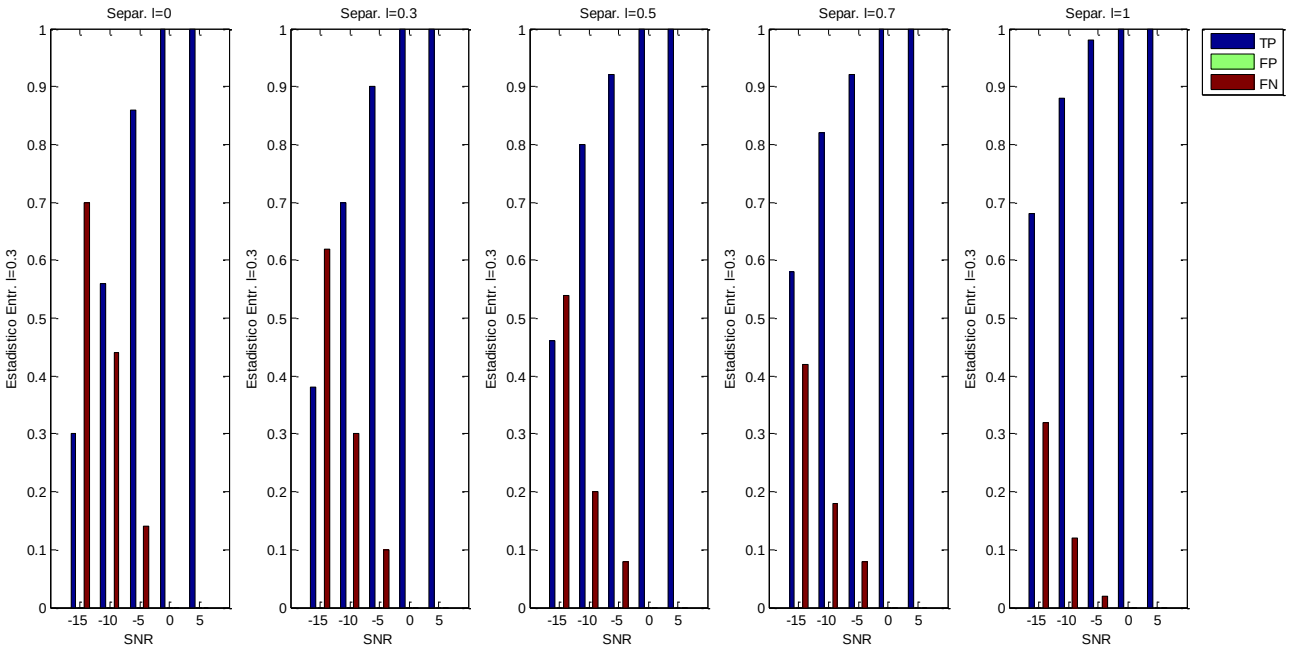


Figura 46. Diagrama de barras para $\lambda=0.3$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos de ruido BR.2.

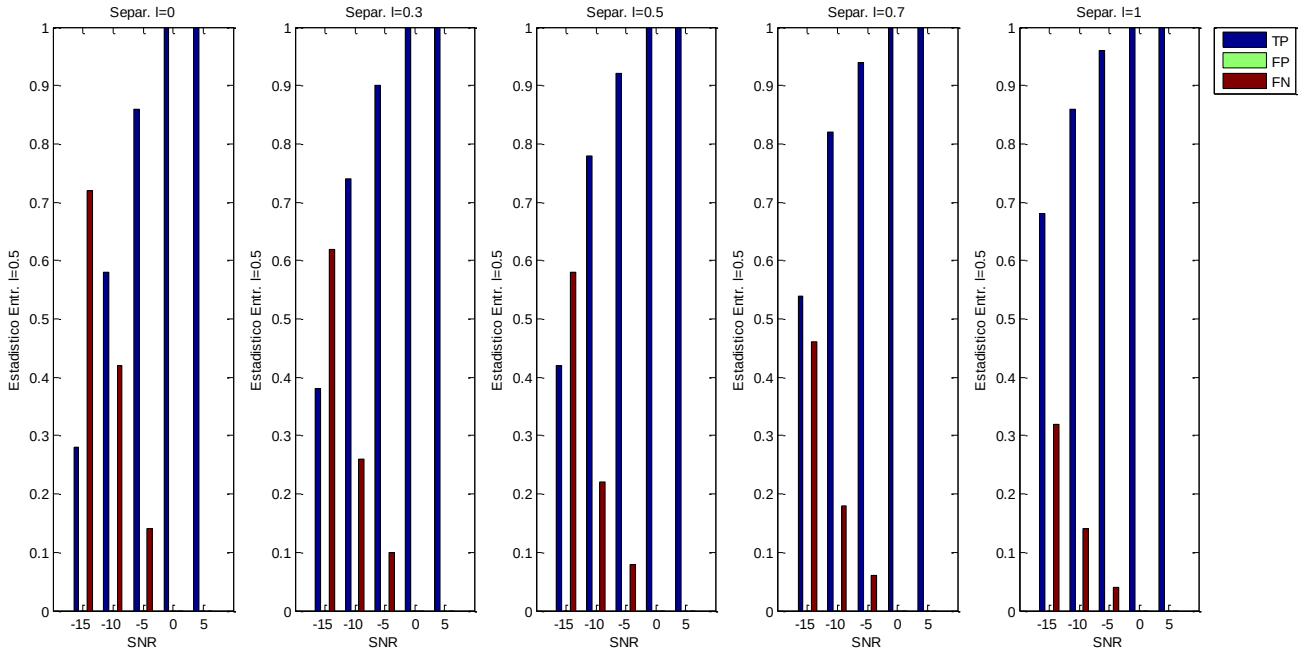


Figura 47. Diagrama de barras para $\lambda=0.5$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos de ruido BR.2.

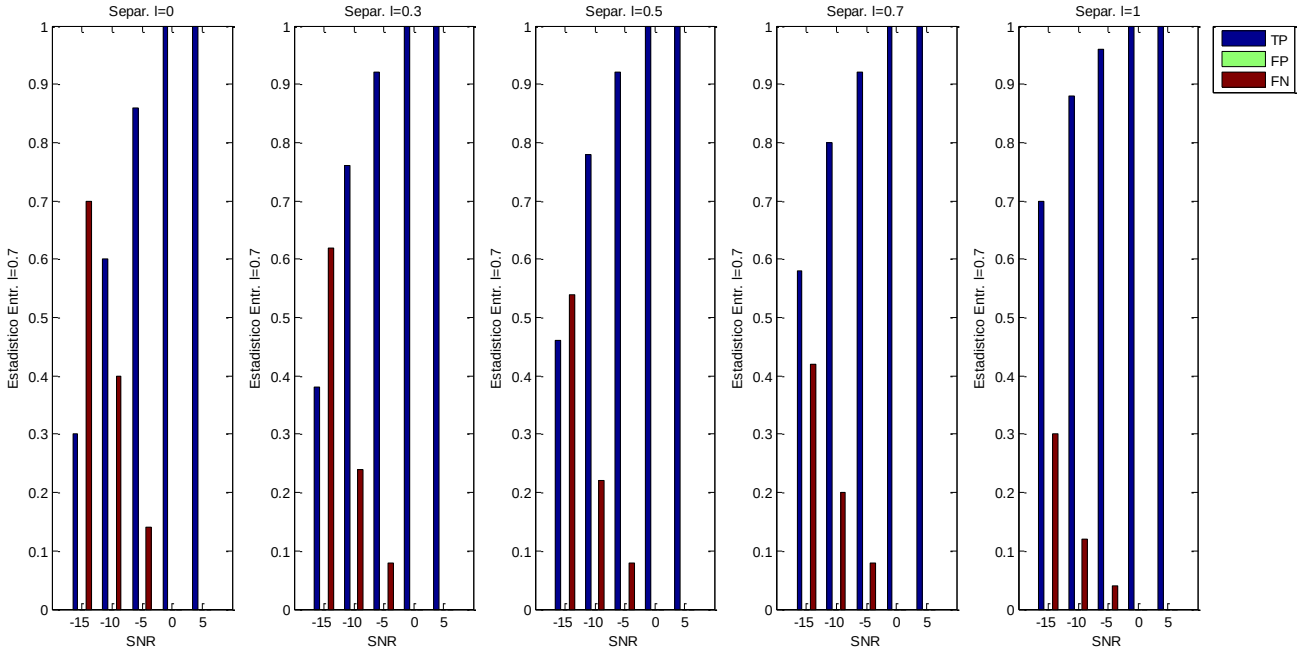


Figura 48. Diagrama de barras para λ=0.7 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.

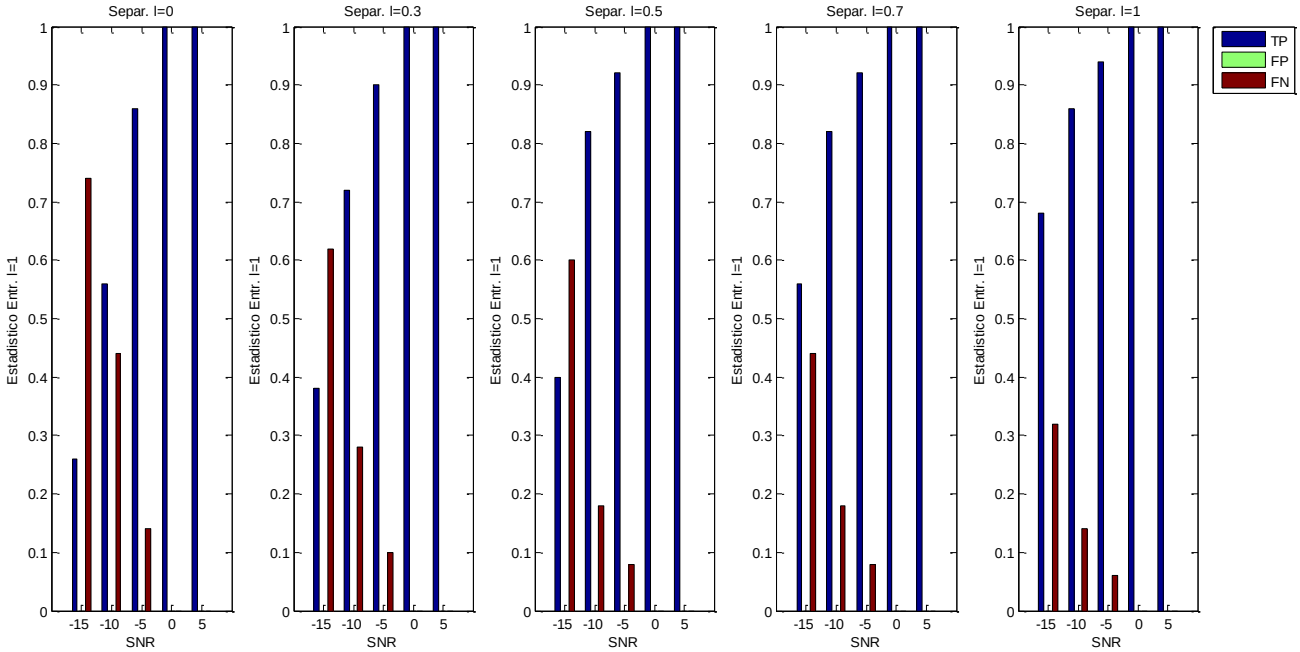


Figura 49. Diagrama de barras para λ=1 en la etapa de entrenamiento y λ=[0, 0.3, 0.5, 0.7,1] en la etapa de separación. Base de datos de ruido BR.2.

Con respecto a las gráficas mostradas, tanto de la base de datos de ruido formada por grabaciones propias, como la base de datos de ruido de [38], se llega a unas conclusiones en común: el algoritmo funciona bastante bien para SNR mayores de 0 dB, hay que tener presente que en situaciones reales, el disparo estará por encima de los 20 dB sobre el ruido, aunque siempre puede haber excepciones, por lo que se estudian

peores casos; con respecto al *SPARSE*, el algoritmo tiene mejores resultados, disminuyendo los falsos negativos con λ=1 en la etapa de separación y con lambda=0.7 en la etapa de entrenamiento. Se ha elegido esta configuración observando los diagramas de barras. Otro detalle observado es que en esta segunda evaluación no se han obtenido falsos negativos, debiéndose a que el ruido utilizado aquí no presenta en ningún momento la aparición de otros ruidos impulsivos, que en el caso de las grabaciones propias realizadas en la calle sí que puede aparecer, como portazos de vehículos o claxon. A continuación se muestra en Tabla 2 los porcentajes de TP, FP y FN obtenidos en la mejor configuración de parámetros con λ=1 en la etapa de separación y con lambda=0.7 en la etapa de entrenamiento.

Estadísticos(%)	**SNR=-15dB**	**SNR=-10dB**	**SNR=-5dB**	**SNR=0dB**	**SNR=5dB**
TP	70 %	88%	96%	100%	100%
FP	0 %	0 %	0 %	0 %	0 %
FN	30 %	12 %	4 %	0 %	0 %

Tabla 2. Resultados del sistema de detección de señales de disparo usando SPARSE con λ=0.7 en la etapa de entrenamiento y λ=1 en la etapa de separación.

Para esta configuración (λ=1 en la etapa de separación y con lambda=0.7 en la etapa de entrenamiento) se va a calcular unas métricas comentadas en la sección anterior, dadas por las ecuaciones (64) y (65). Según estas, el sistema para la configuración descrita en la Tabla 2 presenta la siguiente exactitud y exhaustividad para cada relación SNR, se refleja en la Tabla 3:

Métricas para $\boldsymbol{\lambda_{entre.} = 0.7}$ $\boldsymbol{\lambda_{separ.} = 0.7}$	**SNR=-15dB**	**SNR=-10dB**	**SNR=-5dB**	**SNR=0dB**	**SNR=5dB**
Exhaustividad (Recall)	0.7	0.88	0.96	1	1
Exactitud (Precision)	1	1	1	1	1

Tabla 3. Métricas para configuración óptima para entornos ruidosos.

Como se puede comprobar el sistema ofrece muy buenos resultados a baja relación SNR y perfectos a partir de 0 dB. Se ha elegido esta configuración como la óptima para entornos ruidosos, ya que es la que mejores resultados ofrece, sin embargo, como se puede observar en los diagramas de barras, cuando se tiene una buena SNR no es necesario el uso de *SPARSE* para obtener muy buenos resultados. Además no existen

falsos positivos por lo que la exactitud es óptima, en el caso anterior también se comprobó que eran mínimos, obteniendo en el peor de los casos FP= 6%. Estos resultados tan pequeños de falsos negativos indican que en ambientes ruidosos no se detectará un disparo sin haberlo.

Sin embargo, a raíz del ruido ambiente con el que se pueda detectar falsos positivos se plantea una cuestión, qué ocurre con aquellos ruidos impulsivos que en frecuencia se asemejan a un disparo, como son palmadas, portazos, claxon, globos explotando, rotura de cristales, etc. Pues bien, se ha hecho un estudio sobre ello, evaluando estos ruidos como si se trataran de disparos, de manera que los TP indican que el sonido impulsivo ha sido interpretado como disparo, FN indica que no se ha detectado disparo (esto es lo que más conviene) y FP indica igualmente que se ha detectado disparo, pero no en el lugar esperado (la posición del disparo en muestras está orientada a la duración y forma de un disparo, por lo que es normal que no encaje para otro ruido impulsivo). Los resultados obtenidos vienen reflejados en las Figuras 50, 51, 52, 53 y 54.

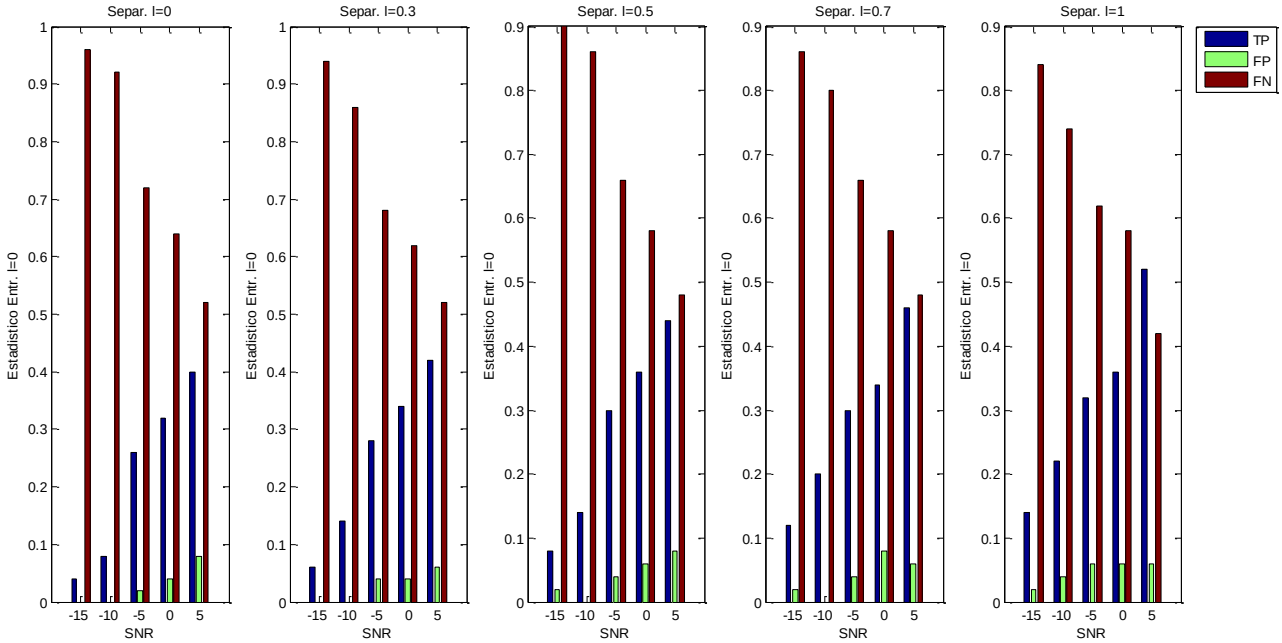


Figura 50. Diagrama de barras para $\lambda=0$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos BI.1.

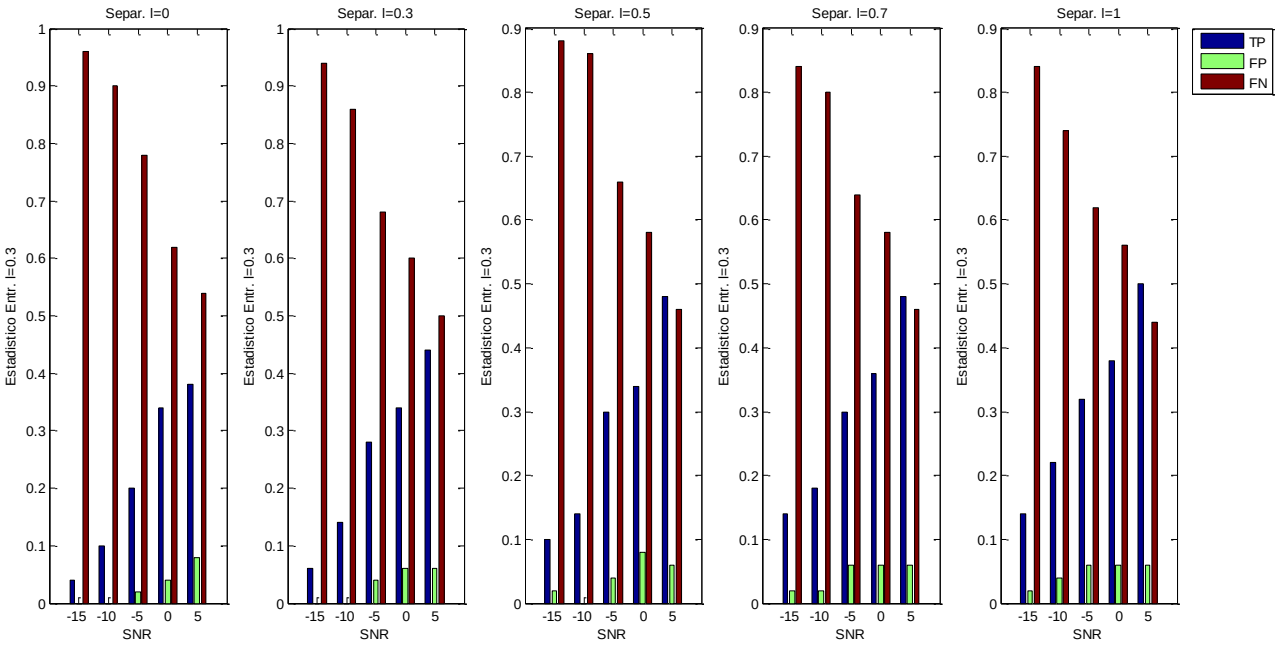


Figura 51. Diagrama de barras para $\lambda=0.3$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos BI.1.

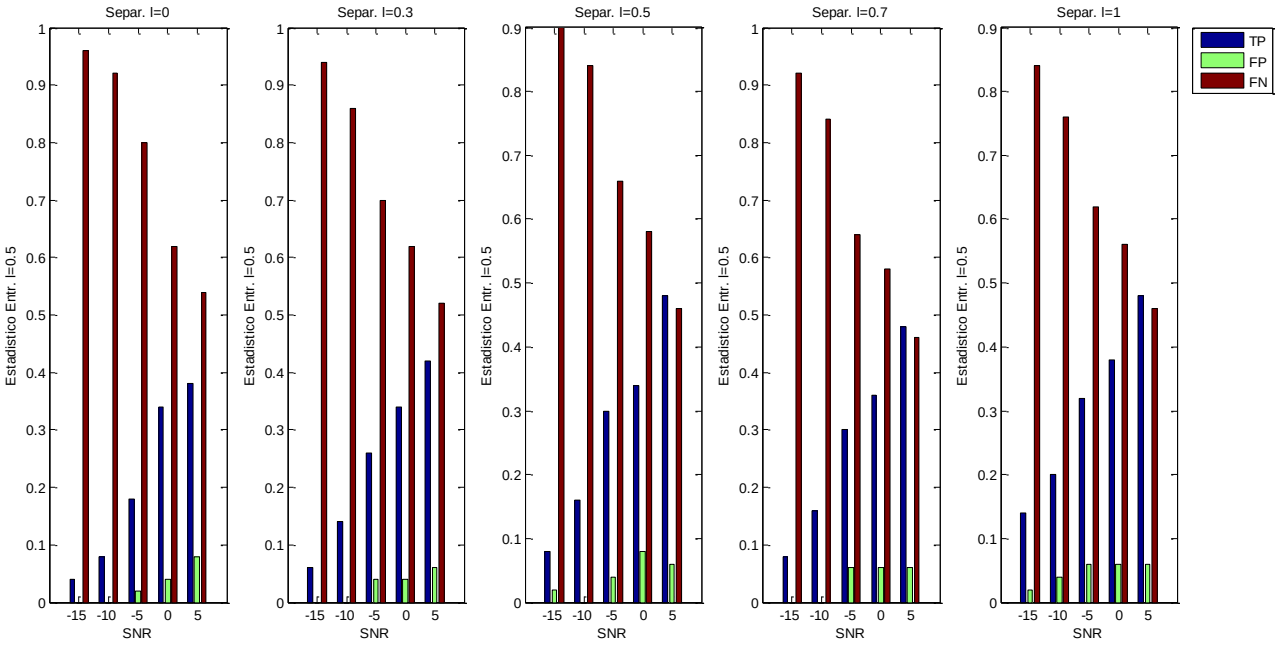


Figura 52. Diagrama de barras para $\lambda=0.5$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos BI.1.

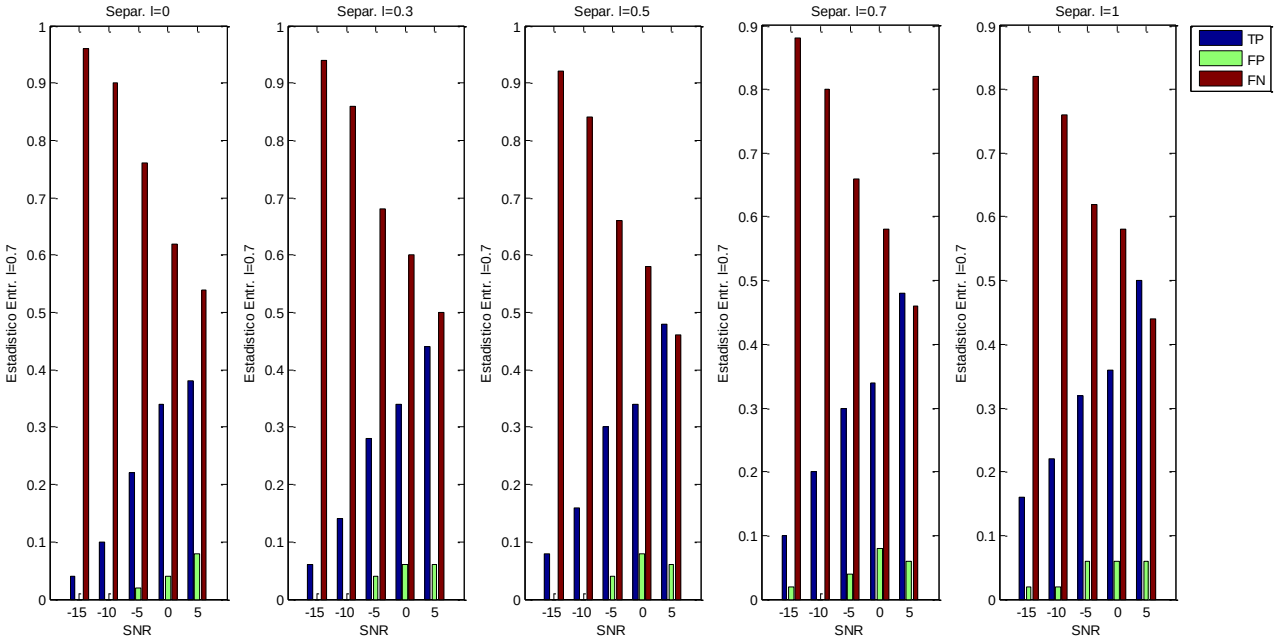


Figura 53. Diagrama de barras para $\lambda=0.7$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos BI.1.

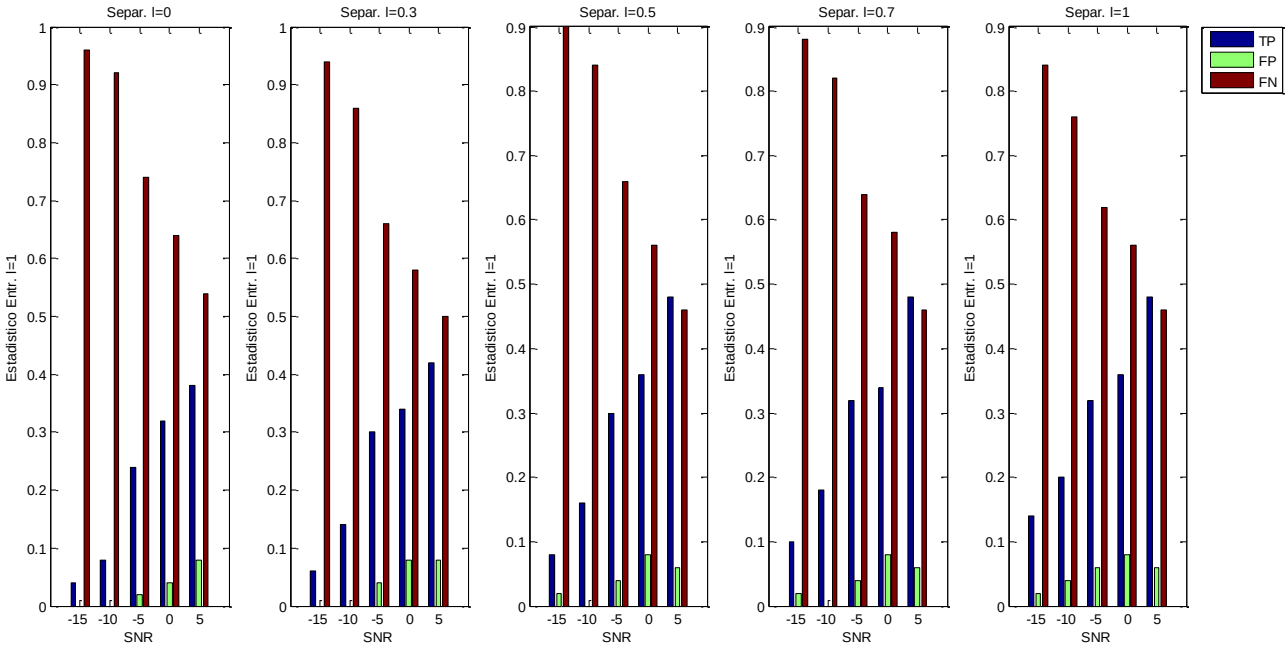


Figura 54. Diagrama de barras para $\lambda=1$ en la etapa de entrenamiento y $\lambda=[0, 0.3, 0.5, 0.7, 1]$ en la etapa de separación. Base de datos BI.1.

Viendo los resultados obtenidos, se observa que cuando la relación señal a ruido es muy pequeña, es decir, cuando el ruido de fondo enmascara al sonido impulsivo, apenas se detecta un ruido impulsivo como disparo, pero aún así existe un pequeño porcentaje. A medida que mejora la SNR y el sonido impulsivo se hace más fuerte se aprecia el incremento de los True Positive, por lo tanto el sistema implementado tiene problemas en

cuanto a la captura de ruidos impulsivos. Se puede ver como a medida que aumenta el parámetro λ en la etapa de separación, el sistema se hace más susceptible a confundir diversos sonidos con señales producidas por armas de fuego, mientras que el incremento de λ en la etapa de entrenamiento no supone prácticamente diferencia. Aún así, para los mismos análisis y simulaciones para disparo y otros sonidos impulsivos, se observa como el algoritmo se comporta mucho mejor con disparos. En la Tabla 4 se plasma la misma configuración, elegida como la óptima, que en la Tabla 2, pudiéndose comprobar las claras diferencias.

Estadísticos(%)	**SNR=-15dB**	**SNR=-10dB**	**SNR=-5dB**	**SNR=0dB**	**SNR=5dB**
TP	16 %	22 %	32 %	36 %	50%
FP	2 %	2 %	6 %	6 %	6 %
FN	82 %	76 %	62 %	58 %	44 %

Tabla 4. Resultados para un sistema de localización de disparos insertando otros sonidos impulsivos usando SPARSE con λ=0.7 en la etapa de entrenamiento y λ=1 en la etapa de separación.

Igual que en el caso anterior se va a calcular la exactitud y exhaustividad para el caso de sonidos impulsivos introducidos en el sistema detector de disparos (ver Tabla 5).

Métricas	**SNR=-15dB**	**SNR=-10dB**	**SNR=-5dB**	**SNR=0dB**	**SNR=5dB**
Exhaustividad (Recall)	0.163	0.224	0.34	0.383	0.532
Exactitud (Precision)	0.888	0.917	0.842	0.857	0.892

Tabla 5. Métricas para el sistema óptimo elegido usando señales de diferentes sonidos impulsivos.

Hay que tener presente que un sonido impulsivo ocasionado por palmadas, portazos o gritos, entre otros, no tiene la potencia que presenta un sonido ocasionado por un arma de fuego, así que es más probable encontrarse estos sonidos impulsivos en baja relación señal a ruido, a diferencia de los sonidos producidos por disparo. En diversas fuentes de información [4] [9] se ha realizado evaluaciones con relaciones señal a ruido de 20 a 30 dB. En este caso se ha comprobado que para disparos el sistema funciona perfectamente desde SNR de 0 dB. Comparando las Tablas 5 y 3, el algoritmo funciona perfectamente para disparos, pero no tan bien para diferentes sonidos impulsivos.

Por último, queda por comparar este algoritmo con otros ya implementados, para ello se usará como referencia [39]. En la sección 5.2.1 se detallaron los diferentes algoritmos

con los que se comparará el sistema implementado usando NMF y las métricas usadas para dicha evaluación. En las Tablas 6 y 7 se muestran los resultados de la evaluación realizada en [39].

Técnica	**Clasificador**	**TPR**	**FPR**
Correlación cruzada máxima	Umbralización	94.580	8.824
MFCC	SVM	97.321	50.000
LPC	SVM	99.702	11.458
LPC + Correlación Cruzada	SVM	99.702	8.333

Tabla 6. Porcentajes de detección para diferentes técnicas con Kern lineal y SVMs.

Técnica	**TPR**	**FPR**
Correlación cruzada máxima	97.917	10.417
MFCC	92.497	15.625
LPC	96.429	1.042
LPC + Correlación Cruzada	97.619	~0

Tabla 7. Porcentajes de detección para diferentes técnicas con RBF Kern y SVMs.

Se comprueba que este paper propone una mejora en la disminución de FPR, que es el porcentaje de señales de gritos, portazos, etc identificadas como disparos. Con la disminución de FPR también se produce un decremento en TPR. A continuación, en la Tabla 8 se realiza la misma evaluación con el algoritmo implementado en este trabajo, usando la misma base de datos que se usa en las Tablas 6 y 7. Estas bases de datos se han mezclado con ruido blanco gaussiano de media 0, y se ha establecido una relación señal a ruido de 20 dB, tal y como se indica en este paper.

Técnica	SPARSE	TPR	FPR
NMF	λ=0.7 en entrenamiento λ=1 en separación	100	87.255
NMF	λ=0 en entrenamiento λ=0 en separación	100	83.333

Tabla 8. Porcentajes de detección para NMF con la configuración óptima y la peor en entornos de ruido.

Como se puede ver, comparando las Tablas 6, 7 y 8, se aprecia como la técnica NMF ofrece mejores resultados de detección de disparos en situaciones de muy buena SNR, ya que no hay información sobre cómo funcionan otros métodos en entornos ruidosos no se ha podido establecer una comparativa en este nivel. Sin embargo, se observa que NMF también ofrece buenos resultados en cuanto a la detección de otros ruidos impulsivos usando un sistema de detección de disparo, sobre esto saca mucha ventaja las técnicas evaluadas en [39].

Teniendo en cuenta los resultados obtenidos, se ha tomado, como ya se ha visto, un *SPARSE* con λ=0.7 en entrenamiento y λ=1 en separación para optimizar el sistema en cuanto a mejora contra el ruido. Considerando este uso de SPARSE, se ha evaluado un parámetro que igualmente podría ofrecer mejoras en la detección según el valor que se le asigne. El número de bases seleccionadas a la hora de establecer la factorizción matricial puede verse modificado. En toda la implementación del sistema se ha tomado K=5, pero ahora se va a proceder a un estudio de diferentes valores y así comprobar si un número de bases mayor o inferior mejora en cuanto a la detección de disparo disminuyendo los falsos negativos y positivos. El análisis se ha realizado para valores de K= 2,3,5 y 7, considerando únicamente el caso óptimo del uso de SPARSE. Igualmente, como en evaluaciones anteriores, este análisis se va a realizar para diferentes relaciones señal a ruido, para así comprobar que tan robustos son. En la Figura 55 se representa un diagrama de barras que indica la proporción de TP, FP y FN para cada una de las SNR consideradas y del número de bases oportuno.

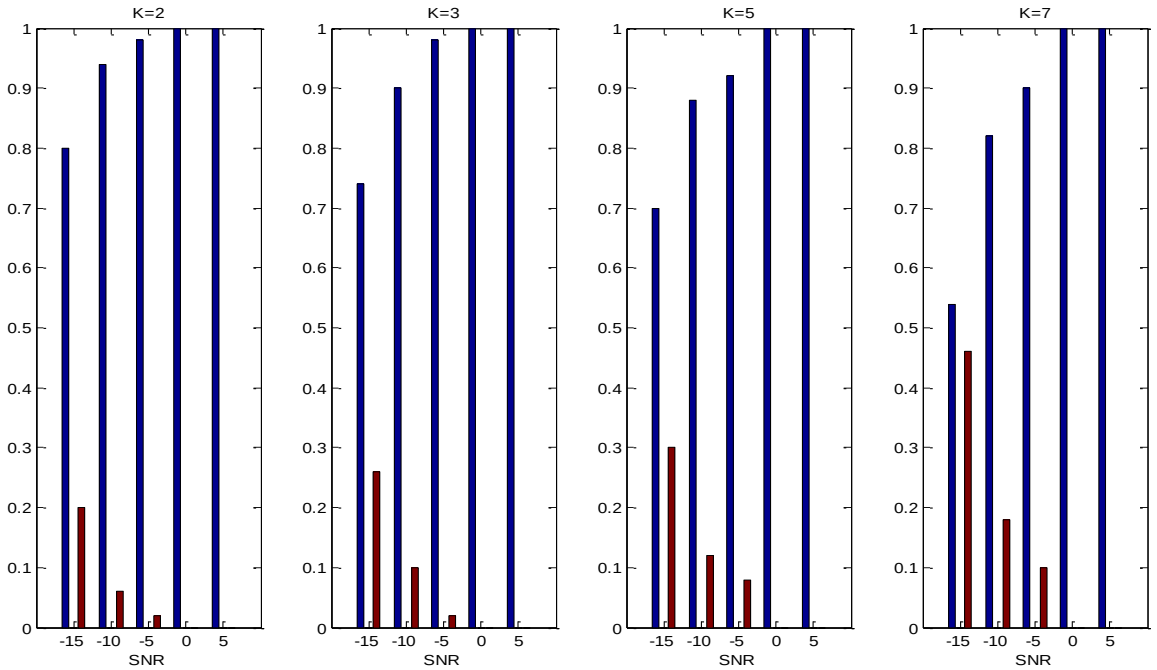


Figura 55. Diagrama de barras para un número de bases K=2, 3, 5 y 7 para diferentes SNR.

Observando los resultados obtenidos, se comprueba como disminuyendo el número de bases, que constituyen la dimensión de columnas de la matriz no negativa W y la de filas de la matriz no negativa H, los falsos negativos que se producen en baja SNR se ven disminuidos, constituyendo una mayor probabilidad de TP, ya que, igual que en evaluaciones anteriores, no se ha reflejado ningún FP. Para realizar esta evaluación se ha utilizado la base de datos 2 de ruido, obtenida de [39]. De modo que, si se precisa de mayor fiabilidad en cuanto a examinar señales de elevado ruido de fondo, se elegirá un número de bases K=2. Para esta evaluación se ha requerido implementar desde cero la etapa de entrenamiento, ya que la matriz no negativa W_d es diferente para cada K evaluado.

Un concepto a tener en cuenta en la etapa de detección es el efecto Doppler, el cual consiste en el aparente cambio de frecuencia de una onda producido por el movimiento de la fuente con respecto al observador o viceversa. Se estudia igualmente este concepto, ya que en el caso dado, el vehículo puede estar en movimiento y ello puede perjudicar a la detección del disparo por un cambio de frecuencia aparente. El cambio de frecuencia producido por el Efecto Doppler viene dado por:

$$f' = f\left(1 \pm \frac{v_0}{v}\right) \tag{69}$$

donde v_0 es la velocidad a la que se mueve el observador, v es la velocidad del sonido (343 m/s)y el operador $\pm$ indica si el objeto captador se aleja o se acerca a la fuente. Suponiendo que el vehículo se aleje de la fuente sonora a una velocidad de 100 Km/h (27.778 m/s) y teniendo en cuenta que la señal de un disparo es una señal en frecuencia de banda ancha donde concentra su energía en torno a f=1 KHz pero puede seguir presentando información hasta los 20 KHz, se realizan los cálculos obteniendo $f' = 917\ Hz$. De modo que sabiendo que la señal de disparo es una señal de banda ancha, no va a influir esta variación aparente de frecuencia.

5.4.2. Resultados de etapa de localización

En las Figuras 56 y 57 aparecen las gráficas en las que se refleja el error absoluto y la desviación típica de cada método con respecto al valor del ángulo real. A la derecha de todas las figuras se informa del color de la línea que representa a cada método, de manera que podemos identificar fácilmente el estimador que se desvía de la localización de la fuente sonora.

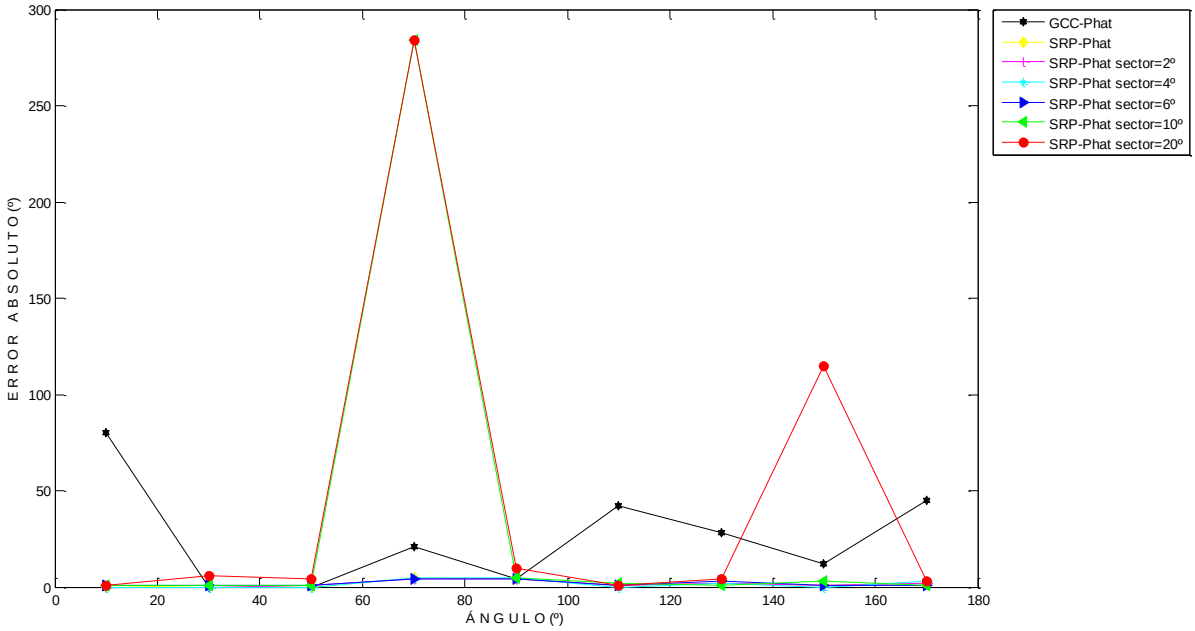


Figura 56. Error absoluto para GCC-Phat, SRP-Phat y SRP-Phat modificado, usando la base de datos de localización.

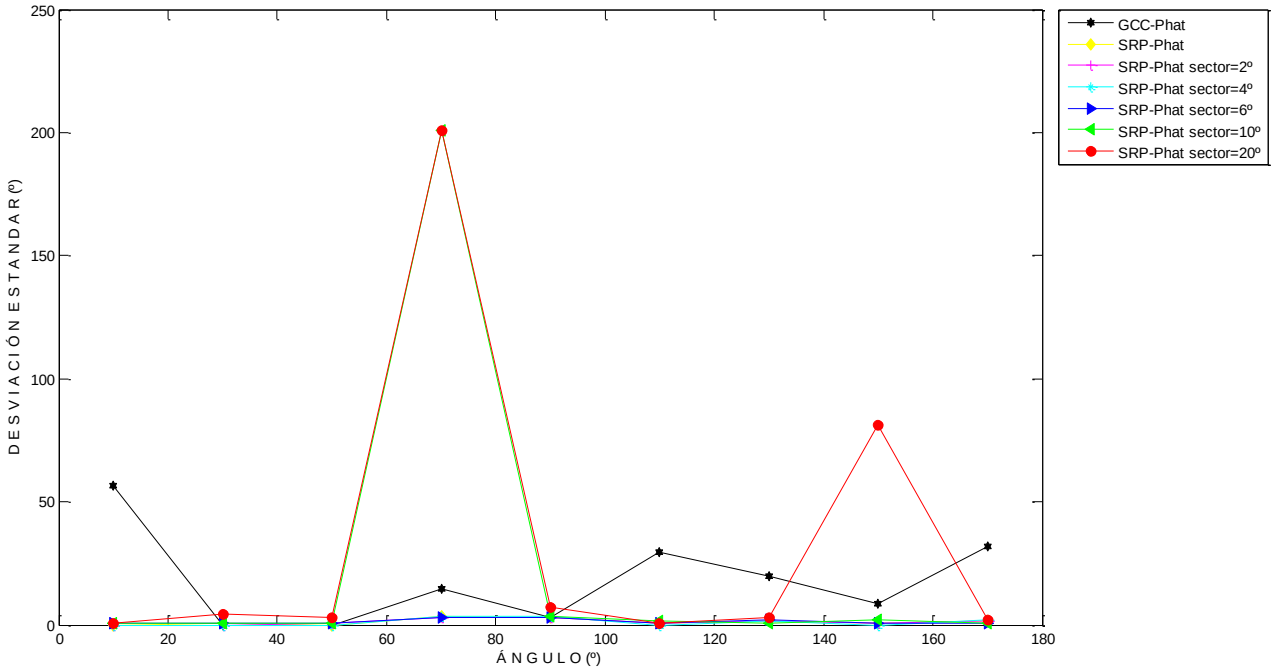


Figura 57. Desviación estándar para GCC-Phat, SRP-Phat y SRP-Phat modificado, usando la base de datos de localización.

MÉTODO	E_{abs} *medio* (°)	σ_s(°)
GCC-Phat	25.7778	18.2276
SRP-Phat	1.7778	1.2571
SRP-Phat sector=2°	1.6667	1.1785
SRP-Phat sector=4°	1.6667	1.1785
SRP-Phat sector=6°	1.8889	1.3356
SRP-Phat sector=10°	33.2222	23.4917
SRP-Phat sector=20°	47.5556	33.6269

Tabla 9. Error absoluto medio y desviación típica media para los métodos implementados.

La Tabla 9 constituye un resumen de los resultados obtenidos de la evaluación de todas las pistas con diferente localización de la fuente sonora. Se observa que el método que presenta un menor error es el SRP-Phat con un sector de 2°, seguido por un sector de 4° y por el método SRP-Phat original. Se puede ver en las Figuras como los métodos SRP-Phat con un sector de 10° y 20° presentan un error enorme para ciertos ángulos de localización, lo que origina que el error medio y la desviación estándar se dispare, ya que para el resto de ángulos se aprecia que ofrecen resultados muy precisos con respecto al ángulo real. Como premisa se tenía que con sectores mayores se tendría mayor robustez en el cálculo de localización, pero aquí se aprecia como en muy buenas condiciones de

señal a ruido, como las encontradas en las grabaciones realizadas en la Escuela Politécnica de Linares, puede obtenerse errores tan extremos que aumenten el error medio y con ello la fiabilidad del sistema. Más adelante se verá qué tal se comporta SRP-Phat para entornos más ruidosos. Mientras, el método GCC-Phat tampoco se comporta muy bien, presenta errores para bastantes localizaciones, por lo que el cálculo de la media de las diferentes métricas es elevado.

A continuación se representa en la Figura 58 es estudio de diferentes escenarios para ciertas localizaciones de la fuente sonora. Estos escenarios vienen dados por el hecho de que el vehículo se encuentre arrancado, acelerando o apagado como en las Figuras anteriores se ha visto.

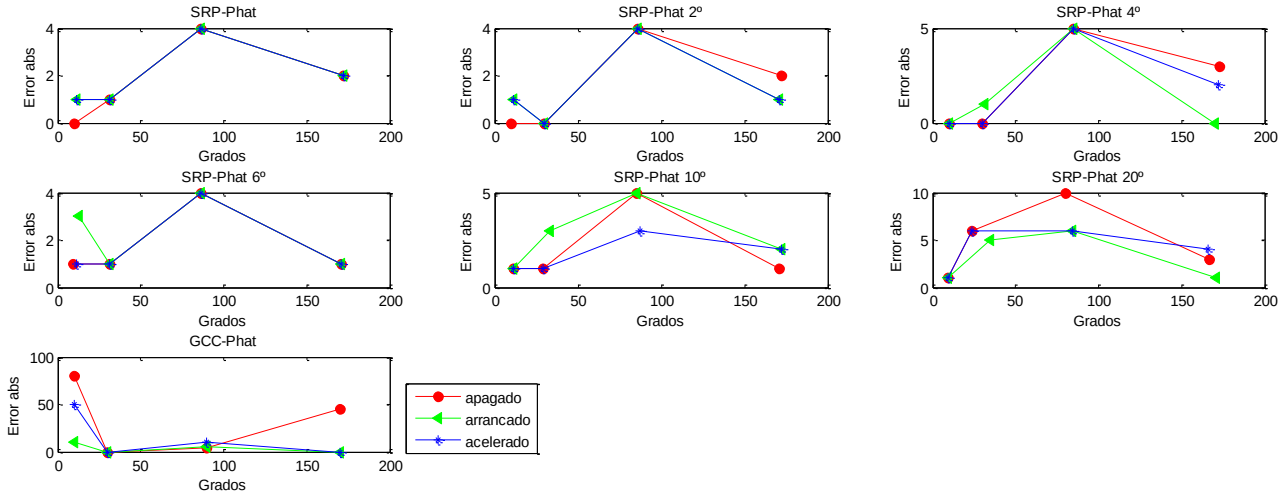


Figura 58. Error absoluto resultante de la evaluación de todos los métodos en los casos de motor apagado, arrancado y acelerado.

En la Figura 58 se aprecia que no hay diferencia a destacar en las tres situaciones planteadas, de manera que los algoritmos se comportan de manera similar. Se puede ver claramente como el método GCC-Phat es el que peor se comporta para estas localizaciones de la fuente, coincidiendo con las gráficas anteriores.

A continuación se va a entrar en un campo de interés para este tipo de algoritmos, y se trata de comprobar cómo se comportan ante diferentes relaciones señal a ruido. Así que el principal objetivo de este análisis es comparar la robustez frente al ruido de los diferentes métodos. Las Figuras 59, 60 y 61 servirán para dicho estudio, donde tanto el error absoluto, el RMSE (raíz del error cuadrático medio) y la desviación típica aportan información sobre cuánto se parece el valor estimado por los algoritmo al valor real.

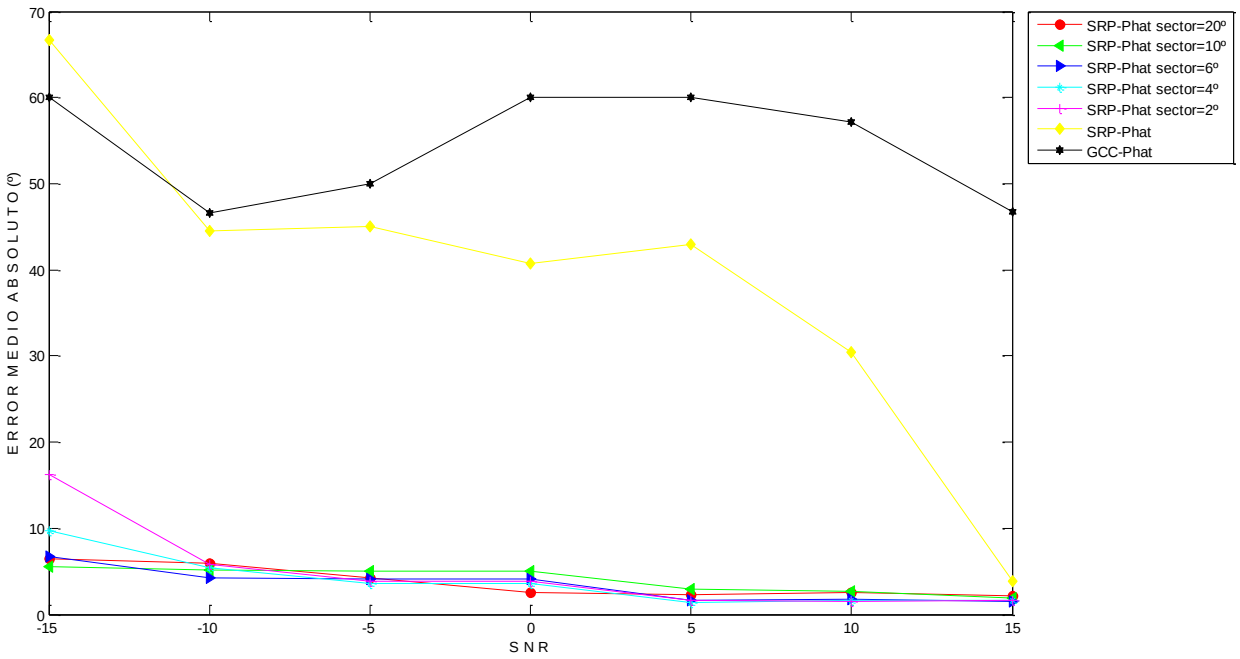


Figura 59. Error absoluto medio de cada método para diferentes relaciones señal a ruido.

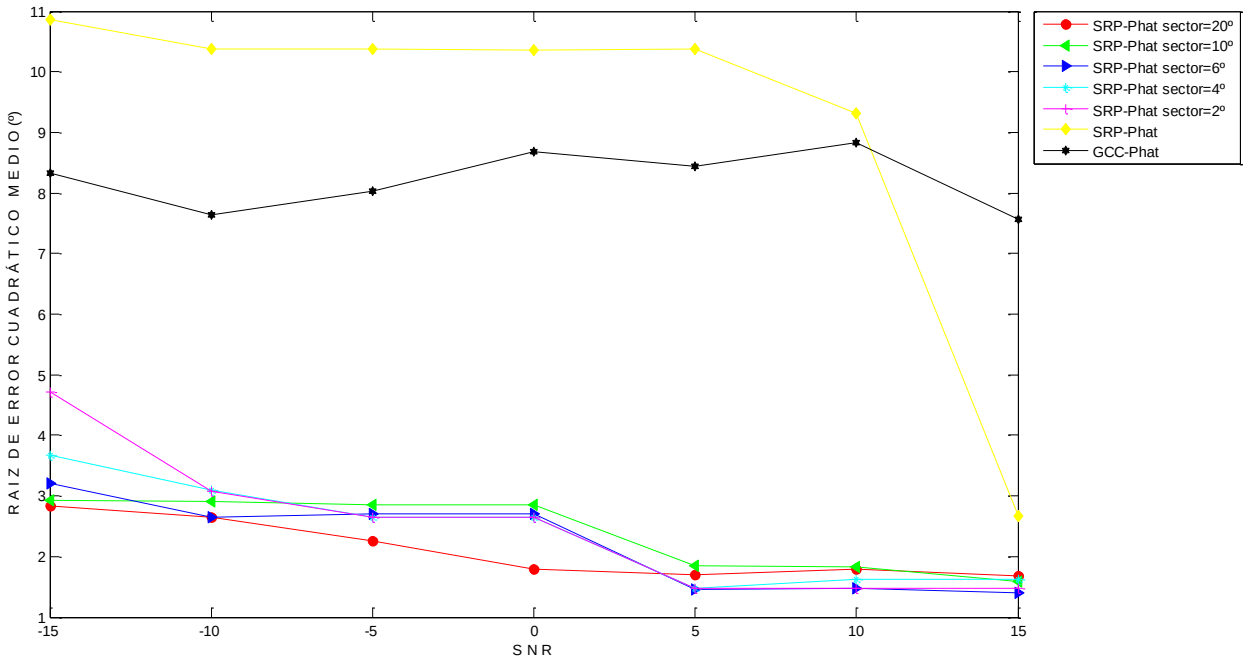


Figura 60. RMSE de cada método para diferentes relaciones señal a ruido.

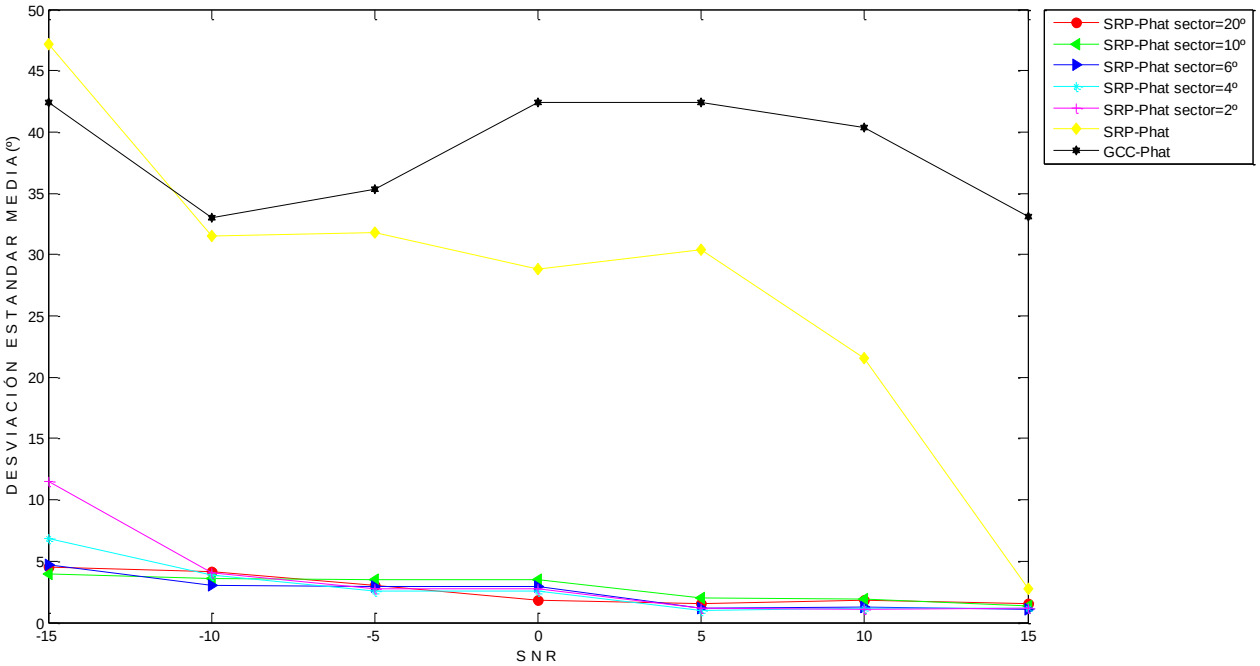


Figura 61. Desviación estándar media de cada método para diferentes relaciones señal a ruido.

En estas gráficas (Figuras 59, 69 y 61) se interpreta a primera vista la gran diferencia en cuanto a robustez que presenta el SRP-Phat modificado. Dicho esto, se va a realizar un estudio más exhaustivo sobre cómo se comportan frente al ruido observando estas gráficas y los diagramas de barras de las Figuras 62, 63 y 64, donde tal vez se observe con más claridad lo que se ha pretendido plasmar en estas gráficas.

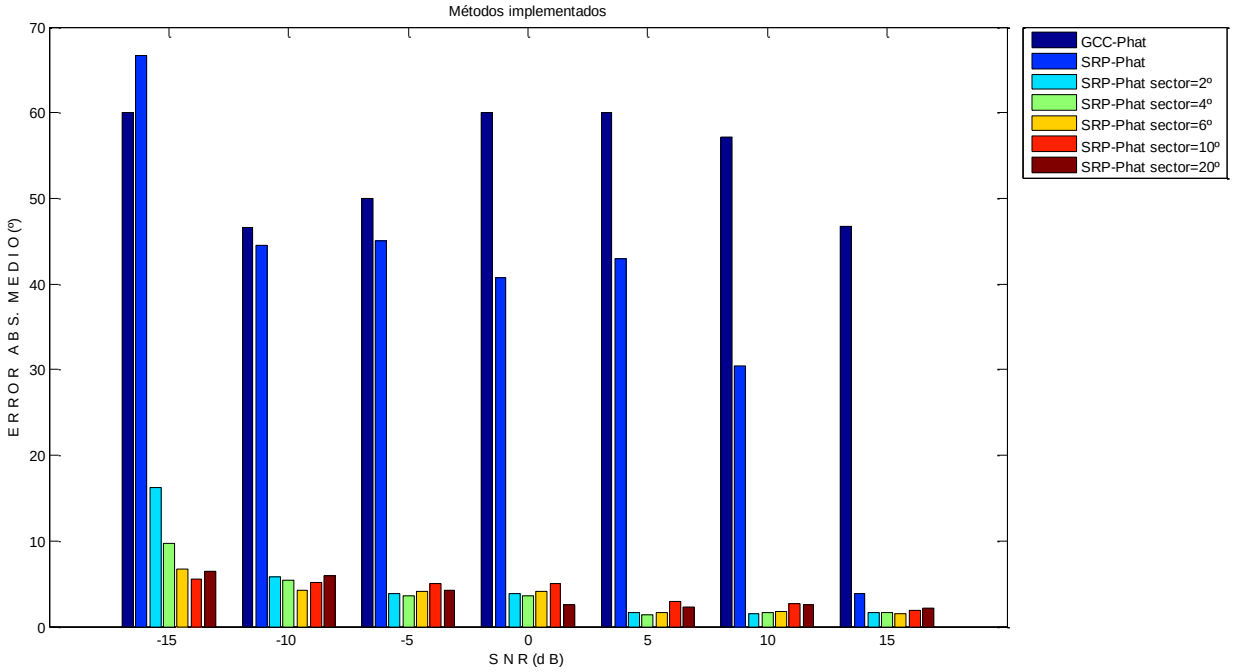


Figura 62. Diagrama de barras del error absoluto medio de cada método para diferentes relaciones señal a ruido.

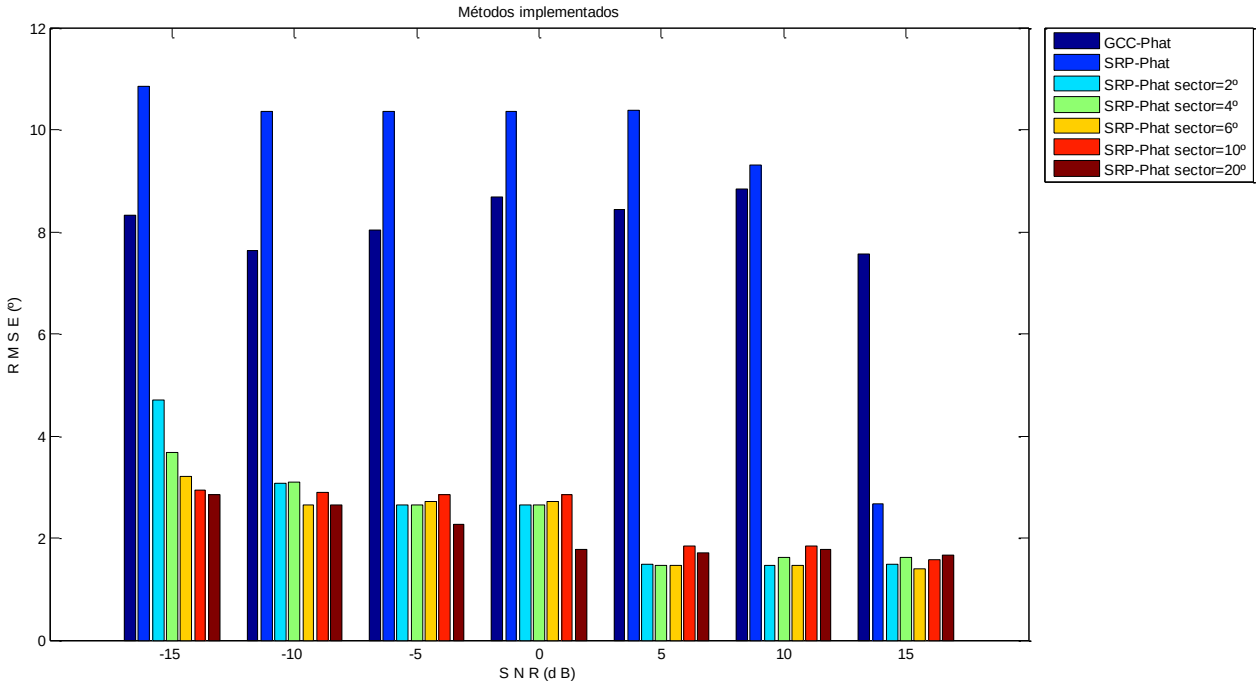


Figura 63. Diagrama de barras del RMSE de cada método para diferentes relaciones señal a ruido.

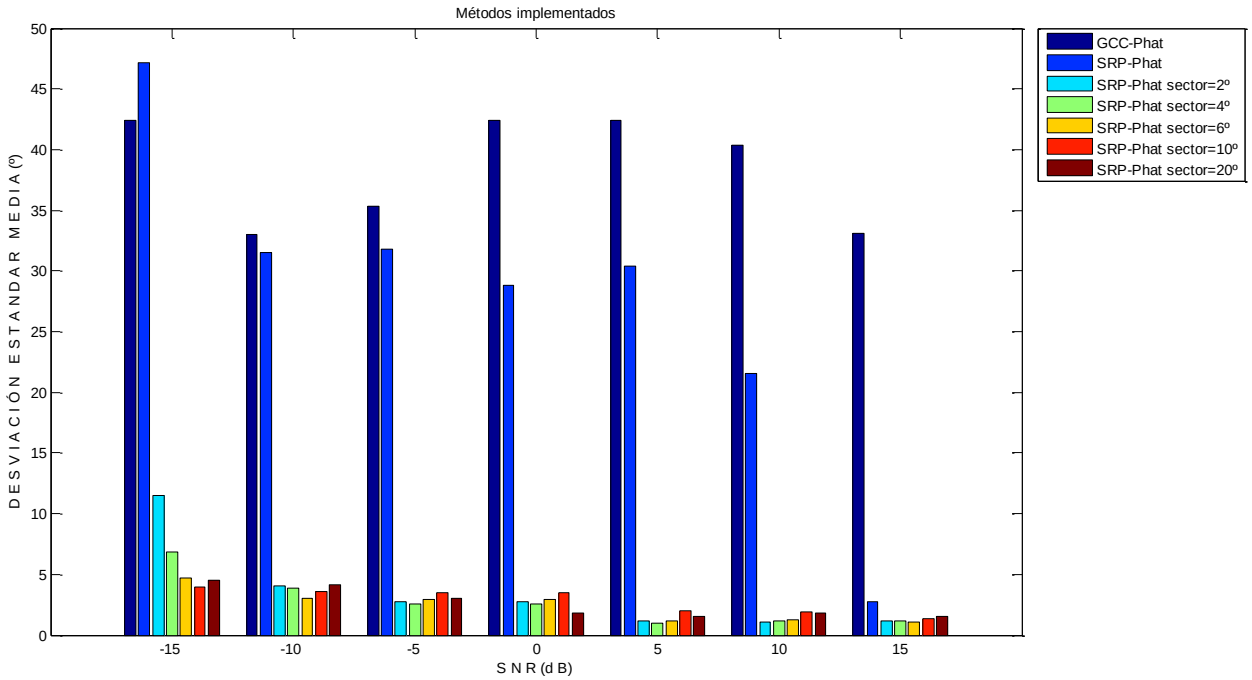


Figura 64. Diagrama de barras del desviación típica media de cada método para diferentes relaciones señal a ruido.

La relación señal a ruido como se puede comprobar va de -15 dB a 15 dB. Con esta vista global lo que se saca en conclusión es como el método GCC-Phat es el que peor funciona para todos los casos planteados si se observa las Figuras 59 y 62, que corresponden a la gráfica y diagrama de barras del error absoluto. El GCC-Phat presenta bastante desviación con respecto al ángulo en que se sitúa la fuente sonora producida por el arma de fuego. Este error se ve ligeramente mejorado al aumentar la relación señal a ruido, de modo que su robustez frente al ruido deja mucho que desear, y en cuanto a la mejor situación planteada, SNR=15 dB, los demás métodos garantizan mejores resultados que el GCC-Phat. Con todo esto, este método no tiene ninguna ventaja respecto al resto, de manera que no será elegido para ningún tipo de configuración.

Como segundo punto, cabe hablar del SRP-Phat original. En los diagramas de barras y gráficas se observa como tampoco presenta buenos resultados en situaciones de muy mala SNR en espacio abierto, según diversas referencias, este método es muy robusto frente a reverberación en salas, pero en el caso estudiado, no es el que mejores resultados presenta. Dicho esto, se observa como su mejora en cuanto a localización es abismal cuando mejora la SNR, a mayor relación señal a ruido mejor funciona este método. Si se hubiera seguido mejorando la SNR, este habría obtenido los mejores resultados, pero se han estudiado situaciones no tan favorables como las que habría en una sala de conferencia [18][22]. Otro punto a señalar es que, observando la Figura 63, se aprecia como SRP-Phat es el que peores resultados en cuando a RMSE presenta,

hasta que se produce esa increíble mejora al verse aumentada la SNR; esto es debido a que se han producido errores puntuales en la estimación de ciertos ángulos tan grandes que han hecho que el RMSE se dispare. Los errores producidos por el SRP-Phat no son tan constantes como los de GCC-Phat.

Por último, sólo queda el estudio del método SRP-Phat modificado, implementado para diferente sectorización. De manera global, los resultados que proporciona este método son muy buenos en baja SNR, ganando a sus métodos antecesores. Examinando el comportamiento a -15 dB, se puede sacar como conclusiones que, dentro de toda la sectorización considerada, un mayor sector considerado se comporta mejor en muy malas condiciones de ruido, ya que se seleccionan muchas más muestras y ello propicia a una mayor robustez al ruido. En el caso de -15 dB, el método que mejores resultados ofrece es el SRP-Phat con un sector de 10°. A medida que va mejorando la relación señal a ruido, mejoran los resultados ofrecidos por el método SRP-Phat modificado y en general del SRP-Phat y GCC-Phat.

Mientras que el método SRP-Phat modificado con una sectorización de 20º es el que resulta más robusto en condiciones de SNR muy bajas. Se puede ver también como el método SRP-Phat original mejora conforme mejora las condiciones de señal a ruido, dando mejores resultados y más exactos que ningún otro método en ambientes con muy bajo ruido (se debe tener en cuenta que las medidas de ángulos consideradas como reales pueden tener fallos de un par de ángulos, de modo que SRP-Phat en buenas condiciones de SNR podría dar aún mejores resultados incluso tener un error medio de 0º si se pudiera obtener la total exactitud de las grabaciones hechas en el patio). Otras conclusiones que se pueden sacar es que el método GCC da errores disparatados en ciertos ángulos haciendo que su error medio se dispare, por tanto se demuestra cómo no es un método robusto frente al ruido, y lo mismo ocurre con el SRP-Phat original en condiciones de mala SNR. El método propuesto en este proyecto experimenta una mejora de resultados obtenidos progresiva, es decir, va mejorando poco a poco conforme mejora la SNR, apreciándose como va ganando en exactitud los métodos con una menor sectorización cuando se está en un ambiente cada vez menos ruidoso. Por ello una configuración única para localización no sería lo más apropiado, se podría elegir el algoritmo empleado para la estimación de la localización dependiendo del ruido de fondo, así que en situaciones con muy poco ruido se optaría por el empleo del método SRP-Phat modificado con un sector de 2º o el original, mientras que si se está en una situación con mucho ruido, que incluso pueda enmascarar la fuente sonora, se elige un método SRP-Phat modificado con un sector de entre 10° a 20°. El resumen se plasma en la Tabla 10.

SNR	Método Óptimo
< -15dB	Método SRP-Phat con sector de 20°
-10 dB a 0 dB	Método SRP-Phat con sector de 10° o 6°
0 dB a 15 dB	Método SRP-Phat con sector de 4° o 2°
>15 dB	Método SRP-Phat original

Tabla 10. Resumen de métodos óptimos según la SNR del entorno.

Otra cuestión a analizar, viendo lo ocurrido, sobre todo con el método SRP-Phat original y el GCC-Phat es que presentan estimaciones muy precisas y otras disparatadas que ocasionan que su error medio se dispare, esto es a causa de su falta de robustez en entornos ruidosos.

En definitiva, en base a todos los resultados obtenidos, cuando se pretende utilizar técnicas de localización basadas en TDOA o beamforming en el exterior, cuando se pueden dar casos de mucho ruido, los datos obtenidos indican que la mayor robustez la consigue el método modificado.

5.4.3. Coste computacional

La técnica NMF no es destacada por su bajo coste computacional, debido a que los cálculos que se realizan son pesados debido al alto número de iteraciones. Igualmente ocurre con el método de localización SRP-Phat, que al tratarse de un beamformer que examina todo el espacio resulta lento en cálculo.

Dicho esto, se puede reducir el coste de la técnica NMF almacenando la matriz de interés de la etapa de entrenamiento W_d y no volviendo a tener que ejecutar esta etapa. En cuanto a la etapa de separación se ha visto reducida en cierta medida con la compresión de los bins en frecuencia del espectro de la señal de entrada. Los datos temporales plasmados a continuación son los obtenidos para una señal de entrada de 4 segundos, que en tiempo real será mucho menor. Para la implementación en tiempo real se deberían tomar tramas de milésimas de segundo, ya que esa es la duración del disparo (3-4 ms).

El proyecto ha sido realizado en un PC con sistema operativo Windows 7, un procesador *intel Core i5* con *CPU* a *2.4 GHz* y una *RAM* de 4 GHz. La prolongación de este proyecto con la beca asociada será la implementación en tiempo real del sistema usando el entorno CUDA y su capacidad de paralelización, que sumado a las cortas tramas temporales procesadas debe ofrecer buenos resultados de tiempo de ejecución.

En la Tabla 11 se muestra los resultados temporales obtenidos en la técnica NMF, tanto para la etapa de separación como para la de entrenamiento con un tamaño de ventana de 2048 muestras y diferente número de bases K. Se observa como variar el número de bases implica una diferencia en el coste computacional de la etapa de separación. En el caso de usar un número de bases K=5 y el tamaño de ventana citado, la etapa de entrenamiento no se ejecutará, ya que los resultados para estos parámetros están almacenados y únicamente se cuenta con el tiempo de la etapa de separación, igualmente se podría almacenar el parámetro W_d para los diferentes casos. Como ya se ha explicado a lo largo de la memoria, en la etapa de entrenamiento se deben procesar 150 señales de disparo de unos y la etapa de separación procesa señales de entrada de 4 segundos.

NMF	**Tiempo de ejecución**
Etapa de entrenamiento K=2	208.802 s
Etapa de separación K'=2*150	6.772 s
Etapa de entrenamiento K=3	125.67 s
Etapa de separación K'=3*150	8. 697 s
Etapa de entrenamiento K=5	0 s
Etapa de separación K'=5*150	23.639 s
Etapa de entrenamiento K=7	123.882 s
Etapa de separación K'=7*150	35.87 s

Tabla 11. Coste computacional de la etapa de detección.

Tal y como se ha mencionado, la etapa de entrenamiento no será necesaria ejecutarla más de una sola vez si se almacena su matriz de patrones espectrales. La etapa de separación usando un número de bases pequeño ofrece la mejor opción temporalmente hablando y en cuento a la disminución de falsos negativos para baja SNR.

En cuanto a la etapa de localización, el coste computacional de todos los métodos descritos será similar y se muestra en la Tabla 12.

Método de Localización	Tiempo de ejecución
GCC-Phat	6.997 s
SRP-Phat	4.068 s
SRP-Phat modificado. Sector 2°	6.143 s
SRP-Phat modificado. Sector 4°	7.034 s
SRP-Phat modificado. Sector 6°	6.586 s
SRP-Phat modificado. Sector 10°	4.587 s
SRP-Phat modificado. Sector 20°	6.285 s

Tabla 12. Coste computacional para la etapa de localización

CAPÍTULO 6. CONCLUSIONES

El Trabajo Fin de Grado desarrollado pretende crear un sistema que estime la localización de una fuente sonora producida por un arma de fuego disparada por un francotirador a una distancia considerable del vehículo para que se pueda suponer condiciones de campo lejano (>10 m). Los objetivos marcados en el Capítulo 1 se han superado con creces, ya que se ha conseguido tanto localizar la fuente sonora predominante, como detectar que esa fuente sonora proviene de un arma de fuego para así ejecutar su localización.

En vista a los resultados obtenidos, analizados en el Capítulo 5, se concluye que la detección del disparo usando NMF en situaciones normales de ruido de fondo son inmejorables, es decir, se detecta el 100% de las fuentes sonoras producidas por un disparo, superando así los resultados obtenidos con otros métodos de detección, que también han sido analizados en dicho capítulo. Por otro lado, la gran precisión en la detección de disparos, implica que otros sonidos impulsivos, con características similares a las de una señal de disparo, puedan ser detectados erróneamente como disparo, ya que no se ha implementado ningún mecanismo que evite confundir ruidos impulsivos con sonidos producidos por armas de fuego. Este hecho no va a ser muy relevante en el sistema diseñado, ya que un sonido impulsivo situado a cierta distancia de los sensores de captura no tendrá muy buena relación señal a ruido y será más fácilmente enmascarado por el ruido de fondo que un disparo, el cual puede superar los 100 dB de ruido. En cuanto a la relación de la potencia de la señal de disparo y el ruido de fondo, está claro que en SNR muy bajas se producirá un aumento de falsos negativos. Para el caso de baja SNR se optimiza la detección eligiendo un *SPARSE* de $\lambda=0.7$ para la etapa de entrenamiento y $\lambda=1$ para la etapa de separación, obteniendo así un 70 % de exhaustividad y un 100% de precisión (Tabla 2) para una SNR= -15 dB. Por otro lado, también se ha comprobado como un número de bases K menor disminuyen los falsos positivos en situaciones de baja relación señal a ruido.

En cuanto al coste computacional de la etapa de detección implementada en Matlab, es algo que se debe mejorar, ya que la técnica NMF conlleva un tiempo de ejecución bastante alto, debido al número de iteraciones y la dificultad de las ecuaciones a resolver. Se ha comprobado como el uso de un número de bases inferior K disminuye el coste computacional de la etapa de localización, además el procesar un menos número de muestras también propiciará a su futura implementación en tiempo real.

Por otro lado, atendiendo a los resultados obtenidos para la localización de un disparo, se concluye que, cuando se está en ambientes muy poco ruidosos, con muy buena relación señal a ruido, va a ganar en precisión aquellos métodos descendientes del SRP-Phat que tomen una menor sectorización o el SRP-Phat original, ya que el error entre el ángulo de localización estimado y el real es menor, mientras que en ambientes muy ruidosos, donde la señal de interés pueda incluso enmascararse, el resultado aportado por el método SRP-Phat modificado con una sectorización de 10° o algo mayor ofrece mayor robustez frente al ruido y por tanto más fiabilidad. De modo que el sistema final deberá adaptarse a las circunstancias de ruido y elegir el tipo de configuración más adecuado según lo explicado.

Igualmente, se debe considerar una conclusión sobre el coste computacional de la etapa de localización. El escaneo espacial para determinar de entre todas las posiciones candidatas la dirección de la fuente es lo que más tiempo consume. Debido a que los tres métodos usan la misma técnica, el coste computacional es similar para todos, de unos 6 segundos.

CAPÍTULO 7. LINEAS DE FUTURO

Las posibles líneas futuras de este Trabajo Fin de Grado son las que se expresan a continuación:

- Algo interesante para continuar con el sistema diseñado es llegar a su implementación en tiempo real, esto se puede realizar con ayuda del dispositivo FPGA. En [31] se indica cómo podría ser el sistema a implementar. Este debe consistir en un módulo de muestreo, otro de control (DPRAM, Dual-Port Random Access Memory) y otro último que debe realizar los cálculos para la localización de fuentes sonoras; como es obvio se necesita un módulo más para la detección de la fuente sonora producida por un arma de fuego.
- En base a las limitaciones que presenta el diseño propuesto en Matlab, como puede ser el coste computacional de la implementación de la técnica de detección basada en NMF y la técnica de localización basada en SRP-Phat, se puede seguir trabajando en este nivel para conseguir un menor tiempo de compilación con resultados igualmente buenos.
- Por otro lado, otro de los contras que se plantean es la alta tasa de falsos positivos resultantes del sistema de detección frente a sonidos impulsivos como gritos, roturas de cristal, portazos, entre otros. De modo que, una buena propuesta sería implementar nuevos métodos y técnicas que disminuya el porcentaje de falsos positivos, combinando técnicas de NMF con propuestas que podemos encontrar en [9] y [39].
- Viendo el avance de la tecnología móvil y cómo ésta se usa prácticamente para todo debido a la multitud de aplicaciones e interacciones con otros dispositivos digitales, podría incluirse este campo como línea futura del proyecto implementado. El dispositivo móvil estaría continuamente conectado al sistema montado en el vehículo, de manera que éste devolviera la información de localización del francotirador, avisando al personal de seguridad e incluso podría lanzar un aviso a los equipos policiales para avisar del delito.
- Por otro lado, sería interesante conseguir una localización de la fuente sonora producida por un disparo en tres dimensiones, para así contar con una mayor información

Como se puede ver, es posible seguir trabajando en esta línea de investigación en futuros proyectos, ampliando el campo o mejorando las limitaciones surgidas, buscándoles la mejor solución posible.

CAPÍTULO 8. BIBLIOGRAFÍA

[1] Jain, P., Orhan, A. And Wright, O. 'A Shot In The Dark: Point-Source Localization of a Gunshot', Final Project, Digital Communication and Signal Processing System Design, Carnegie Mellon Engineering, 2012.

[2] Maher, R. 'Acoustical Characterization of Gunshots', IEEE Workshop on Signal Processing Applications for Public Security and Forensics, 2007.

[3] Molnar, K., 'Muzzle blast detection via short time fourier transform', Minisymposium 2005 of the Department of Measurement and Information Systems of the Budapest University of Technology and Economics, 2005

[4] Chacon-Rodriguez, A., Julian, P., Castro, L., Alvarado, P. and Hernandez, N., 'Evaluation of Gunshot Detection Algorithms', IEEE Transactions on Circuits and Systems I: Regular Papers, 2011, vol. 58, pp. 363 – 373.

[5] Bandi, A.K. ; Rizkalla, M. ; Salama, P., 'A novel approach for the detection of gunshot events using sound source localization techniques', IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), 2012.

[6] Sallai, J., Hendgecock, W., Volgyesi, P., Nadas,A., et al. 'Weapon clasification and shooter localization using distributed multichanel acoustic sensor'. Journal of System Architecture- *Embedded Systems Design, 2010, 57(10):869-885*.

[7] Decreto 395/75. Reglamento de la Ley Nacional de Armas y Explosivos 20.429.

[8] Reglamento de Armas. Real Decreto 137/1993, de 29 de enero, BOE 55/1993.

[9] Freire, I. and Apolinario J., 'Gunshot detection in noisy environments', 7th International Telecommunications Symposium (ITS 2010), 2010.

[10] Machmer, T., Swerdlow, A., Kroschel, K. Moragues, J., Vergara, L., Gos´albez, J., 'Robust impulsive sound source localization by means of an energy detector for temporal alignment and pre-classification', EUSIPCO, Glasgow, Scotland, 2009

[11] Dufaux, A. 'Detection and Recognition of Impulsive Sound Signals', D. Sc. thesis, Institute of Microtechnology. University of Neuchâtel, Neuchâtel, Switzerland, 2001.

[12] Févotte, C., Bertin, N., Durrien, J., Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis'. Neural Computation, 2009, Vol. 21, Nº. 3: 793-830.

[13] Lee, D., Seung, H., 'Algorithms for Non-negative Matrix Factorization', 2001.

[14] Schmidt, M., Larsen, J. and Hsiao, F., 'Wind Noise Reduction using non-negative sparse coding', Machine Learning for Signal Processing, 2007 IEEE Workshop on, pp. 431-436.

[15] Eguchi, S. and Kano, Y. 'Robustify maximum likelihood estimation'. Technical report, Institute of Statistical Mathematics, June 2001.

[16] Tellakula, A. K. 'Acoustic source localization using time delay estimation'. PhD thesis, Indiand Institute of Science, August 2007.

[17] D. H. Johnson and D. E. Dudgeon, 'Array Signal Processing: Concepts and Techniques'. Prentice Hall, New Jersey, 1993.

[18] Torres, A., Cobos, M., Pueo, B., et al, 'Robust acoustic source localization based on modal beamforming and time-frequency processing using circular microphone arrays', Acustical Society of America, pp. 1511-1520, 2012

[19] M. S. Brandstein, 'A framework for speech source localization using sensor arrays', Ph.D. dissertation, Brown University, USA, 1995

[20] Chen, J., Benesty, J., and Huang, Y. 'Time delay estimation in room acoustic environments: an overview'. EURASIP Journal on Applied Signal Processing 2006 (2006), 1–19.

[21] Knapp, C. and Carter, G. 'The generalized correlation method for estimation of time delay', IEEE Trans. on Acoustic, Speech and Signal Processing, vol. ASSP-24, pp. 320–327, 1976.

[22] Zhang, C., Florencio, D. and Zhang, Z. 'Why does PHAT work well in low noise,reverberant environments?' in Proc. of ICASSP, Las Vegas, Nevada, 2008, pp. 2565–2568.

[23] Blandin, C., Ozerov, A and Vincent, E. 'Multi-source TDOA estimation in reverberant audio using angular spectra and clustering', Signal Processing, Elsevier, 2012, 92, pp. 1950-1960.

[24] Dibiase, J., 'A high accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays'. PhD thesis, Brown University, Providence, RI, May 2000.

[25] Martí, A., 'Multichannel Audio Processing for Speaker Localization, Separation and Environments', Doctoral Thesis, Valencia (Spain), July 2013.

[26] Cobos, M., Marti, A., and Lopez, J. J. 'A modified srp-phat functional for robust real-time sound source localization with scalable spatial sampling'. IEEE Signal Processing Letters, vol. 18, no. 1, 2011.

[27] Schmidt, R. 'Multiple emitter location and signal parameter estimation', IEEE Trans. on Antennas and Propagation, vol. 34, pp. 276–280, 1986.

[28] CUDA Programming Guide, NVIDIA Corporation, 2010, http://developer.nvidia.com.

[29] Vicente Peruffo Minotto, Claudio Rosito Jung, Luiz Gonzaga da Silveira, Jr and Bowon Lee, 'GPU-based approaches for real-time sound source localization using the SRP-PHAT algorithm', International Journal of High Performance Computing Applications, 2012.

[30] Virtex-4 Family Overview, DS112 (v3.1) August 30, 2010, http://www.xilinx.com/support/documentation/data_sheets/ds112.pdf

[31] Seunghun Jin, Dongkyun Kim, Hyung Soon Kim, Chang Hoon Lee, Jong Suk Choi and Jae Wook Jeon, 'Real-time Sound Source Localization System based on FPGA', the IEEE International Conference on Industrial Informatics (INDIN 2008), 2008.

[32] Dahlan, B., Mansoor, W., Abbasi, M., et al, 'Sound Source Localization for Automatic Camera Steering'. Networked Computing and Advanced Information Management, pp. 20-25, Gyeongju, 2011.

[33] Johanson, A., Grbic, N., and Nordholm, S., 'Speaker Localisation Using the Far-Field SRP-Phat in Conference Telephony '.

[34] Kwon, B., Park, Y., and Park, Y., 'Sound Source Localization for Robot Auditory System Using the Summed GCC Method', International Conference on Control, Automation and Systems 2008, Korea.

[35] Valin, J., Michaud, F., Rouat, J., and Létourneau, D., 'Robust Sound Localization Using a Microphone Array on a Mobile Robot', Canada.

[36]Pessentheiner, H., Petrik, S., Romsdorfer, H., 'Beamforming Using Circular Arrays for Distant Speech Recognition in Reverberant Environment and Double-Talk Scenarios', Proc. InterSpeech 2012, Portland, Oregon, USA.

[37] Johnson, Adam. 'Gun Sound Effects. Royalty Free Sound Effects' [en linea]. Actualizada: 6 de Febrero 2014. Fecha consulta: Marzo 2014. Disponible en: SFXsource.com

[38] 'Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database'. *ETSI* EG 202 396-1 V1.2.2 (2008-09).

[39] Ahmed, T., Uppal, M., Mhhamad, A., 'Improving Efficiency And Reliability Of Gunshot Detection Systems'. *IEEE ICASSP*, pp. 513-517, 2013.

[40] 'SoundBible.com. Free Sound Clips, Sound Bites, and Sound Effects'. Actualizada: 5 de Diciembre 2013. Fecha de consulta: Mayo 2014.Disponible en: www.soundbible.com

[41] Music Technology Group of Universitat Pompeu Fabra. 'Freesound'. Actualización: Marzo 2014. Fecha de consulta: Mayo 2014. Disponible en: http://www.freesound.org.

[42] Sallay, J., Lédeczi, A., Völgyesi, P. 'Acoustic Shooter Localization with a Minimal Number of Single-Channel Wireless Sensor Nodes'. *SenSys'11,* November 1–4, 2011, Seattle, WA, USA.

[43] 'Sniper Location & Gunshot Detection System'. Creación: 2008. Disponible en: http://defense-update.com/features/2008/november/231108_sniper_detection.html.

[44] 'Sniper Location System Turns Heads'. Actualización: 2010. Disponible en: http://www.vanderbilt.edu/magazines/vanderbilt-magazine/2009/08/sniper-location-system-turns-heads/.

ANEXO. Manual de usuario

Se ha implementado una sencilla aplicación en entorno Matlab del sistema diseñado en este proyecto. Para ello se ha usado la herramienta GUIDE de Matlab, permitiendo poner al alcance de cualquier usuario el análisis de señales de disparo que se han podido ver a lo largo de esta memoria.

Para el uso de esta aplicación es preciso contar con las bases de datos idóneas para poder evaluar distintas señales de audio, tanto para la detección como para la localización. Esta aplicación pretende simular la etapa de detección, introduciendo los parámetros necesarios según el interés del usuario y una vez se ha detectado la presencia de disparo, permite la localización del esta fuente sonora, siempre y cuando los archivos de audio usados estén compuestos por los seis canales correspondientes a los micrófonos.

A continuación se muestra un manual de usuario a seguir para poder llegar a la localización de una fuente sonora producida por un arma de fuego. En primer lugar se va a mostrar el entorno de trabajo usado, tal y como se aprecia en la Figura 65.

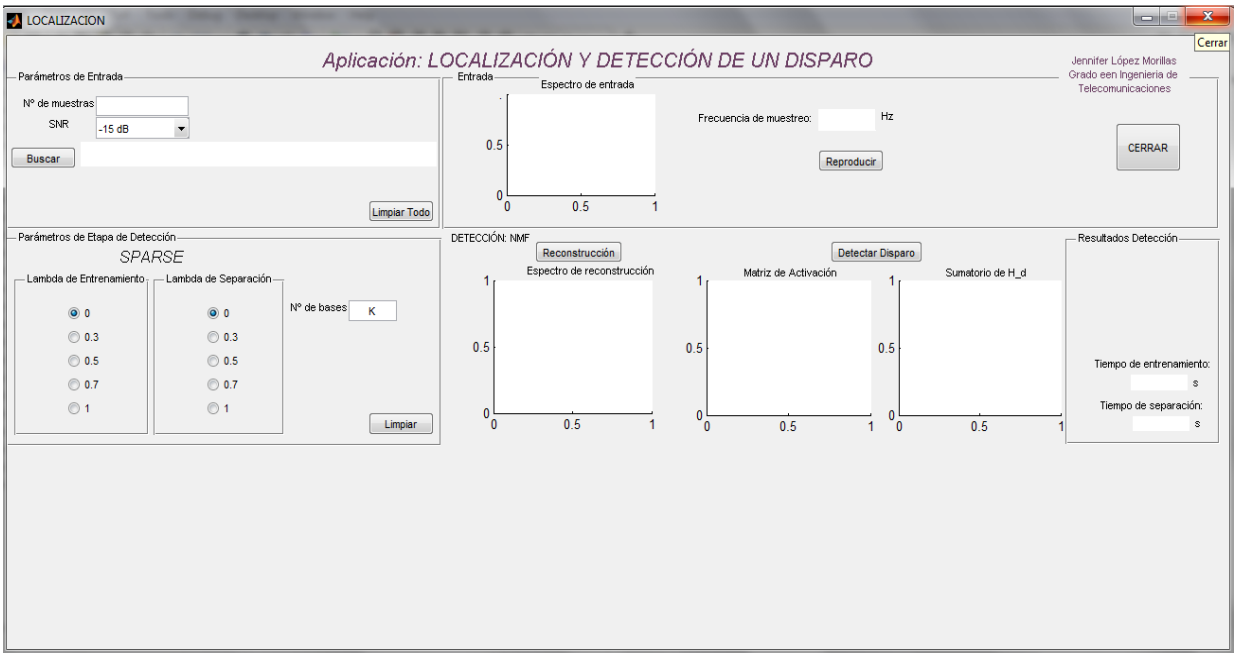


Figura 65. Vista de la pantalla principal de la aplicación para el usuario.

Como se puede ver, este consta de un panel de introducción de parámetros de entrada, otro de parámetros de entrada para la etapa de detección y la etapa de localización no se hará visible hasta que no se garantice que se ha producido disparo,

intentando así simular el sistema en su implementación real. En la parte central de la pantalla aparecen las gráficas y otros datos de interés y en la parte izquierda de la pantalla se muestra un panel con los resultados útiles de la implementación de la etapa de detección. Una vez se detecte la señal como disparo, también aparecerá un panel de resultados a la izquierda de la pantalla de la etapa de localización.

Los parámetros con los que se debe iniciar el programa son: el tamaño de ventana, la relación señal a ruido que tomará la señal elegida. Las señales a usar forman parte de una amplia base de datos, para saber trabajar con ella se debe tener en cuenta que se cuenta con una base de datos únicamente para detección y otra base de datos para localización, que se debe usar primeramente en la etapa de detección para analizar la presencia de disparo. En la Figura 66 se muestra las opciones viables de relación señal a ruido, estas opciones son las elegidas a lo largo de toda esta memoria, por lo que se cuenta con librerías para cada una de estas SNR a considerar.

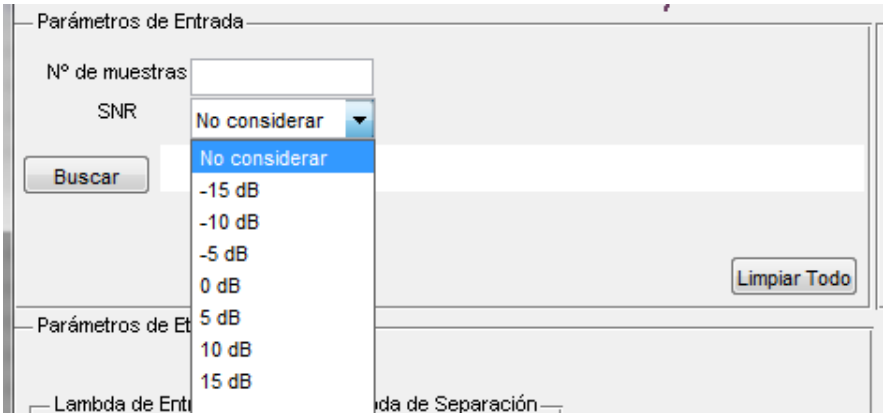


Figura 66. Selección de la relación señal a ruido como parámetro de entrada.

Se aprecia que existe un campo de 'No considerar', este campo debe ser seleccionado cuando se desea que se cargue una señal que solo pueden ser evaluadas para la etapa de detección, ya que éstas están almacenadas con su correspondiente SNR en el nombre del archivo. Sin embargo, para aquellas señales ideadas para pasar por la etapa de localización, se debe seleccionar el archivo que tenga por nombre el ángulo real de localización. Si en este último caso se selecciona el campo 'No considerar', se cargarán las señales originales que se grabaron en el patio de la Escuela Politécnica de Linares y, sin embargo, si se marca un campo determinado para SNR, la aplicación se encargará de buscar los archivos correspondientes. En la Figura 67 se muestra una selección de las grabaciones cuando la fuente sonora se sitúa en 50°, previamente seleccionado que se desea una SNR= 15dB. Para ello se debe pulsar el botón Buscar, abriéndose así una ventada que permite buscar el archivo de audio deseado.

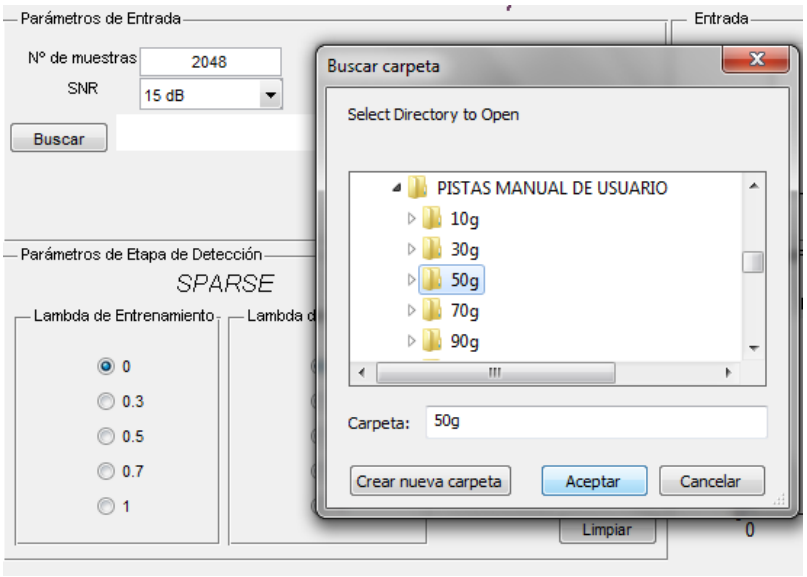


Figura 67. Selección de la carpeta que contiene los archivos de audio a analizar.

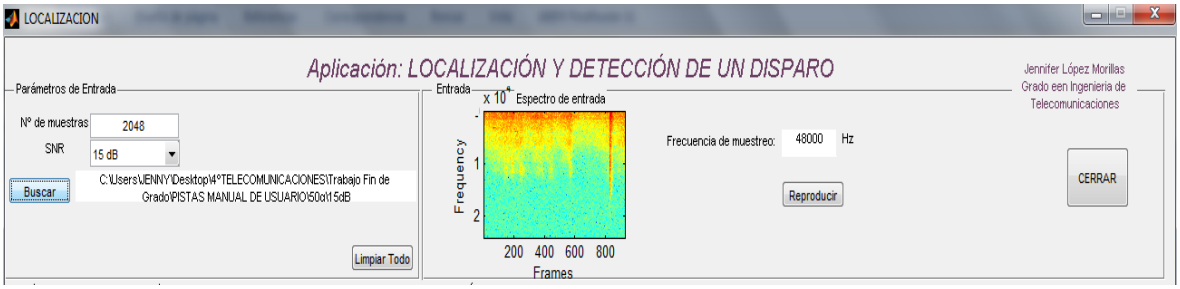


Figura 68. Vista del análisis previo de la señal de entrada.

En esta última imagen se puede ver como al cargar las señales deseadas, aparece el espectro de una de las señales de entrada correspondiente a uno de los seis canales, uno por micrófono, de los que se compone el archivo, detallando su frecuencia de muestreo y dando la opción a reproducir dicho archivo, pulsando en el botón Reproducir. Para inicializar todos los parámetros desde cero y cargar nuevas señales de audio se debe pinchar en Limpiar Todo. Igualmente si se ha evaluado las etapas de este sistema, en este botón se eliminarán los resultados permitiendo empezar una evaluación desde cero. El botón Cerrar situado a la izquierda cierra la aplicación directamente, de modo que para abrirla de nuevo se debe llamar a la aplicación desde la ventana de comandos de Matlab, pinchando en el archivo generado .fig o compilando desde el archivo .m. Si el ordenador no cuenta con plataforma Matlab, el proceso será diferente.

A continuación, una vez se han cargado desde la aplicación las señales de entrada se procede a evaluar si se detecta una señal producida por un arma de fuego. Para ello, en primer lugar se debe seleccionar los parámetros de entrada de la etapa de detección. Se

elegirá el SPARSE que se desea en cada una de las subetapas del algoritmo NMF, ya que como se ha ido viendo a lo largo del proyecto el SPARSE produce una disminución en les falsos negativos resultantes. En el ejemplo mostrado en la Figura 70 se ha marcado los parámetros óptimos de detección $\lambda=0.7$ para la etapa de entrenamiento y $\lambda=1$ para la etapa de separación. En el caso de que no se marque ninguno de los campos de SPARSE, por defecto se tendrá $\lambda=0$ en ambas etapas. Igualmente se considera el número de bases de activación del sistema. Es necesario introducir este parámetro, ya que si no se mostrará un aviso que impide realizar la detección de disparo, como se refleja en la figura 70.

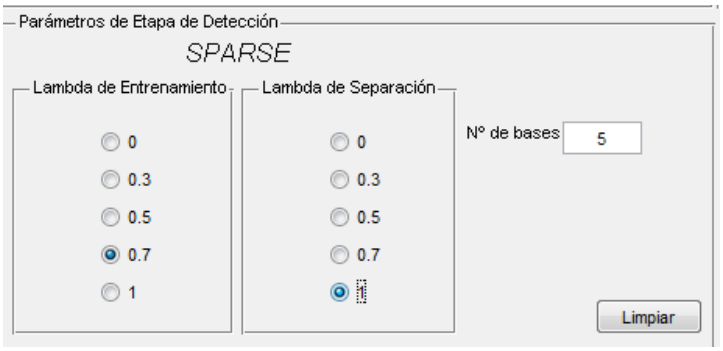


Figura 69. Parámetros de la etapa de detección.

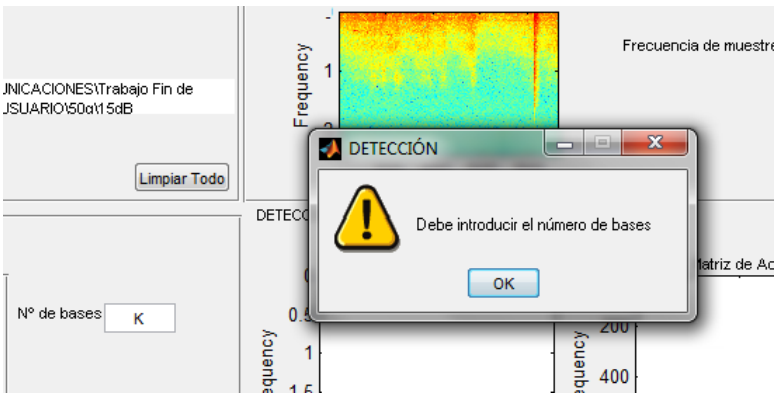


Figura 70. Mensaje de aviso para introducir debidamente el parámetro K.

En la Figura 69 se observa que existe un botón Limpiar, cuya función es borrar los cálculos de esta etapa, permitiendo seguir trabajando con la misma señal de entrada.

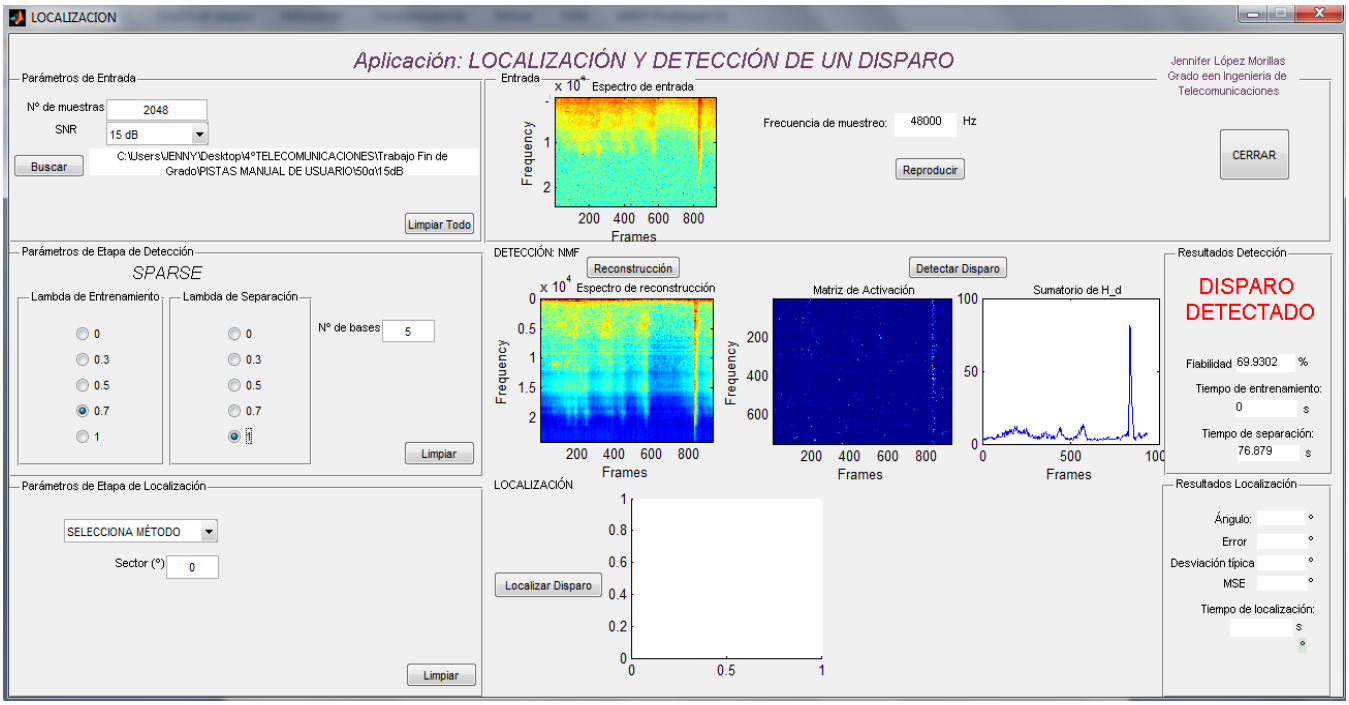


Figura 71. Obtención de resultados de la etapa de detección de disparo.

En la Figura 71 se puede ver a la izquierda el panel que indica que existe disparo, dejando visible la segunda de las etapas de este sistema, ya que cuando hay disparo se procede a la localización. Las graficas que se muestran en la parte central se generan pulsando un determinado botón. Para generar el espectro de reconstrucción de la señal del canal uno del archivo de audio de una fuente sonora situada a 50° (en este caso específico) se pincha en Reconstrucción, generándose el espectro de la señal Y=WH siguiendo la divergencia de Kullback-Leibler, que fue elegida como la que mejores resultados ofrecía y menor iteración necesitaba. Esta reconstrucción será únicamente espectral, ya que para la detección del disparo, la técnica NMF es útil para determinar si existe o no disparo, y para esto se recurre a la matriz de activaciones H_d. Para comprobar si existe disparo se picha en Detectar Disparo, construyéndose así la matriz de activaciones H_d y su sumatorio, de manera que visualmente se comprueba las muestras que el disparo activa. Si existe disparo se notifica un aviso y un porcentaje de fiabilidad de esta detección, que viene dado por la comparación del máximo de la señal a analizar y el umbral máximo calculado, tal y como se indicó en el Capítulo 4. Además aparece información del tiempo de ejecución de la etapa de entrenamiento y la de separación.

En cuanto al tiempo de ejecución de la etapa de entrenamiento, en el caso dado, es prácticamente cero. Esto se debe a que se ha implementado la aplicación con un número de muestras igual a 2048 y un número de bases K=5, que son para las que se ha diseñado el sistema final y por tanto el parámetro W_d que genera la etapa de

entrenamiento está almacenado y no tiene porque ser recalculado. Si en la aplicación se introducen otros parámetros de entrada diferentes, el tiempo de ejecución de entrenamiento será mayor que el de separación, ya que tiene mayor coste computacional.

Como se ha podido comprobar, se ha desplegado paneles para la etapa de localización, donde a la derecha se introducen los parámetros de entrada, como son el método a usar y el sector si es necesario, en la parte central se representará la localización de la fuente sonora gráficamente y a la derecha se pueden ver los resultados de interés obtenidos, como el ángulo estimado, el error absoluto cometido, la desviación típica, el error cuadrático y el tiempo de localización. En la Figura 72 se observa lo descrito.

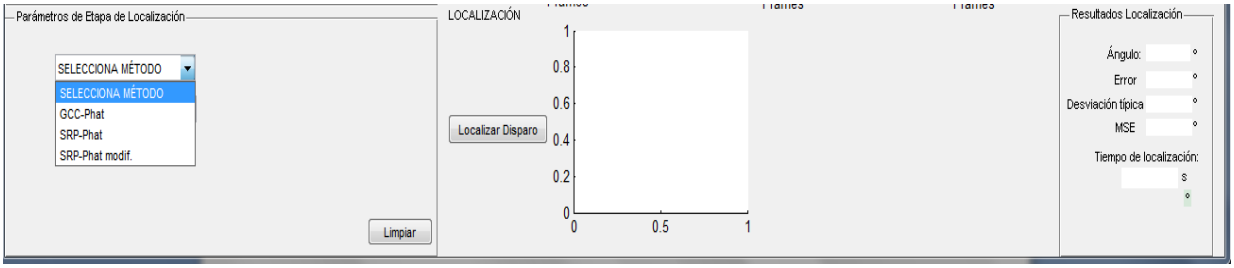


Figura 72. Aparición de paneles para la etapa de localización al detectar la señal como disparo.

Se va a seleccionar el método SRP-Phat modificado con un sector de 10°, obteniéndose los resultados que se muestran en la Figura 73.

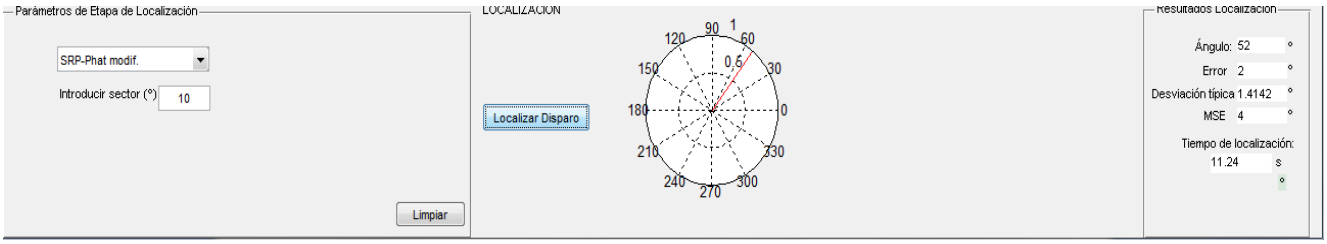


Figura 73. Resultados obtenidos en la etapa de localización con el método SRP-Phat modificado y un sector de 10°, para una fuente situada en 50°.

Seleccionando el botón limpiar para esta etapa se podrá repetir el proceso de localización para los diferentes métodos y usando diferente sectorización. En la Figura 74 se observa la pantalla completa de la aplicación usando el método SRP-Phat original para localizar.

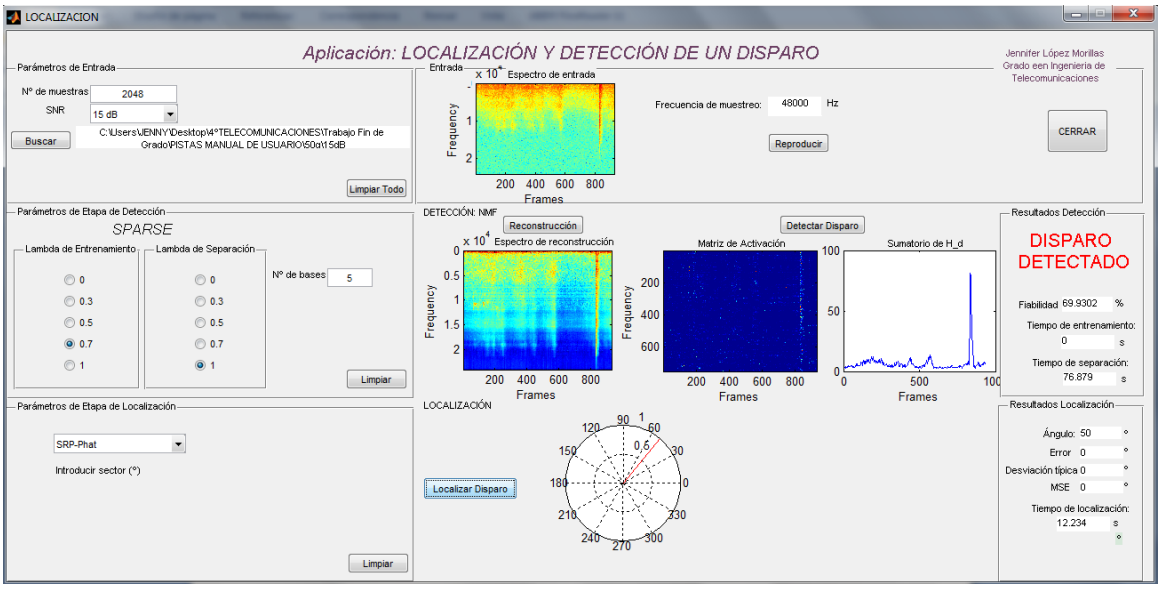


Figura 74. Detección y Localización de una señal de disparo situada a 50°.

Si se presiona el botón de Localizar Disparo sin previamente haber seleccionado un método, aparece un mensaje de aviso como se indica en la Figura 75.

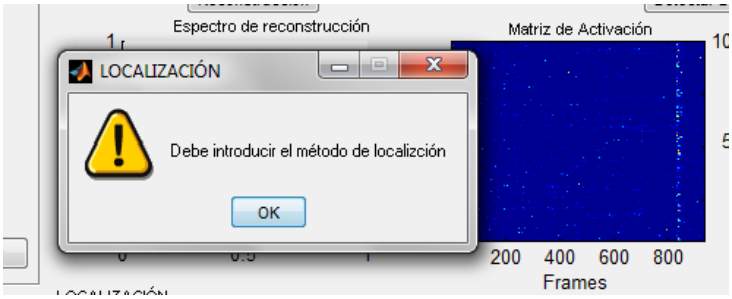


Figura 75. Mensaje de aviso por no haber seleccionado el método de localización.

El valor real del ángulo de localización es conocido debido a que dicha información es aportada por el nombre del archivo de entrada cargado, debido a que la librería de pistas de grabaciones para localización están clasificadas según la posición de la fuente. En el caso de que se introdujese un archivo sin estas características en el nombre, no se devolvería el valor del error absoluto, ya que la aplicación no tendría manera de saber cuál es el ángulo real de localización. En casos reales este error tampoco se podría conocer, por ello se debe disponer del método más preciso posible. En el ejemplo plasmado, el ángulo real de posición de la fuente sonora es de 50º, con buena relación señal a ruido (15 dB) comprobándose así que el error obtenido para el método SRP-Phat modificado es muy pequeño, solamente de dos grado. El método SRP-Phat mejora aún más la precisión de localización.

Hay que tener en cuenta que la relación señal a ruido de los archivos ha sido producida por la ampliación o atenuación de ruido que ha sido añadido a las grabaciones de disparo reales de cada canal, con lo que la señal se ve alterada y el disparo puede ser enmascarado aún más de lo que se debiera. Los mejores resultados de localización y detección se observan con las señales limpias de ruido añadido, presentando únicamente el ruido de fondo capturado por los micrófonos (que es muy leve).

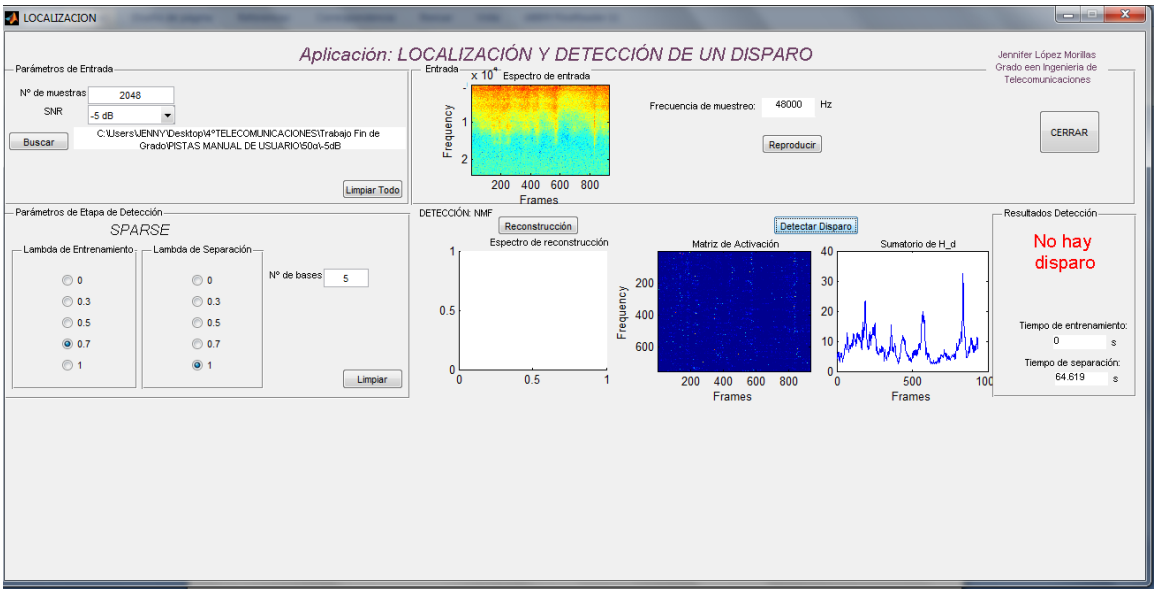


Figura 76. Disparo no detectado con una señal con SNR=-5 dB y la fuente situada en 50°.

Como se puede ver en la Figura 76 no se detecta el disparo y por tanto no se abre el panel de localización. Igualmente, si se desea insertar pistas para evaluar únicamente el disparo según su relación señal a ruido, pulsando el botón Buscar iremos al directorio deseado.