



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
MAESTRÍA EN INGENIERÍA ELÉCTRICA – PROCESAMIENTO
DIGITAL DE SEÑALES

FILTRADO DIRECCIONAL DE SEÑALES DE VOZ EN
TIEMPO REAL

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
ÁLVAREZ FERNÁNDEZ LUIS ALBERTO

TUTOR
M. I. LARRY HIPÓLITO ESCOBAR SALGUERO
FACULTAD DE INGENIERÍA

CIUDAD DE MÉXICO, SEPTIEMBRE 2017

JURADO ASIGNADO:

Presidente:	Dr. Savage Carmona Jesús
Secretario:	Dr. Rascón Estebané Caleb
1er. Vocal:	M. I. Escobar Salguero Larry
2do. Vocal:	Dr. Pérez Alcázar Pablo Roberto
3er. Vocal:	Dr. Psenicka Bohumil

Laboratorio de Procesamiento Digital de Señales, en el Posgrado de Ingeniería de la UNAM.

TUTOR DE TESIS:

M. I. LARRY HIPÓLITO ESCOBAR SALGUERO

FIRMA

Agradecimientos

A mis padres por el apoyo a lo largo de los años, por alentarme siempre a ser una mejor persona y por enseñarme a esforzarme para cumplir las metas que me proponga. A mis hermanos por estar siempre conmigo animándome y alentándome a seguir adelante.

Al M.I. Larry Escobar por el apoyo y la paciencia durante la elaboración de este trabajo, y por las enseñanzas que me ha brindado a lo largo de los años.

Al Dr. Caleb Rascón por la ayuda y los consejos que me dio tanto para el desarrollo del marco teórico como el experimental.

A mis compañeros del *Laboratorio de Procesamiento Digital de Señales*: Mickey, Nacho, Michel, Ivan y Yolo, gracias por su amistad.

Quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico, ya que sin éste me hubiera sido imposible llevar a cabo mis estudios. Así mismo, al proyecto PAPIME PE100616, *Servidor para prácticas de procesamiento digital de señales en tiempo real* por el apoyo con recursos y equipo necesarios para el desarrollo de mi trabajo.

Resumen

En este trabajo se presenta el diseño e implementación de un sistema de filtrado direccional de señales de voz, el cual es capaz de funcionar en tiempo real.

El sistema implementado consta de un Formador de Haz y una etapa de post-filtrado multicanal. Debido a que la respuesta en frecuencia del formador de haz no es constante, se empleó modelado numérico para obtener un conjunto de coeficientes óptimos y de esta manera tener un Patrón de Dirección Espacial constante en el intervalo de frecuencias de la señal de voz. Tanto el formador de haz como el postfiltro se calculan de forma adaptable para diferentes bloques de la señal, de tal manera que se emplean las características estadísticas de ésta para mejorar la respuesta del sistema.

Para la evaluación del sistema se creó un *corpus* de señales de voz, las cuales arribaban a un arreglo de micrófonos desde diferentes direcciones, estas señales se grabaron en un ambiente ruidoso y reverberante buscando aproximarse a ambientes reales. Las pruebas se realizaron en diferentes escenarios, en los cuales variaban tanto la Relación Señal a Ruido como la Relación Señal a Interferencia, obteniéndose resultados sobresalientes en la reducción de ruido e interferencias.

Contenido

1	Introducción	1
1.1	Objetivo	2
1.2	Metodología	2
1.3	Organización del Trabajo	2
2	Fundamentos de Acústica	5
2.1	Física de ondas	5
2.1.1	Ecuación de Onda	6
2.1.2	Propiedades de las Ondas	8
2.2	Acústica de Recintos	13
2.2.1	Tiempo de Reverberación	14
2.2.2	Respuesta Acústica al Impulso	14
2.3	Señal de Voz	16
2.4	Percepción del Sonido	17
3	Formadores de Haz	21
3.1	Tipos de Formadores de Haz	23
3.2	Características de los Formadores de Haz	23
3.2.1	Potencia de Salida	25
3.2.2	Patrón de Dirección Espacial	26
3.2.3	Patrón de Respuesta Direccional	27
3.2.4	Ancho del Haz	27
3.2.5	Aliasing Espacial	28
3.2.6	Ganancia del Arreglo	29
3.2.7	Ganancia de Ruido Blanco	30
3.2.8	Factor de Sensibilidad	30
3.3	Diseño de Formadores de Haz	31
3.3.1	Atenuación de Lóbulos Laterales Empleando Ventanas	31

3.3.2	Diseño Empleando Filtros FIR	32
3.3.3	Formador de Haz de Bartlett	34
3.3.4	Formador de Haz de Respuesta sin Distorsión de Mínima Varianza	35
3.3.5	MVDR-PHAT	37
3.3.6	Formador de Haz de Mínima Varianza Linealmente Restrungido	39
3.3.7	Cancelador de Lóbulos Laterales Generalizado	40
3.4	Formador de Haz de Banda Ancha	43
3.4.1	Fomador de Haz Empleando Subarreglos	45
3.4.2	Fomador de Haz de Filtro y Suma	46
3.4.3	Remuestreo Espacial	47
4	Filtrado Espacial	51
4.1	Postfiltrado	51
4.1.1	Filtro de Wiener Multicanal	53
4.1.2	Filtro de Zelinski	54
4.1.3	Filtro de McCowan	56
4.1.4	Postfiltro APES	57
4.1.5	Postfiltro APAB	58
4.2	Medidas de Desempeño	59
4.2.1	Relación Señal a Ruido	61
4.2.2	Relación Señal a Interferencia	61
4.2.3	Relación Señal a Distorsión	62
4.2.4	Relación Señal a Artefactos	62
4.3	Simulaciones	62
4.3.1	Comparación de Formadores de Haz	63
4.3.2	Comparación de Algoritmos de Postfiltrado	66
5	Diseño e Implementación del Sistema	71
5.1	Diseño	72
5.1.1	Diseño del Formador de Haz	72
5.1.2	Diseño del Postfiltro	74
5.2	Implementación	74
5.2.1	Adquisición de las Señales	74
5.2.2	Software	77

<i>CONTENIDO</i>	xi
6 Pruebas y Resultados	79
6.1 Resultados con Una Fuente	80
6.2 Resultados con Dos Fuentes	81
6.3 Resultados con Tres Fuentes	85
6.4 Evaluación del Formador de Haz Implementado	87
7 Conclusiones	91
Lista de símbolos	99
Acrónimos	101

Índice de Figuras

2.1	Reflexión especular [1].	8
2.2	Reflexión Difusa [1].	9
2.3	Ejemplo de difracción de una onda [2].	9
2.4	Ejemplo de la ley cuadrática inversa.	11
2.5	Ejemplo de múltiples fuentes emitiendo la misma señal [3]. . .	12
2.6	Patron de radiación para M=6 [3].	12
2.7	Linea de retardo para redireccionar el arreglo [3].	13
2.8	Ejemplos de respuesta al impulso para dos habitaciones con diferentes tiempos de reverberación [4].	15
2.9	Aparato humano productor de voz [5].	16
2.10	Efecto de la supresión de frecuencias en la señal de voz [6]. . .	17
2.11	Respuesta en frecuencia del oído [6].	18
3.1	Ejemplo de patrón de radiación de un formador de haz (0° a 180°).	21
3.2	Analogía entre estructura de filtro FIR y un arreglo de micrófonos lineal.	22
3.3	Ejemplo de patrones de radiación con diferentes aperturas (0° a 180°).	24
3.4	Arreglo lineal de Micrófonos.	25
3.5	Ejemplo de un patrón de dirección espacial.	26
3.6	Ejemplo de un patrón de respuesta direccional.	27
3.7	Ancho del haz.	28
3.8	Ejemplo de aliasing espacial.	29
3.9	SDP para diferentes tipos de ventanas.	32
3.10	SDP mediante diseño de filtros FIR	33
3.11	Ejemplo de desfase para suma constructiva y destructiva de señales	34

3.12	Ejemplo formador de haz de Bartlett.	35
3.13	Ejemplo formador de haz de Capon.	36
3.14	SDP de Capon para once señales.	37
3.15	Ejemplo formador de haz MVDR-PHAT.	38
3.16	Ejemplo formador de haz LCMV.	40
3.17	Diagrama de bloques del GSC.	41
3.18	Estructura del formador de haz GSC.	42
3.19	Ejemplo formador de haz GSC.	43
3.20	SDP Bartlett de banda ancha.	44
3.21	SDP Capon de banda ancha.	44
3.22	Ejemplo de un arreglo empleando subbandas.	45
3.23	Estructura general para un formador de haz de banda ancha. .	46
3.24	SDP Capon de banda ancha empleando remuestreo espacial. .	50
4.1	Postfiltrado monocanal.	52
4.2	Postfiltrado multicanal.	52
4.3	Postfiltrado de Zelinski.	56
4.4	Postfiltrado APES.	58
4.5	Postfiltrado APAB.	59
4.6	Señal de entrada al arreglo.	63
4.7	Señal de entrada al arreglo con ruido.	64
4.8	Señal de salida de los BF sin ruido.	65
4.9	Señal de salida de los BF con ruido.	66
4.10	Señal de salida de los postilftro sin ruido.	67
4.11	Señal de salida de los postilftro con ruido.	68
5.1	Diagrama general del sistema.	71
5.2	SDP MVDR de banda ancha empleando remuestreo espacial. .	73
5.3	Tarjeta 8SoundsUSB [7].	75
5.4	Movimiento angular del arreglo [8].	76
5.5	Sistema de adquisición.	76
5.6	Ejemplo de OLA de una señal con traslape de 50% [9].	77
6.1	Señal original.	79
6.2	Señal captada sin interferencia.	80
6.3	Salida del sistema sin interferencia.	81
6.4	Señal captada con interferencia en 20°.	82
6.5	Salida del sistema con interferencia en 20°.	83

6.6	Señal captada con interferencia en 35°	84
6.7	Salida del sistema con interferencia en 35°	85
6.8	Señal captada con interferencia en 20° y 35°	86
6.9	Salida del sistema con interferencias en 20° y 35°	87
6.10	Comparación de SNR y SIR para diferentes BF.	88

Capítulo 1

Introducción

Actualmente, en espacios acústicos se tiene especial interés en la separación de señales provenientes de diferentes fuentes que se encuentran altamente correlacionadas entre sí, tal como en el *efecto fiesta de cóctel*, donde existen señales de voz de distintas personas, las cuales se encuentran hablando al mismo tiempo. En este caso, el sistema de audición humana es capaz de localizar la fuente de sonido de interés y al mismo tiempo ignorar el resto de las señales, es decir, el ser humano tiene la capacidad de concentrarse en una sola conversación en medio de una sala ruidosa.

Se han buscado métodos para que los sistemas digitales puedan eliminar el ruido proveniente de direcciones diferentes a la deseada, esto con el fin de mejorar la calidad de las señales de voz para su posterior análisis o almacenamiento. Una forma de lograr esto es considerando que, si bien, las señales de voz de diferentes personas están correlacionadas entre sí de forma temporal, estas se encuentran decorrelacionadas de forma espacial [10], de tal manera que si se tienen sensores colocados en diferentes puntos del espacio, se puede obtener información temporal y espacial de las señales de interés [11].

Esto ha dado lugar a los algoritmos de filtrado espacial, como el Formador de Haz (BF), el cual está compuesto por un conjunto de señales provenientes de un arreglo de sensores. Este tipo de sistema es capaz de filtrar señales provenientes de distintas direcciones o ángulos, lo cual se logra “enfocando” el *haz* en la dirección deseada y de esta manera atenuar las señales no deseadas (interferencias).

1.1 Objetivo

Diseñar e implementar un sistema capaz de filtrar direccionalmente señales de voz, mediante un arreglo de sensores. El sistema podrá ser reconfigurado para filtrar señales provenientes de diferentes direcciones o ángulos de arribo.

1.2 Metodología

El sistema está comprendido por un arreglo de micrófonos, el cual se emplea para adquirir señales de sensores localizados en diferentes puntos del espacio, para posteriormente realizar el filtrado de la señal en una dirección específica, para esto el sistema se compone de dos bloques principales:

1. Formador de Haz de banda ancha.
2. Postfiltro multicanal.

El BF se calcula de forma adaptable para diferentes *frames* de la señal y para lograr que trabaje con señales de banda ancha, como es el caso de la señal de voz, se emplea modelado numérico. El postfiltro se encarga de disminuir el ruido incoherente presente en la señal de salida del Formador de Haz de banda ancha.

1.3 Organización del Trabajo

En el capítulo 2: *Fundamentos de Acústica*, se da una breve descripción de las características de las ondas acústicas, así como una forma de describir la respuesta acústica de un recinto; de igual manera se muestran algunas cualidades del sistema de audición humano.

En el capítulo 3: *Formadores de Haz*, se describen las características de los BF, así como algunas de las técnicas de diseño comúnmente empleadas en este tipo de algoritmos, tanto para señales monotónicas como para señales de banda ancha.

En el capítulo 4: *Filtrado Espacial*, se presentan algunos algoritmos de post-filtrado multicanal, los cuales se emplean para mejorar el nivel de reducción

de ruido incoherente de los Formadores de Haz. También se presentan las medidas empleadas para evaluar el desempeño de los algoritmos de filtrado espacial y separación de señales.

En el capítulo 5: *Diseño e Implementación del Sistema*, se emplea la teoría descrita en los capítulos anteriores para describir la implementación del sistema, así como la forma en la que se adquirieron las señales para las pruebas posteriores.

En el capítulo 6: *Pruebas y Resultados*, se describen las pruebas que se realizaron con el sistema implementado, así mismo se muestran los resultados obtenidos y una comparación con algunos de los algoritmos descritos en el capítulo 3.

Finalmente, en el capítulo 7: se presentan las *conclusiones* obtenidas con base en los resultados obtenidos en el capítulo 6.

Capítulo 2

Fundamentos de Acústica

En este capítulo se mencionan fenómenos relacionados con las señales acústicas, entre las que se encuentran: reflexión, difracción, interferencia, etc. Así mismo, se definen parte de los parámetros empleados para describir la acústica de un recinto. Posteriormente, se muestran algunas de las características del oído humano, así como de la voz, con el fin de describir las características de la señal con la que se trabajará en capítulos posteriores.

2.1 Física de ondas

El sonido es una onda mecánica la cual se propaga a través de un medio compresible y es el resultado de una perturbación variante en el tiempo, de las variables dinámicas y termodinámicas que describen al medio [12]. La transmisión de las ondas se basa en la existencia de partículas, las cuales pueden desplazarse de su posición de equilibrio; si dicho desplazamiento es perpendicular a la dirección de propagación de la onda se dice que es una onda *transversal*, por otro lado, si el desplazamiento es paralelo a la dirección de propagación se dice que la onda es *longitudinal*.

Las ondas acústicas generalmente se describen por las variaciones de presión que existen en el medio. Mientras la onda atraviesa el medio, esta causa variaciones en la presión las cuales hacen que la distancia entre las moléculas o átomos del medio varíe de forma periódica; la presión total está dada por la ecuación:

$$P(x, t) = P_0(x, t) + P_1(x, t) \quad (2.1)$$

donde P_0 representa la presión ambiental (en equilibrio) del medio y P_1 representa la fluctuación en la presión causada por la onda acústica.

2.1.1 Ecuación de Onda

La transmisión del sonido puede ser descrita mediante la *ecuación de onda* (2.2), la cual modela la propagación de ondas a través del espacio y a lo largo del tiempo [13].

$$\nabla^2 P - \frac{1}{c^2} \frac{\partial^2 P}{\partial t^2} = 0 \quad (2.2)$$

La forma más simple de solución para la ecuación de onda se da cuando se considera que la señal se propaga únicamente en una dirección, de esta manera, la ecuación (2.2) se puede expresar de forma unidimensional como:

$$\frac{\partial^2 P}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 P}{\partial t^2} = 0 \quad (2.3)$$

donde x es el eje de propagación de la onda. La solución general para la ecuación (2.3), conocida como *solución de d'Alembert*¹, esta dada por la ecuación (2.4) [14].

$$P(x, t) = F(\psi) + G(\eta) = F(x - ct) + G(x + ct) \quad (2.4)$$

Es decir, la solución de la ecuación de onda unidimensional es la suma de una función F que se transmite en la dirección positiva del eje de propagación y una función G que se propaga en la dirección negativa. Una solución particular a la ecuación de onda es la *onda plana monocromática*, representada por la ecuación (2.5) [4].

$$P(\vec{x}, t) = P_0 e^{-j(\omega t - \vec{k}^T \vec{x})} \quad (2.5)$$

donde P_0 es la amplitud de la onda, ω es la Frecuencia Angular y $\vec{k}$ es el Vector de onda. A la magnitud del vector de onda se le conoce como Número de onda ($|\vec{k}| = k = 2\pi/\lambda$), y representa el número de ciclos de la onda en la dirección de propagación, se puede considerar como la “frecuencia

¹Sin considerar condiciones iniciales.

espacial” de la onda. En la tabla 2.1 se muestran algunas de las analogías entre el número de onda y la frecuencia temporal para señales espaciales y temporales.

Señal Temporal	Señal Espacial
$t \xleftrightarrow{FT1D} \omega$	$\vec{x} \xleftrightarrow{FT3D} \vec{k}$
$\omega = 2\pi/T$	$k = \vec{k} = 2\pi/\lambda$
Filtrado temporal	Filtrado espacial
Respuesta en frecuencia	Patrón de radiación
$T \leq T_s/2$	$d \leq \lambda/2$

Tabla 2.1: Analogía entre señales temporales y espaciales

La ecuación (2.5) considera que la señal tiene un frente de onda plano, sin embargo, esta condición no siempre se cumple, ya que el frente de onda de una señal acústica puede ser considerado esférico si la fuente es de tamaño inferior a la longitud de onda y si el punto de observación se encuentra cercano a la fuente, por lo que una segunda solución a la ecuación de onda es la *onda esférica monocromática* (2.6).

$$P(r, t) = \frac{P_0}{r} e^{-j(\omega t - kr)} \quad (2.6)$$

donde r es la distancia entre la fuente y el punto de observación de la onda. Los puntos de amplitud constante en la ecuación anterior se encuentran en esferas concéntricas. Una señal acústica puede ser modelada por las ecuaciones (2.5) y (2.6), dependiendo de la distancia a la que se encuentren la fuente y el receptor. El área cercana a la fuente se conoce como región de *campo cercano* o de Fresnel. Si el punto de observación se encuentra alejado de la fuente, entonces se puede considerar que el frente de onda de la señal es plano y a esta región se le conoce como *campo lejano* o región de Fraunhofer [12]. En el caso de un arreglo de sensores, una fuente puede considerarse en el campo lejano si cumple con [15], [16]:

$$r > \frac{2D^2}{\lambda} \quad (2.7)$$

donde λ es la longitud de onda y D es la apertura del arreglo.

2.1.2 Propiedades de las Ondas

Existen ciertos fenómenos asociados a las ondas, cuando éstas se encuentran en los límites del medio de propagación, entre ellos se encuentran la reflexión, la refracción, la difracción, etc., a continuación se describen brevemente dichos fenómenos.

Reflexión

La reflexión se da cuando una onda alcanza la frontera entre dos medios diferentes, en cuyo caso, una parte de la onda es reflejada, mientras que otra se transmitirá a través del nuevo medio². Las características de la onda reflejada dependerán de la frecuencia y ángulo de arribo de la onda original, así como de las propiedades del medio. Existen dos tipos de reflexiones, *especulares* y *difusas*:

- Reflexiones especulares: este tipo de reflexión se da, comúnmente, en superficies rígidas y lisas, y se caracterizan por el hecho de que la onda reflejada tiene las mismas características que la onda incidente, excepto que la onda reflejada se origina en un punto ficticio localizado detrás de la superficie, y cada parte de la onda incidente “rebota” en una dirección específica. En la figura 2.1 se muestra un ejemplo de este tipo de reflexión.

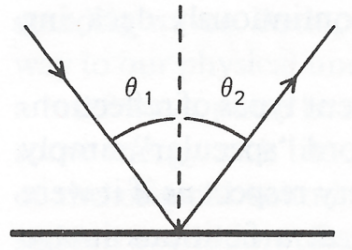


Figura 2.1: Reflexión especular [1].

- Reflexiones difusas: cuando la superficie tiene imperfecciones, estas pueden³ causar perturbaciones y ocasionar la propagación de ondas en múltiples direcciones, como se muestra en la figura 2.2.

²El nuevo medio podría incluso absorber parte de la onda.

³Dependiendo si las irregularidades son pequeñas en comparación con la longitud de onda.

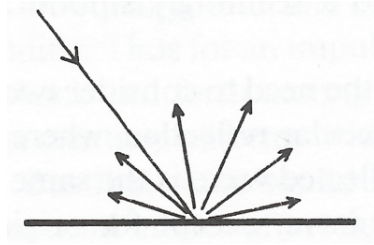


Figura 2.2: Reflexión Difusa [1].

Difracción

La difracción es la desviación de la onda alrededor de un borde o apertura. Esta desviación depende del tamaño relativo de la apertura con respecto a la longitud de la onda, así mismo, este fenómeno depende del frente de la onda; cuando la difracción se da en ondas planas, se conoce como *difracción de Fresnel*, mientras que si se da en ondas esféricas se conoce como *difracción de Fraunhofer* [12]. En la figura 2.3 se muestra un ejemplo de este fenómeno.

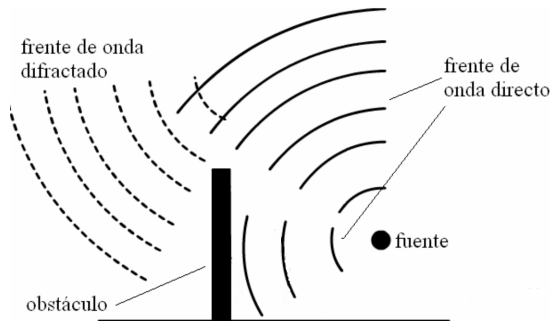


Figura 2.3: Ejemplo de difracción de una onda [2].

Refracción

La refracción es el cambio en la dirección de propagación de la onda cuando esta pasa de un medio a otro, esto se debe al cambio en la velocidad de propagación de la onda en los diferentes medios. Este fenómeno se describe empleando la ley de Snell (2.8).

$$\frac{\sin(\theta_1)}{c_1} = \frac{\sin(\theta_2)}{c_2} \quad (2.8)$$

donde θ_1 y c_1 son la dirección de propagación y velocidad de la onda en el primer medio, y θ_2 y c_2 son la dirección de propagación y velocidad de la onda en el segundo medio.

Atenuación

Los frentes de onda con frecuencia de propagación constante tienen amplitudes que decrecen exponencialmente a medida que aumenta la distancia de propagación [12]:

$$A = A_0 e^{-\alpha x} e^{-j(\omega t - kx)} \quad (2.9)$$

donde α es el coeficiente de absorción del medio, el cual es una propiedad intrínseca del material, y se encuentra asociado con la transferencia de energía de la onda hacia la energía interna del medio. Sin embargo, una forma de medir la atenuación de una onda es mediante la **ley cuadrática inversa**, la cual considera una onda esférica, cuya Intensidad Acústica (I_A) se define como:

$$I_A = \frac{W}{4\pi r^2} \quad (2.10)$$

donde W es la potencia de la onda, y r es la distancia entre la fuente y el frente de onda. Si se considera un medio uniforme, sin reflexiones ni reverberación, entonces, se tendrá que para dos distancias diferentes (r_1 y r_2), se tendrán dos segmentos esféricos los cuales tendrán la misma potencia de sonido, sin embargo, la I_A disminuirá para secciones con la misma área, como se muestra en la figura 2.4. Para obtener la relación entre dichas intensidades se emplea la ecuación (2.11) [2].

$$\frac{I_{A1}}{I_{A2}} = \frac{4\pi r_1^2}{4\pi r_2^2} = \frac{r_2^2}{r_1^2} \quad (2.11)$$

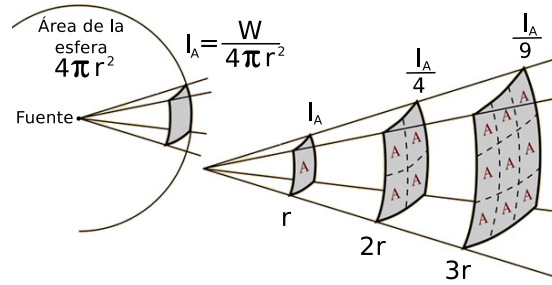


Figura 2.4: Ejemplo de la ley cuadrática inversa.

Interferencia

La interferencia es el fenómeno que se da cuando dos o más señales coherentes⁴, se propagan a través del mismo espacio. La amplitud de la combinación de estas señales cambiará de acuerdo al punto en el espacio donde se encuentre el receptor, esto se debe al tiempo que le toma a cada frente de onda para llegar al receptor. Para el caso particular de dos señales, con amplitudes A_1 y A_2 , la amplitud de la onda resultante podrá tomar cualquier valor entre $A_{min} = (A_1 - A_2)^2$ y $A_{max} = (A_1 + A_2)^2$, dependiendo de la fase de cada una de las ondas en el punto de medición. Cuando la diferencia entre las fases de las dos ondas es de 180° , se considera que las señales se encuentran *fuera de fase*, por lo que la amplitud total será A_{min} , y se dice que existe una *interferencia destructiva*; por el contrario, si el desfase entre las señales es de 0° , se considera que las señales se encuentran en *fase*, por lo que la interferencia será *constructiva* y se tendrá como amplitud total a A_{max} [1].

Una aplicación que hace uso de este fenómeno es aquella en la que se desea irradiar una onda en una dirección específica, empleando diferentes fuentes (bocinas) y se desfasando la señal de cada una de estas de tal manera que la combinación de todas las señales emitidas se encuentren en fase para la dirección deseada, mientras que para las otras direcciones las señales se encuentren fuera de fase.

Las diversas fuentes pueden organizarse empleando diferentes geometrías. En la figura 2.5 se muestra un arreglo lineal de P fuentes, las cuales se encuentran separadas entre si una distancia d y el ángulo entre la recta normal al arreglo

⁴Las señales coherentes son aquellas que tienen una relación de fases definida y fija [1].

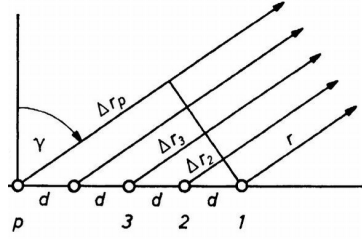


Figura 2.5: Ejemplo de múltiples fuentes emitiendo la misma señal [3].

y el punto de observación⁵ es γ . Se puede observar que cada una de las ondas debe recorrer diferentes distancias para llegar al receptor:

$$\Delta r_p = (p - 1)dsin(\gamma) \quad (2.12)$$

Por lo cual, la amplitud total de la onda, para cada ángulo γ se puede expresar como [1], [3]:

$$A(\gamma) = \left| \frac{sin(P\phi)}{Psin(\phi)} \right| \quad (2.13)$$

donde P es el número de fuentes, y $\phi = \frac{kd}{2}sin(\gamma)$. En la figura 2.6 se muestra el *patrón de radiación* para un arreglo lineal, como el mostrado en la figura 2.5, con seis elementos. Se puede observar que los máximos se localizan en la dirección $\phi = 0^\circ$ y $\phi = 180^\circ$, o lo que es lo mismo, $\gamma = 0^\circ$ y $\gamma = 180^\circ$, es decir, la amplitud máxima de la onda se dará en la dirección normal al arreglo.

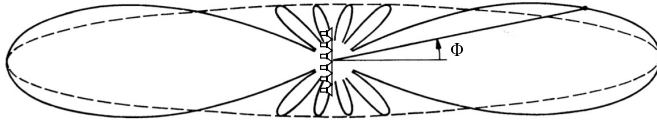


Figura 2.6: Patron de radiación para M=6 [3].

Cuando se quiere que la onda se propague en una dirección diferente a la normal, se requiere “redireccionar” el arreglo, lo cual se puede lograr, de forma electrónica, agregando líneas de retardo entre cada una de las fuentes,

⁵Considerando que el receptor se encuentra en la zona de campo lejano.

para esto se emplean arreglos LC entre cada una de las fuentes⁶, como se muestra en la figura 2.7. El retardo agregado a cada uno de los elementos se obtiene como:

$$\tau = \sqrt{LC} \quad (2.14)$$

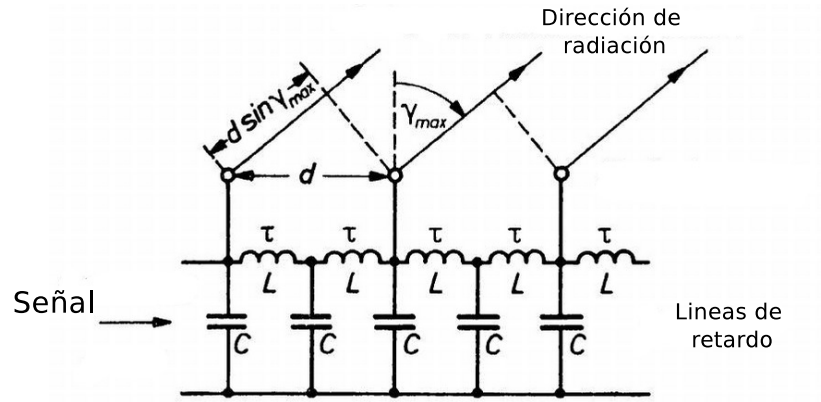


Figura 2.7: Línea de retardo para redireccionar el arreglo [3].

Si se desea redireccionar el patrón de radiación en la dirección γ_d , el retardo se calculará como [3]:

$$\tau_d = \frac{d \sin(\gamma_d)}{c} \quad (2.15)$$

Este fenómeno se emplea en el Formador de Haz (BF) de retraso y suma, el cual se explicará con mayor detalle en el capítulo 3.3.3.

2.2 Acústica de Recintos

La acústica del ambiente juega un rol importante en la forma en que se perciben las señales acústicas, ya que, en una habitación, la mayoría de las ondas de sonido que llegan a un receptor, arriban mediante la reflexión de la onda en la superficie de las paredes del recinto [12]. Debido a estas reflexiones

⁶Este filtro paso bajas se emplea para señales cuyo ancho de banda se encuentre por debajo de la frecuencia de corte del filtro.

la señal presenta atenuación y distorsión espectral, también llamada *reverberación*, la cual puede resultar agradable en las señales de voz; sin embargo, en ambientes altamente reverberativos la inteligibilidad de estas señales se puede ver reducida considerablemente [17].

2.2.1 Tiempo de Reverberación

Este parámetro puede emplearse para caracterizar de forma global a una habitación. Este concepto fue definido en el primer cuarto del siglo XX por Wallace Clément Sabine como el tiempo entre la extinción de la fuente de sonido y la completa atenuación del sonido en el recinto [18]. Actualmente se define como el tiempo necesario para que la presión de sonido decaiga a una millonésima parte de su valor original ($-60dB$) y se simboliza como T_{60} , una habitación típica puede tener un tiempo de reverberación en el orden de $200 - 500 \text{ ms}$.

Si el volumen de la habitación (V) y los coeficientes de absorción de los materiales (α_i) de cada una de las superficies son conocidos (S_i), entonces se puede aproximar el tiempo de reverberación empleando la fórmula de Eyring [17], [18]:

$$T_{60} = \frac{0.163V}{-\sum_i S_i \ln \left(1 - \frac{\sum_i S_i \alpha_i}{\sum_i S_i} \right)} \quad (2.16)$$

Otra forma de calcular el tiempo de reverberación, cuando los coeficientes de absorción son pequeños, es mediante la fórmula de Sabine [3], [17]:

$$T_{60} = \frac{0.163V}{\sum_i S_i \alpha_i} \quad (2.17)$$

2.2.2 Respuesta Acústica al Impulso

La propagación del sonido en un recinto puede ser descrita mediante la respuesta espacio-temporal al impulso [4], la cual puede ser modelada como un filtro lineal el cual describe los efectos de la reverberación entre dos puntos de una habitación. En la figura 2.8 se muestran ejemplos de la respuesta acústica al impulso para dos habitaciones con tiempos de reverberación diferentes. Esta respuesta consta de tres partes [17]:

1. Tiempo muerto: es el tiempo necesario para que la onda acústica se propague desde la fuente hasta el receptor (micrófono) a través de la trayectoria más corta, a esta onda se le conoce como *sonido directo*, y la intensidad de esta depende unicamente de la distancia entre el emisor y el receptor, considerando la ley cuadrática inversa.
2. Reflexiones tempranas: sus amplitudes y retrasos están fuertemente determinadas por la forma de la habitación y de la posición de la fuente y el receptor.
3. Reflexiones tardías: decaen exponencialmente con el tiempo.

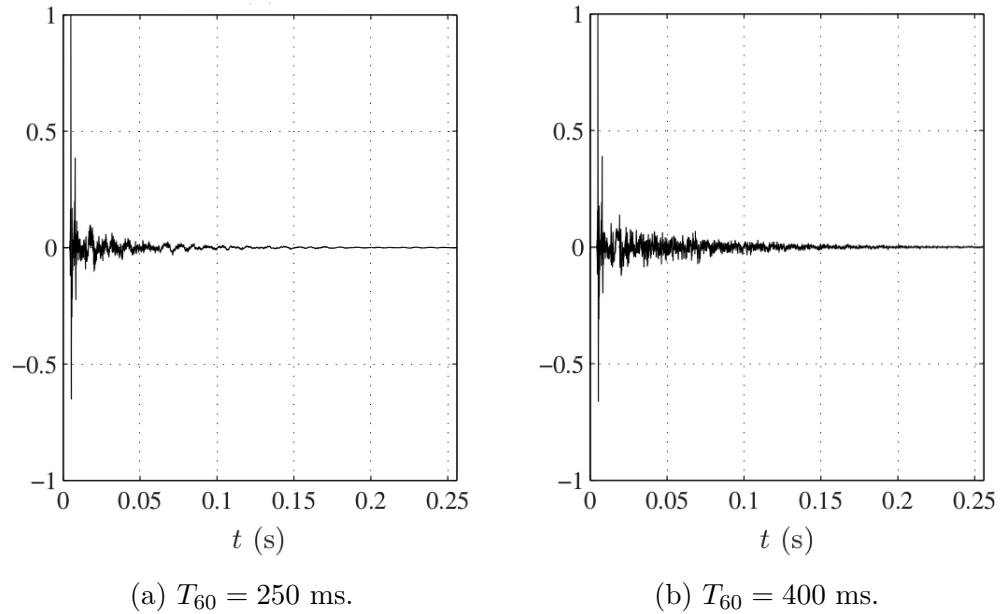


Figura 2.8: Ejemplos de respuesta al impulso para dos habitaciones con diferentes tiempos de reverberación [4].

Esta respuesta puede ser aproximada mediante el uso de filtros FIR, sin embargo, se ha demostrado que en ambientes reales, este tipo de filtros no son de fase mínima, por lo que no existe un filtro inverso causal que pueda emplearse para realizar la dereverberación de la señal [17], [19].

2.3 Señal de Voz

La señal de voz esta compuesta por una secuencia de sonidos, los cuales sirven como una representación simbólica de la información [20]. El sistema de producción de voz, figura 2.9, está constituido por los pulmones, la traquea, las cuerdas vocales, la laringe, faringe y las cavidades bucal (oral) y nasal. Los diferentes sonidos producidos se dan por el movimiento de dichos órganos, así como del movimiento de una serie de elementos articulatorios como son los dientes, los labios, la lengua, etc.

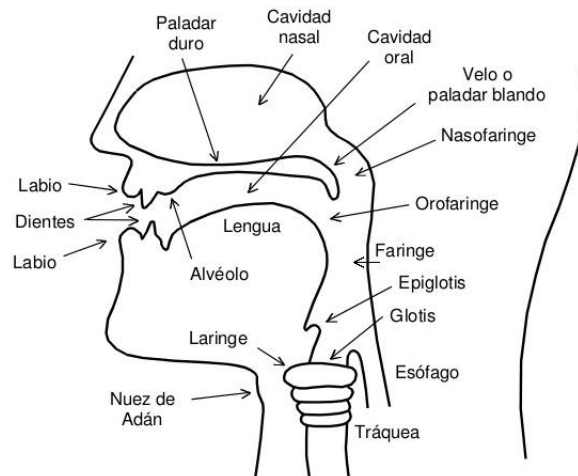


Figura 2.9: Aparato humano productor de voz [5].

La voz tiene un espectro de frecuencias que va desde los 100 Hz hasta los 6 KHz, sin embargo, la mayoría de la energía de la señal se encuentra entre los 200 Hz y los 4 KHz [20]. En la figura 2.10 se muestran las curvas de Crandall y MacKenzie, las cuales representan el porcentaje de sílabas entendidas cuando se suprimen frecuencias en la señal; la curva *L* se obtuvo manteniendo las frecuencias inferiores a la indicada en el eje de las abscisas (suprimiendo las frecuencias superiores), mientras que la curva *H* se obtiene al mantener las frecuencias superiores a la indicada (suprimiendo las frecuencias inferiores)[6].

Debido a que la voz es una señal aleatoria, para poder analizarla es necesario considerar que sus propiedades varían lentamente, es decir, se puede

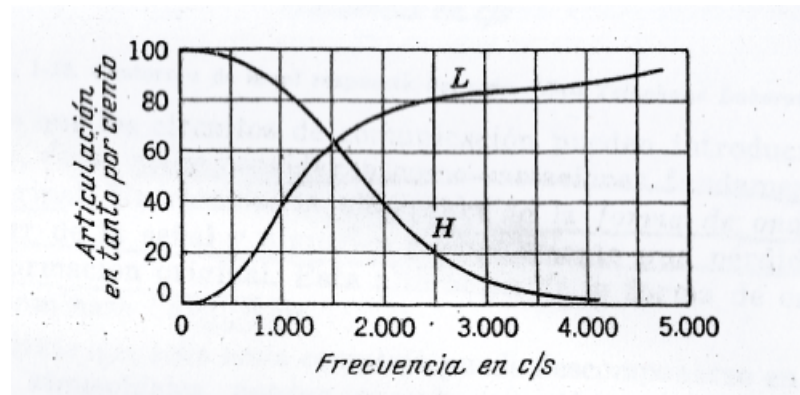


Figura 2.10: Efecto de la supresión de frecuencias en la señal de voz [6].

considerar a la señal de voz como un proceso Estacionario en Sentido Amplio (WSS) cuando se emplean segmentos de 10 a 30 ms [20]; así mismo, la señal de voz puede ser considerada como *intermitente*, es decir, existe una gran cantidad de silencio entre las palabras. En una conversación típica, mas del 50% del tiempo consiste en pausas [17].

2.4 Percepción del Sonido

El oído es parte vital del sistema de audición humana, al grado de que, generalmente, cuando se habla de sonido se hace en el contexto de las vibraciones que el oído puede captar [21]; así mismo, el sistema auditivo humano es sumamente robusto a condiciones adversas, ya que somos capaces de concentrarnos en señales de voz bajo condiciones severas de ruido y reverberación [17], esta característica se debe a la habilidad del oído para separar, internamente, las fuentes acústicas superpuestas en diferentes señales u objetos.

La Psicoacústica es la disciplina que se encarga de medir y modelar la relación entre las diferentes propiedades físicas del sonido con sus respectivos atributos subjetivos; en la tabla 2.2 se listan algunas de las propiedades físicas de las ondas de sonido y su equivalente a nivel perceptual.

El humano puede percibir sonidos con presiones desde $20\mu Pa$ hasta $200Pa$ [6], debido a esto, es común emplear escalas logarítmicas para representar el Nivel de Presión de Sonido (SPL), el cual se calcula como:

Atributo Físico	Atributo Perceptual
Intensidad	Volumen
Frecuencia fundamental	Tonalidad
Forma espectral	Timbre
Tiempo de inicio y fin	Ritmo
Fase	Ubicación

Tabla 2.2: Relación entre atributos físico y perceptuales del sonido [9].

$$SPL = 20 \log_{10} \left(\frac{P}{P_0} \right) \quad (2.18)$$

donde P_0 es la presión mínima que puede captar el oído humano, $P_0 = 20 \mu Pa$.

El intervalo de audición humano va de 20 Hz a 20 KHz, sin embargo, el oído no responde igual a todas las frecuencias, en la figura 2.11 se muestran las gráficas de Fletcher y Munson, las cuales muestran un juego de curvas de igual respuesta auditiva. Las frecuencias cuyas amplitudes se encuentren por debajo de la curva 0 no producirán sensación alguna, por lo que a esta se le conoce como *umbral auditivo* [22].

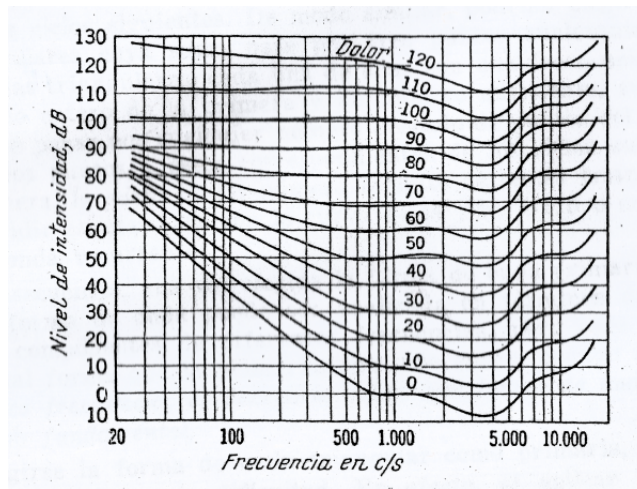


Figura 2.11: Respuesta en frecuencia del oído [6].

Resumen

En este capítulo se mostraron las propiedades básicas de las ondas acústicas, reflexión, refracción, difracción, etc., así como su forma de propagación; de igual manera se definieron los conceptos de *Tiempo de reverberación* y *Respuesta acústica al impulso*, los cuales se emplean para modelar la respuesta de un recinto a las ondas acústicas. Se mostraron algunas de las características de la señal de voz y de percepción del sonido.

Capítulo 3

Formadores de Haz

El procesamiento de señales de arreglos de sensores se ocupa de procesar señales provenientes de ondas acústicas, electromagnéticas, etc., las cuales son sensadas en diferentes puntos en el espacio [23]. De esta manera, se pueden estimar parámetros de la señal de interés al emplear información temporal y espacial [11].

El término *beamforming* o Formador de Haz (BF) deriva del hecho de que los primeros filtros espaciales se diseñaron para generar un haz con forma de lápiz, para poder recibir una señal proveniente de una ubicación específica y atenuar las señales provenientes desde otros puntos [24]; en la figura 3.1 se observa un ejemplo del patrón de radiación de un formador de haz.

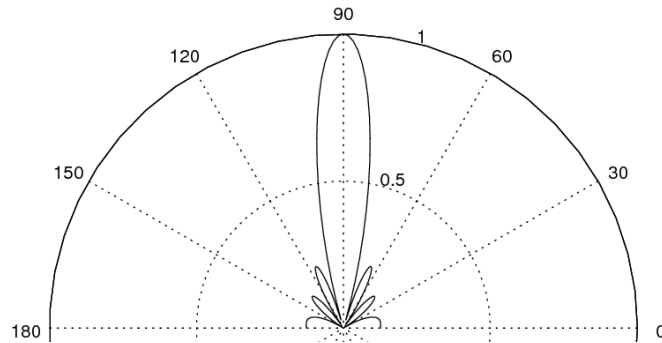


Figura 3.1: Ejemplo de patrón de radiación de un formador de haz (0° a 180°).

Los BF tiene una amplia gama de aplicaciones, entre las que se encuentran, Dirección de Arribo (DOA), reducción de ruido, dereverberación en señales, etc. En este trabajo se planea realizar el filtrado direccional de señales de voz, por lo cual, primero se debe estimar la DOA, para determinar las señales que se encuentran presentes y posteriormente filtrar “espacialmente” para obtener la señal de interés.

En este capítulo se presenta la teoría empleada del Formador de Haz, sus características, así como algunos métodos de diseño, tanto para tonos puros como para señales de banda ancha.

El formador de haz convencional (Bartlett) data de la Segunda Guerra Mundial, y es una aplicación de la Transformada de Fourier (FT) a señales espacio-temporales [11], posteriormente se buscaron métodos para mejorar la capacidad de los formadores de haz, con el fin de diferenciar entre señales cuyas fuentes se encuentren cercanas, lo que implica reducir el ancho del lóbulo principal del haz.

Los formadores de haz lineales de banda angosta pueden ser comparados con los filtros Respuesta Finita al Impulso (FIR), donde la entrada se encuentra desfasada espacialmente, en la figura 3.2 se observa la similitud entre estas estructuras.

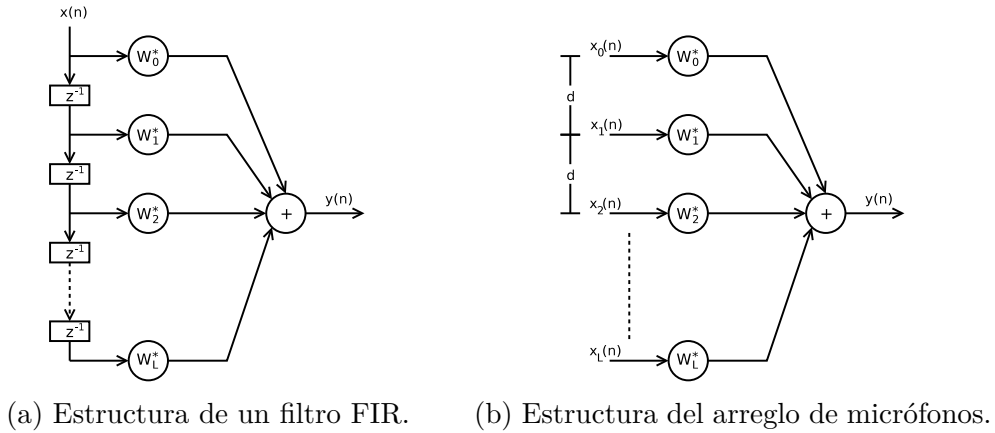


Figura 3.2: Analogía entre estructura de filtro FIR y un arreglo de micrófonos lineal.

La salida $y(n)$, en ambos casos, se expresa como una combinación lineal de la señal de entrada con desfases (temporales o espaciales) constantes; en la ecuación (3.1) se expresa esta relación, donde $\vec{x}$ es el vector de muestras de la señal de entrada, y $\vec{w}^H$ es el conjugado transpuesto del vector de pesos.

$$y = \vec{w}^H \vec{x} = \sum_{l=0}^L w_l^* x_l \quad (3.1)$$

3.1 Tipos de Formadores de Haz

Los formadores de haz pueden ser clasificados en: *independientes de los datos* y *estadísticamente óptimos*. Dicha clasificación se basa en la forma en la que se eligen los coeficientes del vector de pesos de la ecuación (3.1) [24]. Para los formadores independientes de los datos, los coeficientes no dependen de las características de la señal de entrada, y se eligen para obtener una respuesta específica para todo tipo de señales. Los formadores de haz estadísticamente óptimos se basan en determinar los coeficientes de acuerdo con las características estadísticas de la señal para “optimizar” la respuesta del arreglo de micrófonos.

Así mismo, los formadores de haz se pueden dividir de acuerdo con la geometría del arreglo, los cuales pueden estar espaciados de forma regular o aleatoria [21]:

- Arreglos lineales (unidimensionales).
- Arreglos planares (bidimensionales).
- Arreglos volumétricos (tridimensionales).

Los arreglos lineales con espaciado regular se conocen como Arreglos Lineales Uniformes (ULA) y en el resto de este capítulo, los ejemplos mostrados se realizaron empleando este tipo de arreglo.

3.2 Características de los Formadores de Haz

En general, el desempeño del formador de haz depende directamente del tamaño físico del arreglo (apertura), a mayor tamaño mejor resolución espa-

cial y viceversa, esto se observa en la figura 3.3.

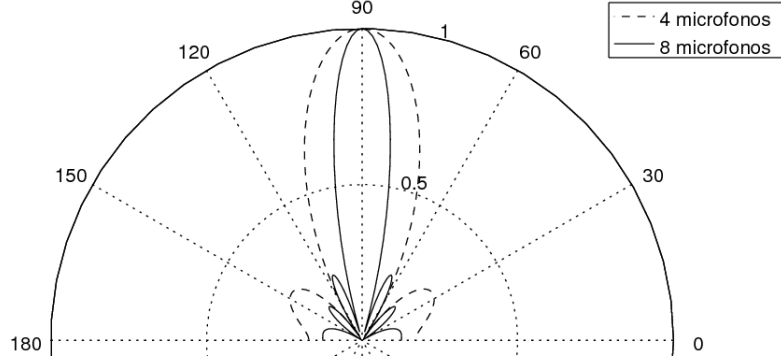


Figura 3.3: Ejemplo de patrones de radiación con diferentes aperturas (0° a 180°).

Considérese el arreglo de micrófonos mostrado en la figura 3.4, el cual está formado por M micrófonos, espaciados uniformemente una distancia d . Se tiene una señal $S(t)$, localizada a un ángulo θ respecto al centro del arreglo.

La dirección $\theta = 90^\circ$ se conoce como *broadside*, mientras que las direcciones $\theta = 0^\circ$ y $\theta = 180^\circ$ se conocen como *endfire* [13]. Si la señal $S(t)$ se encuentra suficientemente alejada del arreglo para considerarla en la región de campo lejano, las señales $x_m(t)$, para $m = 0, 1, \dots, M-1$, son versiones retrasadas de $S(t)$, es decir:

$$x_m(t) = S(t - \tau_m(\theta)) \quad (3.2)$$

donde $\tau_m(\theta)$, en el caso de una señal discreta, representa el retraso en muestras de la señal, definido como: $\tau_m(\theta) = df_s \sin(\theta)/c$ [17], con f_s la Frecuencia de muestreo. Expresando $x_m(t)$ en el dominio de la frecuencia, $X_m(\omega)$, la ecuación (3.2) se puede escribir como:

$$X_m(\omega) = S(\omega)e^{-j\omega\tau_m(\theta)} \quad (3.3)$$

Sustituyendo la ecuación (3.3) en (3.1), se obtiene:

$$Y(\omega) = \vec{W}^H(\omega)\vec{a}(\omega, \theta)S(\omega) \quad (3.4)$$

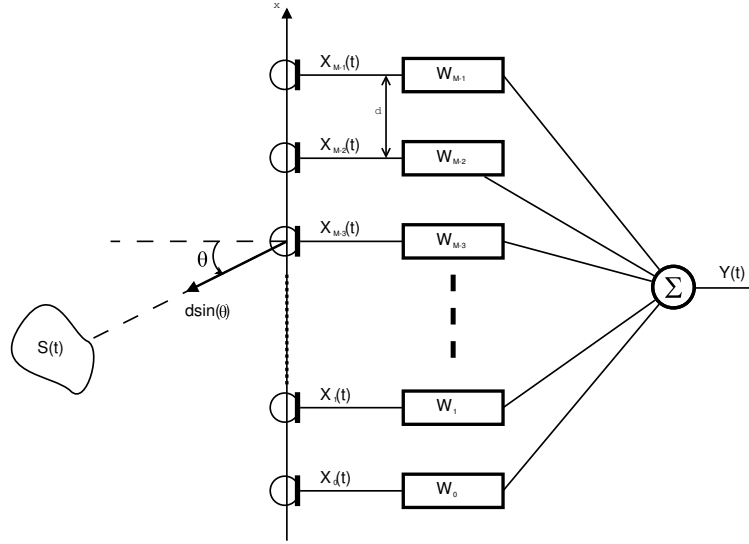


Figura 3.4: Arreglo lineal de Micrófonos.

al vector $\vec{a}(\omega)$ de la ecuación anterior se le conoce como **Vector de Dirección**, y se expresa de la siguiente manera:

$$\vec{a}^T(\omega, \theta) = \begin{bmatrix} e^{-j\omega\tau_0(\theta)} & e^{-j\omega\tau_1(\theta)} & e^{-j\omega\tau_2(\theta)} & \dots & e^{-j\omega\tau_{M-1}(\theta)} \end{bmatrix} \quad (3.5)$$

3.2.1 Potencia de Salida

La potencia de salida del haz, teniendo N muestras de la señal de salida, se calcula mediante la siguiente ecuación [11]:

$$P(w) = \frac{1}{N} \sum_{n=0}^{N-1} |y(n)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \vec{w}^H \vec{x}(n) \vec{x}(n)^H \vec{w} = \vec{w}^H \hat{\mathbf{R}}_{\mathbf{xx}} \vec{w} \quad (3.6)$$

Donde $\hat{R}$ es una estimación de la matriz de correlación de $\vec{x}(n)$, ecuación (3.7), con N siendo el número de muestras de $x(t)$.

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n)^H \quad (3.7)$$

3.2.2 Patrón de Dirección Espacial

El Patrón de Dirección Espacial (SDP), también conocido como patrón de directividad [25], de un BF con pesos $\vec{W}(\omega)$, está definido como la función de transferencia entre la salida $Y(\omega)$ y la entrada $S(\omega)$ [17], expresándose a partir de la ecuación (3.4) como:

$$H(\omega, \theta) = \frac{Y(\omega)}{S(\omega)} = \vec{W}^H(\omega) \vec{a}(\omega, \theta) \quad (3.8)$$

En la figura 3.5 se muestra un ejemplo de SDP, el cual provee una caracterización completa de la entrada-salida del sistema [25], es decir, la forma que tendrá la salida del arreglo de micrófonos cuando los coeficientes $\vec{W}(\omega)$ se encuentran “apuntando” a una dirección en particular.

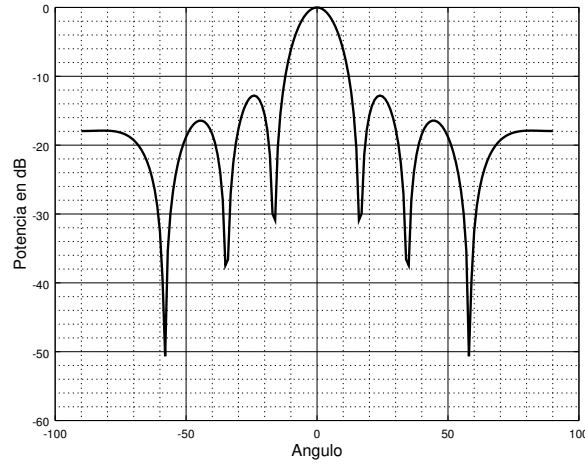


Figura 3.5: Ejemplo de un patrón de dirección espacial.

Este parámetro se puede emplear para analizar el comportamiento del arreglo para señales provenientes de direcciones diferentes a la dirección deseada, ya que se puede observar la atenuación de las señales provenientes en las direcciones del endfire.

3.2.3 Patrón de Respuesta Direccional

El Patrón de Respuesta Direccional (SRP) mide la salida del formador de haz cuando los coeficientes $\vec{W}(\omega)$ se varían sistemáticamente en función del ángulo θ en la región de interés (endfire) [25]. Esta respuesta se mide empleando la potencia de salida del BF, ecuación (3.6), con la diferencia de que $\vec{W}(\omega) = \vec{W}(\omega, \theta)$. En la figura 3.6 se muestra un ejemplo de SRP con dos señales arribando al arreglo.

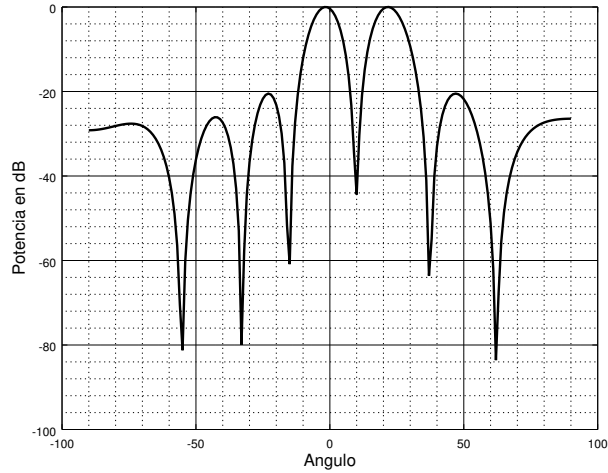


Figura 3.6: Ejemplo de un patrón de respuesta direccional.

Este patrón se puede emplear para estimar el ángulo de arribo de las señales de entrada al arreglo de micrófonos, ya que, cuando los coeficientes $\vec{W}(\omega, \theta)$ coinciden con el ángulo de arribo de la señal $S(t)$, la potencia de salida del BF tendrá una mayor magnitud.

3.2.4 Ancho del Haz

El ancho del haz de un BF se define como la región entre los ángulos donde se encuentran el primer par de *ceros* más cercanos a la dirección deseada [21], [25], figura 3.7.

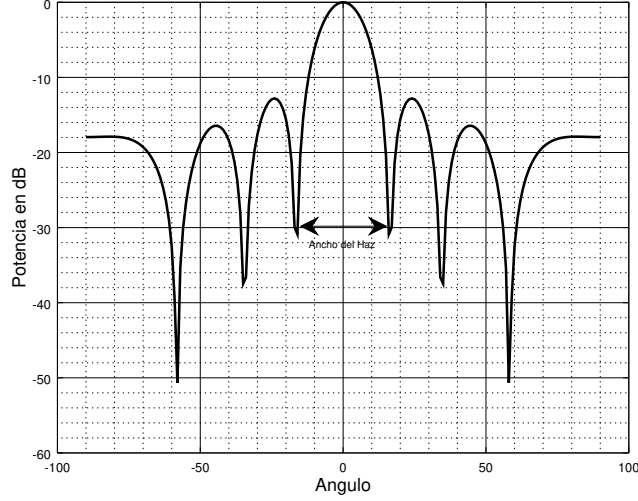


Figura 3.7: Ancho del haz.

3.2.5 Aliasing Espacial

Como se mencionó en el capítulo 2.1.1, la relación que existe entre la distancia y el número de onda es análoga a la que existe entre el tiempo y la frecuencia, por lo tanto se puede considerar que las salidas de los sensores de un ULA son el resultado de “muestrear espacialmente” una onda continua [26], por lo que, así como en las señales temporales donde la frecuencia de muestreo debe ser al menos dos veces más grande que la frecuencia máxima de la señal; en las señales espaciales, el espaciamiento debe ser al menos dos veces menor que la longitud de onda de la señal de mayor frecuencia espacial, la cual va asociada al Número de onda (k). De tal manera que para evitar el *aliasing* espacial, la separación de los elementos debe ser:

$$d_{max} \leq \frac{\pi}{|k|} \quad \text{donde} \quad \frac{\pi}{|k|} = \frac{\lambda}{2} \quad (3.9)$$

Cuando no se cumple con la condición de la ecuación (3.9), se presentan lóbulos gratinados en el SDP como se muestra en la figura 3.8.

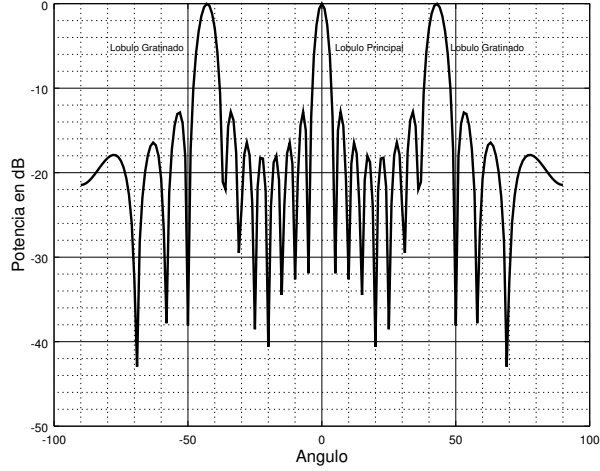


Figura 3.8: Ejemplo de aliasing espacial.

3.2.6 Ganancia del Arreglo

La Ganancia del Arreglo (AG) se define como el aumento de la Relación Señal a Ruido (SNR), debido al arreglo [27]. Es decir, la reducción de ruido en la salida del arreglo, también conocida como SNR de entrada (SNR_{in}), comparada con la señal de un solo micrófono, SNR de salida (SNR_{out}) [28]:

$$AG(\omega, \theta) = 10 \log_{10} \left(\frac{SNR_{arreglo}}{SNR_{sensor}} \right) \quad (3.10)$$

Cuando se considera un campo de ruido isotrópico, se le conoce como *directividad* [27], y puede ser expresada como:

$$AG(\omega, \theta) = 10 \log_{10} \left(\frac{|\vec{W}^H(\omega) \vec{a}(\omega, \theta)|^2}{\vec{W}^H(\omega) \mathbf{\Gamma}_{\mathbf{vv}}(\omega) \vec{W}(\omega)} \right) \quad (3.11)$$

donde $\mathbf{\Gamma}_{\mathbf{vv}}(\omega)$ es la matriz de coherencia del ruido [29], [28]. Si el formador de haz se encuentra alineado con la dirección de propagación de la señal deseada, así se considera la ausencia de imperfecciones en el arreglo y ruido espacialmente decorrelacionado, la ganancia del arreglo dependerá solo de los coeficientes del vector de pesos [30]:

$$AG(\omega) = 10 \log_{10} \left(\frac{|\sum_{m=0}^{M-1} W_m(\omega)|^2}{\sum_{m=0}^{M-1} |W_m(\omega)|^2} \right) \quad (3.12)$$

3.2.7 Ganancia de Ruido Blanco

La Ganancia de Ruido Blanco (WNG) está definida como la ganancia del arreglo en presencia de ruido espacial no correlacionado [17], es decir, es la mejora en SNR del arreglo en relación con el ruido de los sensores (micrófonos) empleados [27]. Si se considera a la matriz de coherencia del ruido blanco no correlacionado espacialmente como:

$$\Gamma_{\mathbf{v}\mathbf{v}} \Big|_{uncorr} = \mathbf{I} \quad (3.13)$$

Sustituyendo dicha matriz de coherencia en la ecuación (3.10), se obtiene:

$$WNG(\omega, \theta) = 10 \log_{10} \left(\frac{|\vec{W}^H(\omega) \vec{a}(\omega, \theta)|^2}{\vec{W}^H(\omega) \vec{W}(\omega)} \right) \quad (3.14)$$

3.2.8 Factor de Sensibilidad

El factor de sensibilidad indica como se ve afectado el SDP debido a las imperfecciones en el arreglo (posición de sensores, errores en la respuesta de los sensores, etc.), y se define como el inverso de la Ganancia de Ruido Blanco [27].

$$T_f(\omega) = \frac{\vec{W}^H(\omega) \vec{W}(\omega)}{|\vec{W}^H(\omega) \vec{a}(\omega, \theta)|^2} \quad (3.15)$$

Si se asume que los coeficientes del formador de haz se calcularon de tal manera que este se encuentra “apuntando” en dirección a la señal deseada, es decir, $\vec{W}^H(\omega) \vec{a}(\omega, \theta) = 1$, aplicando el Teorema de Parseval, se puede obtener el factor de sensibilidad del arreglo $T(k)$ ¹ como [4]:

$$T(k) = \frac{2\pi}{T_s} \vec{W}^T(k) \vec{W}(k) \quad (3.16)$$

¹En tiempo discreto.

3.3 Diseño de Formadores de Haz

Como se mencionó en la sección 3.1, existen diferentes formas de clasificar los BF, ya sea independientes de la señal de entrada o estadísticamente óptimos, en esta sección se abordan varios ejemplos de Formador de Haz, así como sus características.

Para todos los ejemplos de este capítulo se empleó una geometría lineal fija, con $M = 8$ micrófonos, separados $d = 15$ cm, con dos señales senoidales de prueba de $f = 1$ KHz, utilizando una Frecuencia de muestreo $f_s = 8$ KHz.

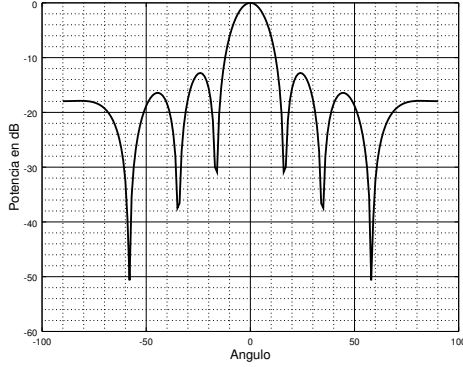
3.3.1 Atenuación de Lóbulos Laterales Empleando Ventanas

Parte del objetivo de este trabajo es minimizar los lóbulos laterales del SDP, una forma de lograrlo es mediante el uso de ventanas; para esto, se eligen los coeficientes $W(\omega)$ de la ecuación (3.1), correspondientes a los coeficientes de la ventana a emplear. En la tabla 3.1 se muestran las definiciones de algunas ventanas comúnmente empleadas y los SDP de estas ventanas se muestran en la figura 3.9.

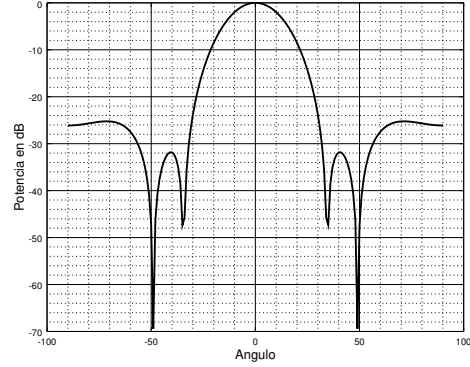
Ventana	Ecuación
Rectangular	$w_i = \begin{cases} 1 & 0 \leq i \leq M-1 \\ 0 & i \geq M \end{cases}$
Ventana Triangular	$w_i = \begin{cases} \frac{2i}{M-1} & 0 \leq i \leq \frac{M-1}{2} \\ 2 - \frac{2i}{M-1} & \frac{M-1}{2} \leq i \leq M-1 \\ 0 & i \geq M \end{cases}$
Ventana Hamming	$w_i = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi i}{M-1}) & 0 \leq i \leq M-1 \\ 0 & i \geq M \end{cases}$
Ventana Hanning	$w_i = \begin{cases} 0.5 - 0.5\cos(\frac{2\pi i}{M-1}) & 0 \leq i \leq M-1 \\ 0 & i \geq M \end{cases}$

Tabla 3.1: Ejemplos de tipos de ventanas.

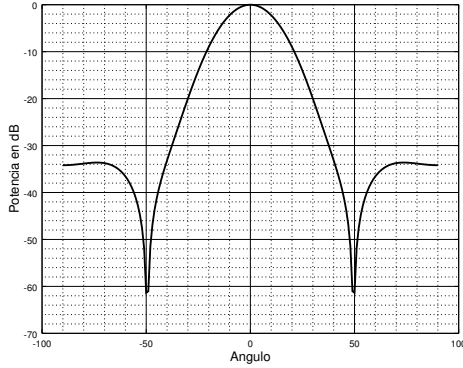
Una de las desventajas de este método es la necesidad de contar con un gran número de sensores para obtener buenos resultados.



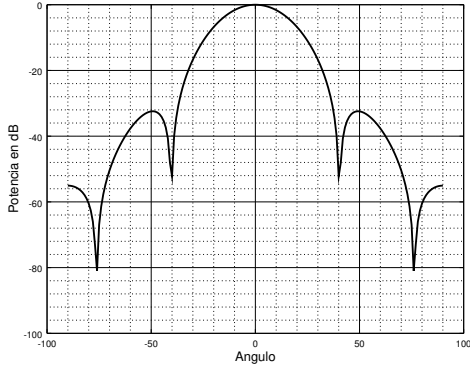
(a) Ventana rectangular



(b) Ventana triangular



(c) Ventana Hamming



(d) Ventana Hannig

Figura 3.9: SDP para diferentes tipos de ventanas.

3.3.2 Diseño Empleando Filtros FIR

Por las similitudes existentes entre los filtros FIR y los BF, las técnicas empleadas en el diseño de este tipo de filtros, pueden ser aplicadas a los formadores de haz; para esto hay que tomar en cuenta ciertas consideraciones entre la resolución espacial (θ) y la frecuencia normalizada (Ω_n). Considerando el endfire entre los ángulos -90° a 90° , y el broadside en $\theta = 0^\circ$,

es decir, perpendicular al arreglo, y considerando $|k| = k = \omega/c$, la ecuación (3.3), se puede reescribir como:

$$X_m(\omega) = S(\omega)e^{-jkdmsin(\theta)} \quad \forall m = 0 \dots M-1 \quad (3.17)$$

Cuando θ varía a lo largo de los ángulos del endfire, entonces $f_s \sin(\theta)$ varía entre $[-f_s, f_s]$, el cual es el mismo intervalo que la frecuencia normalizada del espectro de Fourier. Por ejemplo, para obtener los coeficientes de un Formador de Haz que permita pasar todas las señales provenientes entre los ángulos $\theta = -10^\circ$ y $\theta = 10^\circ$, se consideraría un filtro paso bajas con frecuencia de corte en $f_c = f_s \sin(10^\circ)$, con lo que la frecuencia normalizada quedaría $\Omega_n = [-0.17365f_s, 0.17365f_s]$, para este filtro FIR, con $M = 8$, se obtienen los siguiente coeficientes:

$$\vec{w} = \begin{bmatrix} 0.011471 & 0.052452 & 0.164987 & 0.271090 \\ 0.271090 & 0.164987 & 0.052452 & 0.011471 \end{bmatrix}$$

En la figura 3.10 se muestra el SDP del diseño anterior, la línea punteada representa el patrón deseado, mientras que la línea sólida representa el patrón obtenido.

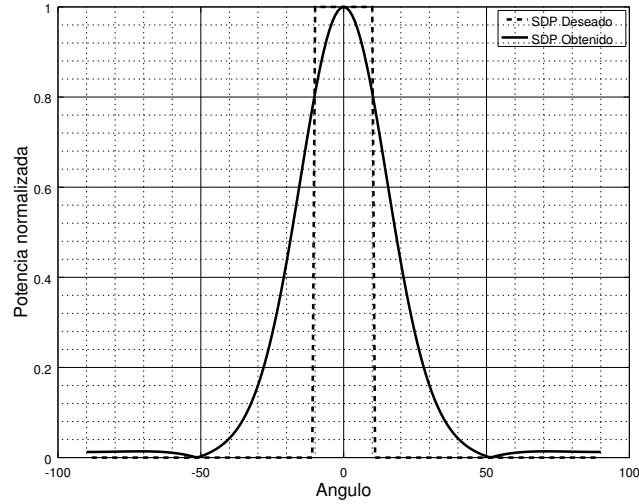


Figura 3.10: SDP mediante diseño de filtros FIR .

3.3.3 Formador de Haz de Bartlett

También conocido como *formador de haz de retraso y suma*, este algoritmo busca maximizar la potencia de salida del formador de haz para una señal de entrada específica, es decir, $\max_w \left(E\{\vec{w}^H \vec{x}(n) \vec{x}(n)^H \vec{w}\} \right)$, la solución a dicha ecuación está dada por [11]:

$$\vec{w}_B(\theta) = \frac{\vec{a}(\theta)}{\sqrt{\vec{a}^H(\theta) \vec{a}(\theta)}} \quad (3.18)$$

Sustituyendo la ecuación (3.18) en (3.6), se obtiene el SRP:

$$P_B(\theta) = \frac{\vec{a}^H(\theta) \hat{\mathbf{R}}_{\mathbf{xx}} \vec{a}(\theta)}{\vec{a}^H(\theta) \vec{a}(\theta)} = \frac{\vec{a}^H(\theta) \hat{\mathbf{R}}_{\mathbf{xx}} \vec{a}(\theta)}{M} \quad (3.19)$$

Este formador de haz busca “retrasar” las señales de los micrófonos de tal manera que al sumarlas, las señales provenientes de la dirección de interés se encuentren en fase y se realice una suma constructiva de la señal deseada, figura 3.11a, mientras que las señales provenientes de otras direcciones se atenúen debido a la suma destructiva, figura 3.11b.

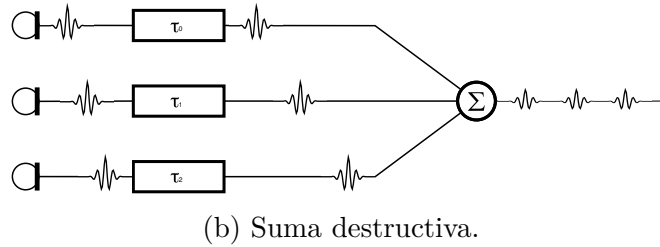
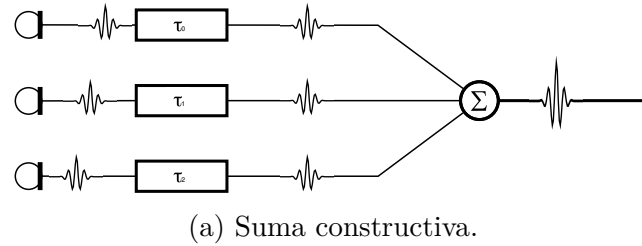


Figura 3.11: Ejemplo de desfase para suma constructiva y destructiva de señales

En la figura 3.12 se muestra un ejemplo de este formador de haz, cuando se emplean cuatro señales senoidales provenientes de direcciones $\theta_1 = -40^\circ$, $\theta_2 = -20^\circ$, $\theta_3 = 20^\circ$ y $\theta_4 = 40^\circ$. Para el SDP se seleccionó como señal de interés la proveniente de la dirección θ_2 .

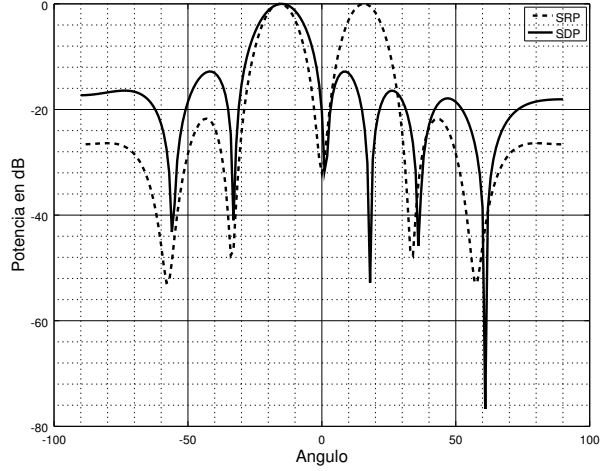


Figura 3.12: Ejemplo formador de haz de Bartlett.

3.3.4 Formador de Haz de Respuesta sin Distorsión de Mínima Varianza

El Formador de Haz de Respuesta sin Distorsión de Mínima Varianza, también conocido como *formador de haz de Capon* o *formador de haz superdirectivo*, fue propuesto en 1969 por Jack Capon como un método para el análisis de datos sísmicos [31]; este Formador de Haz minimiza la potencia de las señales provenientes de direcciones diferentes a θ , manteniendo una ganancia fija en dicha dirección, es decir:

$$\min_w (P(w)) \quad \text{sujeto a} \quad \vec{w}^H \vec{a}(\theta) = 1 \quad (3.20)$$

Los coeficientes óptimos quedan definidos como [25]:

$$\vec{w}_C(\theta) = \frac{\hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \vec{a}(\theta)}{\vec{a}^H(\theta) \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \vec{a}(\theta)} \quad (3.21)$$

El SRP de este formador de haz se obtiene sustituyendo la ecuación (3.21) en (3.6):

$$P_C(\theta) = \frac{1}{\vec{a}^H(\theta) \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \vec{a}(\theta)} \quad (3.22)$$

El SRP y el SDP empleando este algoritmo, para señales iguales a las empleadas en la subsección anterior, se muestra en la figura 3.13.

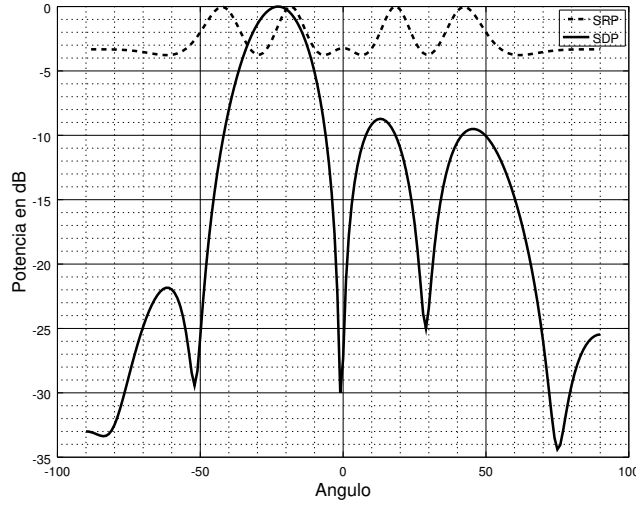


Figura 3.13: Ejemplo formador de haz de Capon.

Una de las características de este formador de haz es que, debido a que se busca minimizar la potencia de salida del arreglo sin afectar a la señal de interés, al aumentar el número de señales que inciden en el arreglo, el rendimiento de este formador de haz aumenta; este fenómeno se puede observar en la figura 3.14, donde se muestra la salida del arreglo para once señales provenientes de direcciones entre $(-50^\circ, 50^\circ)$, espaciadas $\Delta\theta = 10^\circ$.

Otra de las ventajas de este formador de haz es que, al emplear la matriz de correlación de las señales de los micrófonos, se pueden emplear algoritmos adaptables para mejorar la respuesta del arreglo conforme cambien las condiciones de ruido.

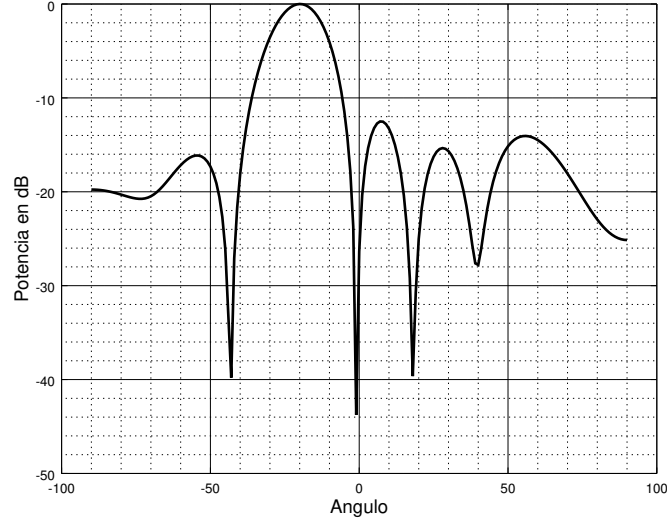


Figura 3.14: SDP de Capon para once señales.

3.3.5 MVDR-PHAT

Hoang Do y Harvey Silverman proponen en [32] una implementación diferente del algoritmo de Capon, en el cual emplean la Transformada de Fase (PHAT) [33], [34] para buscar mejorar el SDP. Esto debido a que esta transformada mejora la resolución de los algoritmos la *correlación cruzada* al calcular la Dirección de Arribo. Como se vio en la sección anterior, los coeficientes del algoritmo MVDR se calculan con base en la matriz de correlación, la cual se calcula empleando la ecuación (3.7). Si la matriz de correlación se obtiene en el dominio de la frecuencia, se obtiene la Matriz de Densidad Espectral de Potencia Cruzada (CPSD), la cual se puede expresar como:

$$\begin{aligned}
 \Phi_{\mathbf{z}}(\omega) &= E\{\vec{X}(\omega)\vec{X}^H(\omega)\} \\
 &= \begin{bmatrix} X_0(\omega)X_0^*(\omega) & X_0(\omega)X_1^*(\omega) & \dots & X_0(\omega)X_{M-1}^*(\omega) \\ X_1(\omega)X_0^*(\omega) & X_1(\omega)X_1^*(\omega) & \dots & X_1(\omega)X_{M-1}^*(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ X_{M-1}(\omega)X_0^*(\omega) & X_{M-1}(\omega)X_1^*(\omega) & \dots & X_{M-1}(\omega)X_{M-1}^*(\omega) \end{bmatrix}
 \end{aligned} \tag{3.23}$$

Aplicando la transformada de fase a la ecuación (3.23) se obtiene una *Matriz de Fase de Densidad Espectral de Potencia Cruzada*:

$$\overline{\Phi}_{\mathbf{z}}(\omega) = \begin{bmatrix} \frac{X_0(\omega)X_0^*(\omega)}{|X_0(\omega)X_0^*(\omega)|} & \cdots & \frac{X_0(\omega)X_{M-1}^*(\omega)}{|X_0(\omega)X_{M-1}^*(\omega)|} \\ \vdots & \ddots & \vdots \\ \frac{X_{M-1}(\omega)X_0^*(\omega)}{|X_{M-1}(\omega)X_0^*(\omega)|} & \cdots & \frac{X_{M-1}(\omega)X_{M-1}^*(\omega)}{|X_{M-1}(\omega)X_{M-1}^*(\omega)|} \end{bmatrix} \quad (3.24)$$

Empleando esta matriz en lugar de la matriz de correlación en la ecuación (3.21) se obtiene:

$$\vec{w}_C(\theta) = \frac{\overline{\Phi}_{\mathbf{z}}^{-1} \vec{a}(\theta)}{\vec{a}^H(\theta) \overline{\Phi}_{\mathbf{z}}^{-1} \vec{a}(\theta)} \quad (3.25)$$

Así mismo, la potencia de salida de este Formador de Haz queda definida como:

$$P_C(\theta) = \frac{1}{\vec{a}^H(\theta) \overline{\Phi}_{\mathbf{z}}^{-1} \vec{a}(\theta)} \quad (3.26)$$

En la figura 3.15 se observa el SRP y SDP empleando este BF para condiciones iguales a las empleadas en la sección 3.3.3.

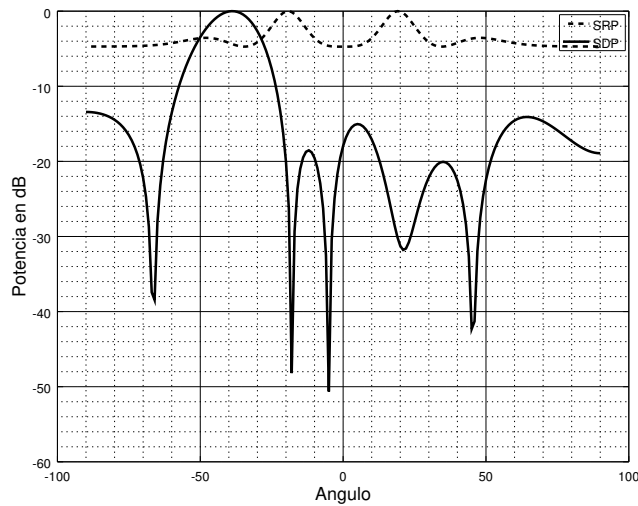


Figura 3.15: Ejemplo formador de haz MVDR-PHAT.

Se puede apreciar que este algoritmo mejora el Patrón de Dirección Espacial, sin embargo, en el Patrón de Respuesta Direccional se observa que se dificulta localizar el DOA de las señales.

3.3.6 Formador de Haz de Mínima Varianza Linealmente Restringido

El formador de haz de Mínima Varianza Linealmente Restringido, busca restringir la respuesta del arreglo de tal manera que la señal de interés tenga una ganancia y una fase específica [24], eligiendo los coeficientes que minimicen la varianza de la potencia de salida de acuerdo con dicha restricción. Se puede observar este método como una generalización del MVDR, dado que se busca:

$$\min_w (P(w)) \quad \text{suje to a} \quad \vec{w}^H \vec{a}(\theta) = G_0^* \quad (3.27)$$

Resolviendo la ecuación anterior, se obtienen los coeficientes:

$$\vec{w}_{LCMV}(\theta) = G_0^* \frac{\hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \vec{a}(\theta)}{\vec{a}^H(\theta) \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \vec{a}(\theta)} \quad (3.28)$$

Este método se puede generalizar para múltiples restricciones, lo cual se logra especificando la dirección de las señales indeseadas; por ejemplo, si la señal de interés se encuentra en la dirección θ_d y se sabe que existe una señal de interferencia en la dirección ϕ , la restricción se puede expresar como:

$$\begin{bmatrix} \vec{a}^H(\theta_d) \\ \vec{a}^H(\phi) \end{bmatrix} \vec{w} = \begin{bmatrix} G_0^* \\ 0 \end{bmatrix} \quad (3.29)$$

Si se tienen $L < M$ restricciones², estas se pueden expresar como:

$$\mathbf{C}^H \vec{w} = \vec{f} \quad (3.30)$$

donde $\mathbf{C}$ es conocida como la *matriz de restricciones*, de tamaño $M \times L$, y $\vec{f}$ siendo el *vector de respuesta*, de tamaño L . La solución de la ecuación (3.27), empleando (3.30), da como resultado los coeficientes [35]:

$$\vec{w}_{LCMV} = \frac{\hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \mathbf{C}}{\mathbf{C}^H \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \mathbf{C}} \vec{f} \quad (3.31)$$

²Se considera que las restricciones son linealmente independientes.

En la figura 3.16 se muestra un ejemplo del SDP para este formador de haz, se considerando un arreglo igual al empleado en la sección 3.3.3 y se usando el vector de respuesta:

$$\vec{f} = \begin{bmatrix} 0.01 \\ 1 \\ 0.01 \\ 0.01 \end{bmatrix} \quad (3.32)$$

Es decir, las señales provenientes de los ángulos θ_1 , θ_3 y θ_4 se minimizarán 40 dB, mientras que la señal de interés, θ_2 no se verá afectada.

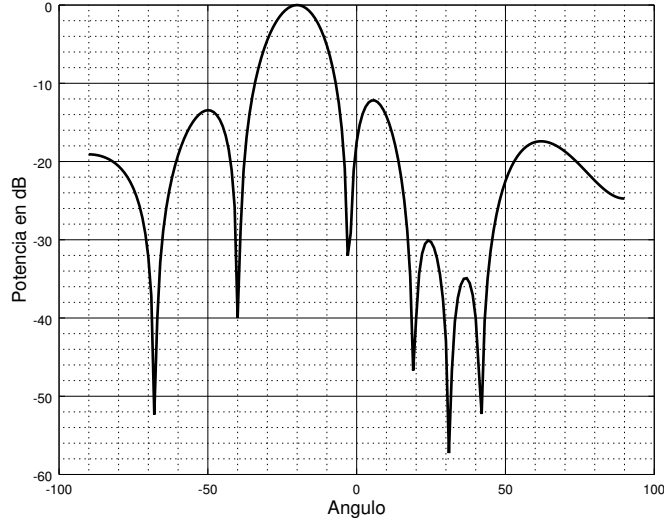


Figura 3.16: Ejemplo formador de haz LCMV.

Existen diferentes maneras de seleccionar las restricciones para este formador de haz, *point*, *derivative*, *eigenvectores*, etc., [24], [35], [36], [37].

3.3.7 Cancelador de Lóbulos Laterales Generalizado

El Cancelador de Lóbulos Laterales Generalizado (GSC), también conocido como *formador de haz de Griffiths-Jim*, puede ser representado como una formulación alternativa del LCMV, descomponiendo $\vec{w}$ en dos vectores ortogonales de tal manera que $\vec{w} = \vec{w}_o - \mathbf{C}_n \vec{w}_n$ [24], donde $\vec{w}_o$ puede verse

como un formador de haz convencional (fijo), $\vec{w}_n$ es un formador de haz independiente de $\vec{w}_o$ y sin restricciones, y a $\mathbf{C}_n$ se le conoce como *matriz de bloqueo*, de dimensiones $J \times M$ (usualmente $J = M - 1$). Esta matriz se encarga de “bloquear” la señal de interés, de tal manera que se obtiene un formador de haz donde los “canales auxiliares” están libres de la señal de interés, la cual se encontrará presente solo en el “canal principal” ($\vec{w}_o$). En la figura 3.17 se muestra el diagrama de bloques de este algoritmo.

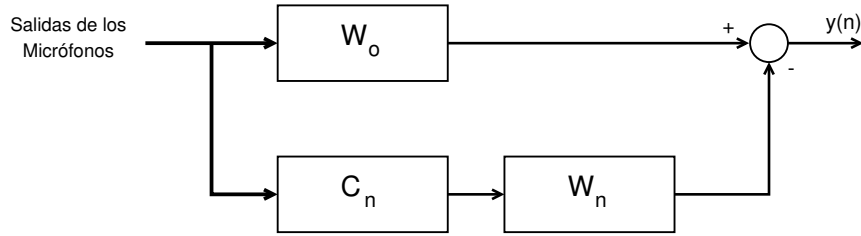


Figura 3.17: Diagrama de bloques del GSC.

En general, es común emplear algoritmos adaptables para la obtención de los coeficientes w_n , de tal manera que la salida de este formador de haz sea una estimación precisa de las señales de ruido e interferencia. En la figura 3.18 se muestra de forma desglosada el diagrama de bloques de la figura 3.17. Los coeficientes w_n se obtienen mediante la minimización del error cuadrático medio entre la salida del formador de haz fijo y la salida del formador de haz adaptable, es decir:

$$\min_{\vec{W}_n} \varepsilon \left\{ |\vec{W}_o^H \vec{X}(\omega) - \vec{W}_n^H \vec{X}|^2 \right\} \quad (3.33)$$

Resolviendo la ecuación anterior se obtiene [17]:

$$\vec{W}_n = \frac{\mathbf{C}_n \hat{\mathbf{R}}_{\mathbf{xx}}}{\mathbf{C}_n \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{C}_n^H} \vec{W}_o \quad (3.34)$$

Obteniéndose de esta manera los coeficientes $\vec{W}$ como:

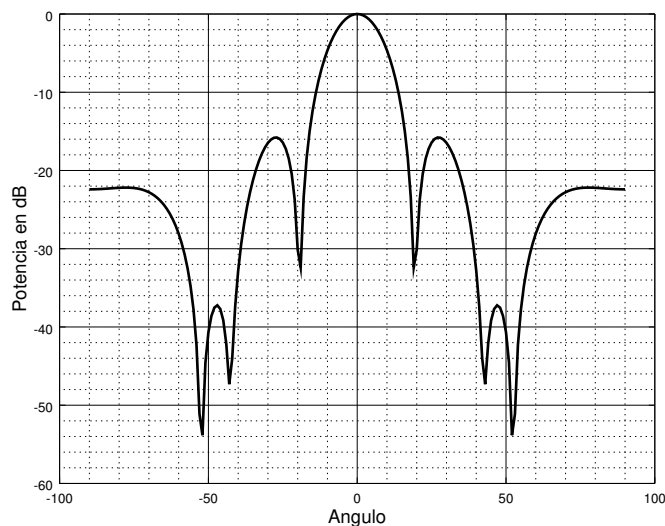


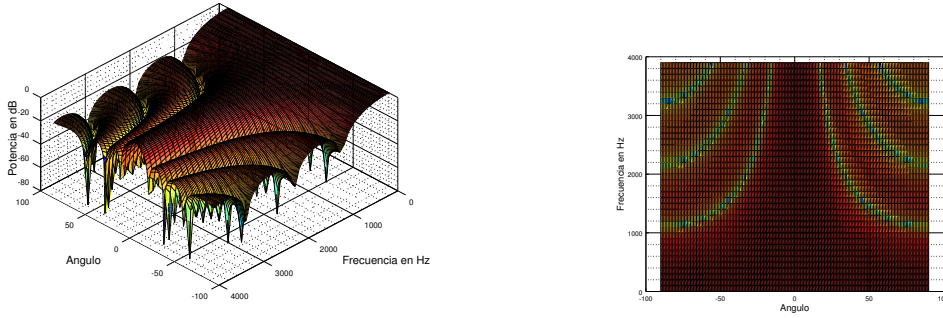
Figura 3.19: Ejemplo formador de haz GSC.

Cabe destacar que este formador de haz considera que la señal de interés arriba al arreglo de micrófonos en un ángulo igual al del broadside, por lo que un paso previo para su implementación es agregar retardos en la señal de entrada de los micrófonos para direccionar electrónicamente el arreglo en dirección de la señal de interés (véase formador de haz de Bartlett en la sección 3.3.3) [17]. Este algoritmo es sumamente útil cuando existe ruido proveniente de una dirección específica, sin embargo, su desempeño disminuye en presencia de ruido difuso no estacionario [38]. Para el correcto funcionamiento de este algoritmo es necesario que el canal auxiliar se encuentre libre de la señal de interés, de otra manera, esta señal se verá atenuada a la salida del BF; por esta razón, el desempeño de este algoritmo se ve altamente afectado por la reverberación.

3.4 Formador de Haz de Banda Ancha

Una de las características más importantes de un formador de haz es su respuesta en frecuencia. Hasta ahora se ha mostrado la respuesta del arreglo de micrófonos a señales compuestas por tonos puros, sin embargo, la respuesta del formador de haz depende directamente de la frecuencia de las señales empleadas, lo cual se puede observar en la figura 3.20, donde el ancho del

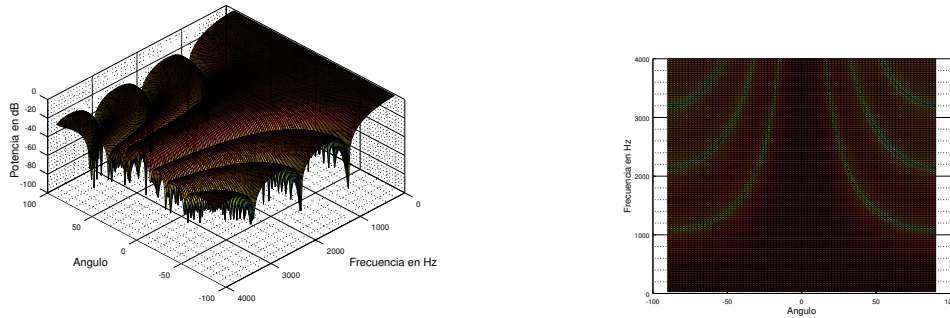
lóbulo principal aumenta conforme disminuye la frecuencia de la señal. Es por esta razón que en muchos casos se considera que la apertura del arreglo es mayor que las longitudes de onda con las que se trabaja [27], sin embargo, esto no siempre es posible.



(a) Respuesta en frecuencia (isométrico). (b) Respuesta en frecuencia (vista superior).

Figura 3.20: SDP Bartlett de banda ancha.

En la figura 3.21 se muestra la respuesta del arreglo empleando el algoritmo de Capon, el cual muestra una mejor respuesta que el de Bartlett; sin embargo, sigue teniendo un comportamiento indeseable a frecuencias bajas. Es por esta razón que, en muchos casos, surge la necesidad de diseñar un Formador de Haz Invariante en Frecuencia (FIBP), en especial cuando se trabaja con señales de voz.



(a) Respuesta en frecuencia (isométrico). (b) Respuesta en frecuencia (vista superior).

Figura 3.21: SDP Capon de banda ancha.

Existen diferentes métodos para obtener formadores de haz de banda ancha, con respuesta invariante en frecuencia, los cuales se pueden dividir en [39]:

- Calcular los coeficientes por bandas de frecuencias.
- Emplear modelado numérico.

En las siguientes secciones se mencionarán algunas de las técnicas empleadas para el diseño de Formador de Haz, las cuales pueden ser englobadas dentro de las categorías anteriores.

3.4.1 Fomador de Haz Empleando Subarreglos

Una metodología para obtener un formador de banda ancha es empleando subarreglos, donde cada subarreglo se diseña para trabajar con un conjunto reducido de frecuencias [25], [40]. Estos subarreglos pueden diseñarse de forma combinada, de tal manera que los sensores puedan emplearse en más de un subarreglo. En la figura 3.22 se muestra un ejemplo de este sistema empleando tres subarreglos.

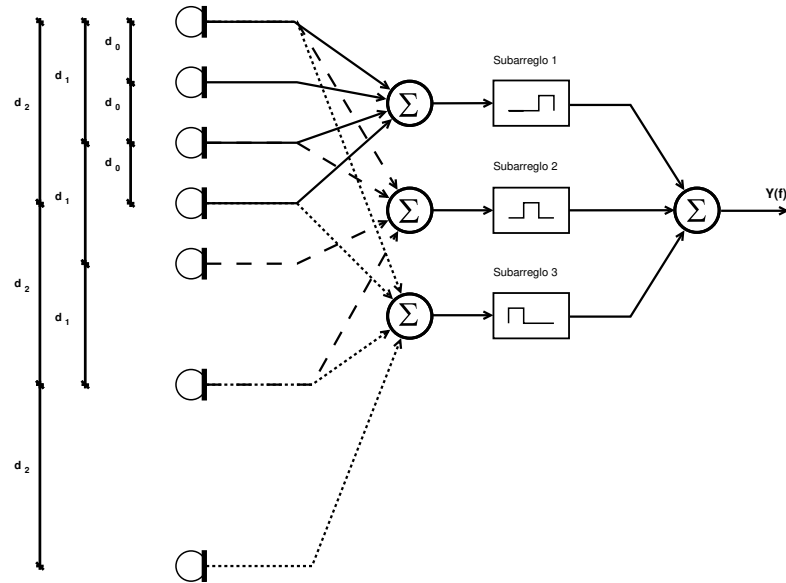


Figura 3.22: Ejemplo de un arreglo empleando subbandas.

Al aumentar la distancia entre los sensores, se pueden trabajar con frecuencias menores, ya que se busca que la relación entre la apertura del arreglo y la longitud de onda se mantenga “constante”, obteniendo así una respuesta en frecuencia invariante.

3.4.2 Formador de Haz de Filtro y Suma

Entre las estructuras más empleadas para el diseño de formadores de haz se encuentra la estructura de *Filtro y suma*, también conocido como Formador de Haz de Frost, la cual emplea filtros FIR para obtener un Patrón de Dirección Espacial predefinido [17], [35], [41]. Es decir se obtienen los coeficientes del filtro FIR a partir de un SDP de tal manera que la respuesta del arreglo, en el intervalo de frecuencias deseado, se ajuste de forma óptima a la respuesta deseada (empleando alguna función de costo predefinida) [42]. En la figura 3.23 se muestra la estructura de este tipo de filtros.

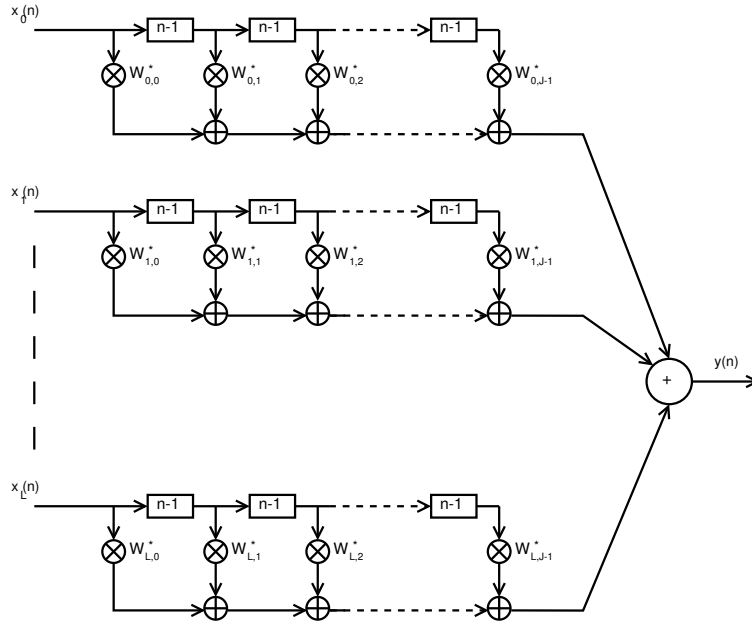


Figura 3.23: Estructura general para un formador de haz de banda ancha.

La ecuación (3.36) describe la salida de este BF, donde L es el número de micrófonos y J es el orden de cada uno de los filtros FIR.

$$y(n) = \sum_{l=0}^L \sum_{j=0}^{J-1} w_{l,j}^* x(n - jT_s) \quad (3.36)$$

Existen múltiples formas de calcular los coeficientes [27], [41], [42], [43]. Una de las desventajas de este método es la necesidad de recalcular los filtros FIR cada vez que se requiera cambiar el vector de dirección, ya que, en general, no existe una correspondencia *uno-a-uno* entre la respuesta del Patrón de Dirección Espacial original y la respuesta direccional deseada [35].

3.4.3 Remuestreo Espacial

Como se ha mostrado, la respuesta en frecuencia del BF depende en gran medida del espaciamiento que existe entre los elementos del arreglo, obteniéndose la mejor resolución para espaciamientos iguales a la mitad de la longitud de onda ($d = \lambda/2$), es por esta razón que surgen los métodos de *remuestreo espacial*, los cuales se basan en el hecho de que las señales de salida de los sensores de un Arreglo Lineal Uniforme pueden ser consideradas como muestras espaciales de una señal. Se puede ajustar esta “tasa de muestreo espacial” como una función de la frecuencia [26], es decir, $d = d(\omega)$, de tal manera que las señales de banda ancha se encuentren “alineadas” en el dominio de la frecuencia.

Esta idea se remonta a 1985 cuando Wang y Kaveh propusieron emplear una transformación lineal $T(\omega, \omega_0)$ para modificar las señales de cada sensor de una frecuencia ω a una frecuencia de referencia ω_0 [44] y de esa manera, aplicar algoritmos de alta resolución para la identificación del DOA de una señal de banda ancha. Esto se logra combinando los subespacios de las señales en diferentes frecuencias para generar un solo subespacio cuyas propiedades algebraicas sean indicativas del número de señales presentes, así como de la Dirección de Arribo de cada una de las señales. Esto es, si se tiene una matriz $\mathbf{A}_{M \times L}$ con M número de sensores y L señales, la salida del arreglo puede ser representada como:

$$\vec{Y}(\omega) = \mathbf{A}^H(\omega) \vec{X}(\omega) \quad (3.37)$$

siendo $\vec{X}^T(\omega) = [X_1(\omega) \ X_2(\omega) \ \dots \ X_M(\omega)]$, el vector de salidas de los micrófonos, entonces existe una matriz $\mathbf{T}(\omega_j)$, $j=1, 2, \dots, J$, tal que:

$$\mathbf{T}(\omega_j)\mathbf{A}(\omega_j) = \mathbf{A}(\omega_0) \quad (3.38)$$

Empleando esta transformación se pueden relacionar las diferentes frecuencias de la señal al mismo subespacio, el cual corresponde con el subespacio de la frecuencia de referencia.

Posteriormente, en 1989, Henri Clergeot y Olivier Michel propusieron que se modificara la frecuencia espacial, de tal manera que la relación d/λ se mantuviera constante [45], lo cual se logra considerando un espaciamiento de referencia d_0 , que corresponde al espaciamiento físico de los sensores, e interpolando los datos para cada frecuencia ω , de tal manera que se tengan “sensores virtuales” de espaciamiento $d = d_0\lambda/\lambda_0$, donde λ_0 es la longitud de onda de referencia.

En ese mismo año, Bienvenu y Fuerxer consideraron a $\mathbf{T}(\omega_j)$ como una matriz de “enfoque” [46], la cual transforma los vectores de dirección $\vec{a}(\omega, \theta)$, en vectores independientes de la frecuencia, ya que estos se encuentran asociados a la frecuencia de referencia ω_0 , es decir:

$$\mathbf{T}(\omega)\vec{a}(\omega, \theta) = \vec{a}(\omega_0, \theta) = \vec{a}(\theta) \quad (3.39)$$

Si se emplea la relación $d = d_0\omega_0/\omega_i$, misma que se consideró en [45], se puede expresar el vector de dirección (3.5) como:

$$\vec{a}^T(\omega_i, \theta) = \begin{bmatrix} 1 & e^{-j\omega_i \frac{\omega_0}{\omega_i} \frac{kd_0 \sin(\theta)}{c}} & \dots & e^{-j\omega_i \frac{\omega_0}{\omega_i} \frac{(L-1)kd_0 \sin(\theta)}{c}} \end{bmatrix} \quad (3.40)$$

Es decir, el vector de dirección para la frecuencia ω_i con espaciamiento d_i es el mismo que el empleado para la frecuencia ω_0 con espaciamiento d_0 , de tal manera que, empleando el vector de dirección correspondiente a cada frecuencia, se puede obtener una matriz de Densidad Espectral Cruzada (CSD) que sea una mejor estimación del eigenspacio de las señales, el cual es independiente de la frecuencia.

En 1990 Krolik propuso un método de remuestreo basado en la interpolación y decimación de las señales entre cada uno de los sensores [26], su algoritmo se detalla a continuación:

1. Insertar $(K_n - 1)$ ceros entre cada par de muestras espaciales, obteniendo $z(m, \omega_n)$.

2. Filtrar $z(m, \omega_n)$ mediante un filtro FIR paso bajas, con frecuencia de corte π/K_n , obteniéndose $z_{lp}(m, \omega_n)$.
3. Decimar $z_{lp}(m, \omega_n)$ por un factor L_n para obtener la secuencia “enfocada espacialmente” $\tilde{y}_d(m, \omega_n)$.

donde L_n y K_n son enteros, los cuales deben garantizar que $\omega_0/\omega_n = L_n/K_n$, de tal manera que remuestrear los datos de la frecuencia ω_n a la frecuencia de referencia ω_0 , corresponde a cambiar la frecuencia de muestreo espacial en un factor K_n/L_n , por lo que la distancia entre sensores, para cada una de las frecuencias queda definida como:

$$d(\omega_n) = d_0 \frac{L_n}{K_n} \quad (3.41)$$

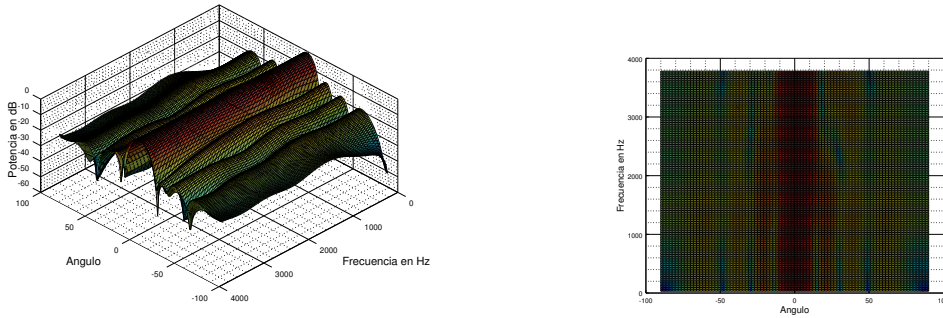
Posteriormente, en 1991 Krolik aplicó su algoritmo de remuestreo espacial empleando el formador de haz de Capon, para obtener el método Mínima Varianza Remuestreado Espacialmente (SRMV) [47], el cual consiste en utilizar el algoritmo mostrado en la sección 3.3.4, empleando las matrices de correlación remuestreadas, obteniendo buenos resultados para la estimación de la Dirección de Arribo de varias señales de banda ancha.

En el 2014, Yang [39] empleó el remuestreo espacial desde el punto de vista del diseño filtros, buscando obtener los coeficientes del Formador de Haz en sus diferentes bandas de frecuencia ($\vec{W}(\omega, \theta)$), mediante la interpolación de los coeficientes de una frecuencia de referencia, obtenidos a partir de cualquiera de los métodos descritos en la sección 3.3. La función de interpolación empleada es:

$$W_m(\omega, \theta) = \frac{d_0}{d_\omega} \sum_{i=1}^M W_i(\omega_0) \frac{\sin[\pi((m-1)d_0 - (i-1)d_\omega)/d_\omega]}{\pi((m-1)d_0 - (i-1)d_\omega)/d_\omega} \quad (3.42)$$

donde $W_i(\omega_0, \theta)$ es el i -ésimo coeficiente obtenido en la frecuencia de referencia ω_0 , d_0 es el espaciamiento de referencia, $d_\omega = \lambda/2$ es el espaciamiento óptimo entre los sensores en la frecuencia ω , $W_m(\omega, \theta)$ es el m -ésimo coeficiente interpolado en la frecuencia ω .

En la figura 3.24 se muestra el BF obtenido empleando este método para interpolar los coeficientes obtenidos mediante el BF de Capon y empleando remuestreo espacial.



(a) Respuesta en frecuencia (isométrico). (b) Respuesta en frecuencia (vista superior).

Figura 3.24: SDP Capon de banda ancha empleando remuestreo espacial.

Se observa que cumple con las características deseadas de un FIBP, ya que el ancho del lóbulo principal se mantiene constante en todas las frecuencias del espectro de la voz. Uno de los problemas al emplear este método es que los coeficientes de referencia se deben de calcular en una banda donde se **garantice** la presencia de la señal de interés.

Resumen

En este capítulo se mostraron las características de los Formadores de Haz, entre las que se encuentran el patrón de dirección espacial, el ancho del haz, la ganancia del arreglo, el factor de sensibilidad, etc. Así mismo, se mostraron algunas técnicas para la obtención de los coeficientes del vector de pesos que mejoraran las características antes mencionadas, entre los que se encuentran el MVDR, LCMV, GSC. También se presentaron algoritmos para la obtención de Formador de Haz Invariante en Frecuencia, los cuales pueden ser empleados para señales de banda ancha, como la señal de voz.

Capítulo 4

Filtrado Espacial

En la sección 3.2 se definió el SDP como la relación salida-entrada del Formador de Haz, mostrándose en la figura 3.5 un ejemplo para ángulos comprendidos dentro de los límites del endfire del arreglo. Esta relación puede ser vista como un **filtro espacial** o **direccional**, siendo el Patrón de Dirección Espacial la “magnitud de la respuesta direccional” de dicho filtro; en otras palabras, para el SDP de la figura 3.5 se tiene que las señales provenientes direcciones diferentes al broadside se atenuarán. El objetivo del filtrado direccional, en este trabajo, es atenuar las señales provenientes de direcciones diferentes a la dirección de interés, así como la minimización del ruido no correlacionado espacialmente.

En este capítulo se analizarán algunos métodos de postfiltrado, los cuales se emplean para mejorar la señal de salida del Formador de Haz, así mismo se definirán las medidas empleadas para evaluar el desempeño de los BF.

4.1 Postfiltrado

A pesar de que los Formadores de Haz mejoran la calidad de las señales de voz al emplear la información espacial de ésta, cuando la señal se encuentra en presencia de ruido espacialmente incoherente o difuso, la reducción de ruido puede ser insuficiente [38], [40], por lo que, para mejorar la calidad de la señal de salida, se emplean algoritmos de postfiltrado, los cuales pueden ser divididos en [48]:

- Monocanal: considera la salida del BF como si fuera la señal proveniente de un micrófono y emplea algoritmos para mejorar la señal de voz. En la figura 4.1 se observa el diagrama de bloques de un sistema empleando este tipo de filtro.

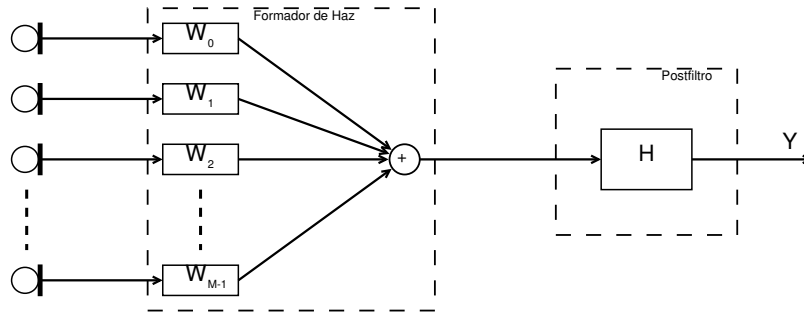


Figura 4.1: Postfiltrado monocanal.

- Multiacanal: emplea la información espacial obtenida del BF para mejorar la distinción entre la señal de voz y el ruido. La figura 4.2 muestra un diagrama de bloques empleando este filtrado.

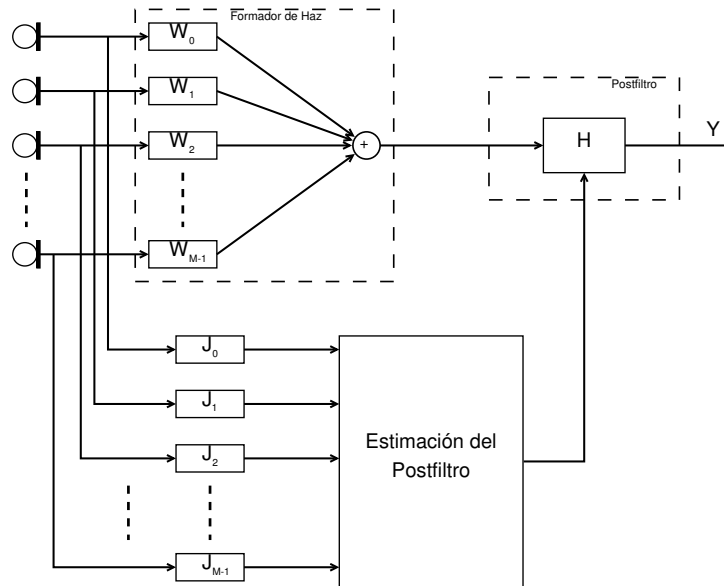


Figura 4.2: Postfiltrado multicanal.

En general, este tipo de algoritmos buscan el mejor filtro lineal en el sentido del Error Cuadrático Medio Mínimo (MMSE), para señales de banda ancha como la voz. La mejor solución para este tipo de filtros es el filtro de Wiener multicanal [28].

4.1.1 Filtro de Wiener Multicanal

Si se considera a $\vec{X}(\omega)$, como el vector de las señales de salida de los micrófonos en la subbanda de frecuencia ω y $\vec{W}(\omega)$ como el vector de coeficientes del BF en la misma banda de frecuencia, entonces $\vec{X}(\omega)$ se puede expresar como:

$$\vec{X}(\omega) = \vec{S}(\omega) + \vec{V}(\omega) \quad (4.1)$$

siendo $\vec{S}(\omega)$ el vector de la señal deseada y $\vec{V}(\omega)$ el vector de ruido aditivo. Se puede obtener un vector de coeficientes $\vec{W}_{opt}(\omega)$, tal que la salida sea la mejor aproximación en el MMSE del escalar $S(\omega)$, donde:

$$\vec{S}(\omega) = S(\omega)\vec{a}(\omega) \quad (4.2)$$

La salida del BF se puede expresar, empleando la expresión 4.1, como¹:

$$Y = \vec{W}^H \vec{X} = \vec{W}^H [\vec{S} + \vec{V}] \quad (4.3)$$

El error entre la salida del Formador de Haz y la señal deseada se obtiene como:

$$e = S - Y = S - \vec{W}^H [\vec{S} + \vec{V}] \quad (4.4)$$

De esta manera, el error cuadrático medio (ϕ_{ee}) se expresa como:

$$\begin{aligned} \phi_{ee} &= E[\{S - \vec{W}^H \vec{X}\}\{S^* - \vec{X}^H \vec{W}\}] \\ &= \phi_{ss} - \vec{W}^H \vec{r}_{xs} - \vec{r}_{xs}^H \vec{W} + \vec{W}^H \mathbf{R}_{xx} \vec{W} \end{aligned}$$

El vector de pesos óptimos se obtiene minimizando la ecuación anterior, con lo que se obtiene la ecuación multicanal de *Wiener-Hopf*:

$$\vec{W}_{opt} = \mathbf{R}_{xx}^{-1} \vec{r}_{xs} \quad (4.5)$$

¹El argumento ω se puede omitir sin pérdida de generalidad.

A esta solución se le conoce como *Filtro de Wiener Multicanal* [28]. Asumiendo que la señal y el ruido no se encuentran correlacionados, el vector de correlación $\vec{r}_{xs}$ se puede reducir a:

$$\vec{r}_{xs} = \phi_{ss}\vec{a} \quad (4.6)$$

por lo que la matriz de correlación se puede expresar como:

$$\tilde{\mathbf{R}}_{\mathbf{xx}} = \phi_{ss}\vec{a}\vec{a}^H + \mathbf{\Phi}_{\mathbf{vv}} \quad (4.7)$$

siendo $\mathbf{\Phi}_{\mathbf{vv}}$ la matriz de correlación del ruido. Empleando las ecuaciones anteriores, la ecuación (4.5) se puede reescribir como:

$$\vec{W}_{opt} = [\phi_{ss}\vec{a}\vec{a}^H + \mathbf{\Phi}_{\mathbf{vv}}]^{-1}\phi_{ss}\vec{a} \quad (4.8)$$

La ecuación anterior puede ser expresada como [28]:

$$\vec{W}_{opt} = \underbrace{\left[\frac{\phi_{ss}}{\phi_{ss} + (\vec{a}^H \mathbf{\Phi}_{\mathbf{vv}}^{-1} \vec{a})^{-1}} \right]}_{\text{H}} \underbrace{\frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1} \vec{a}}{\vec{a}^H \mathbf{\Phi}_{\mathbf{vv}}^{-1} \vec{a}}}_{\text{MVDR}} \quad (4.9)$$

De la ecuación (4.9) se puede observar que el segundo termino corresponde al Formador de Haz MVDR de la sección 3.3.4 multiplicado por un factor escalar. A esta solución se le conoce como *Arreglo Superdirectivo con Post-filtrado de Wiener*. Uno de los inconvenientes de este tipo de filtro es que se debe conocer o estimar ϕ_{ss} .

4.1.2 Filtro de Zelinski

En 1988, Zelinski [49] propuso emplear el filtro de Wiener de forma adaptable para estimar la señal de voz deseada. Esto se logra estimando el espectro de potencia del ruido de los valores que se encuentran fuera de la diagonal principal de la matriz de covarianza entre los micrófonos [50]. Este algoritmo se basa en el hecho de que, en ambientes reverberativos, la correlación entre la señal recibida en los micrófonos disminuye cuando aumenta la distancia entre los micrófonos y la fuente de la señal, por lo que se supone que la señal de interés se encuentra más próxima al arreglo que las señales de interferencia y ruido.

Al estar basado en el filtro de Wiener multicanal, se parte de la ecuación (4.5); sin embargo, el cálculo de la matriz de correlación ($\mathbf{R}_{\mathbf{xx}}$) y del vector de correlación ($\vec{r}_{xs}$) se realiza mediante la Densidad Espectral de Potencia Cruzada (CPSD). En la figura 4.3 se muestra el diagrama de bloques de este algoritmo. Donde $A(k)$ es la densidad autoespectral de los micrófonos, la cual se calcula como:

$$A(k) = \frac{1}{M} \sum_{i=0}^{M-1} |Y_i(k)|^2 \quad (4.10)$$

Dado que se pretende que el filtrado sea adaptable, la matriz de densidad espectral cruzada ($\mathbf{C}(k)$), se estima como:

$$[\mathbf{C}(k)]_{ij} = \Re\{Y_i(k)Y_j^*(k)\} \quad (4.11)$$

$$[\mathbf{R}_{xx}(k)]_{ij} = \Re\{Y_i(k)Y_j^*(k)\} \quad (4.12)$$

El postproceso mostrado en la figura se calcula mediante la suma de los elementos que se encuentran por encima de la diagonal principal de la matriz $\mathbf{C}(k)$.

$$P(k) = \alpha(k) \frac{2}{M(M-1)} \sum_{ij \in IP} [\mathbf{C}(k)]_{ij} \quad (4.13)$$

donde IP representa el conjunto de pares de índices de los elementos de la matriz $\mathbf{C}(k)$ que se encuentran por encima de la diagonal principal, y $\alpha(k)$ es un factor de reducción dependiente de la frecuencia. Posteriormente se aplica la IDFT para obtener $a(n)$ y $c(n)$, las cuales son estimaciones de las funciones de autocorrelación $\vec{r}_{xx}(\cdot)$ y $\vec{r}_{ss}(\cdot)$, respectivamente; con estas funciones es posible obtener los parámetros del filtro de Wiener empleado en el postfiltro [49].

Los coeficientes de este filtro, se pueden expresar, en el dominio de la frecuencia como [48]:

$$H_Z(\omega) = \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M \Re\{Y_i(\omega)Y_j^*(\omega)\}}{\frac{1}{M} \sum_{i=1}^M |Y_i(\omega)|^2} \quad (4.14)$$

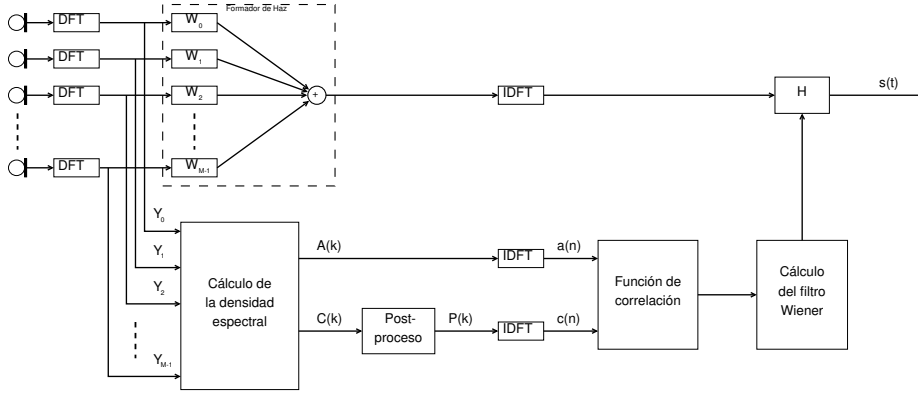


Figura 4.3: Postfiltrado de Zelinski.

Cabe destacar que este algoritmo sobreestima la densidad de potencia del ruido, por lo que la salida no es óptima en el sentido de Wiener [51], [52], [53].

4.1.3 Filtro de McCowan

Este filtro puede verse como una generalización del filtro de Zelinski, debido a que dicho filtro considera la presencia de ruido incoherente ideal, sin embargo, en muchos casos, existe ruido correlacionado en las regiones de baja frecuencia [54]; por esta razón, McCowan et al. proponen el uso de una función de ruido incoherente conocida, de tal manera que el postfiltrado sea más apropiado para diferentes condiciones de ruido [55].

Una forma común de caracterizar el ruido es mediante la *función de coherencia compleja*:

$$\Gamma_{n_i n_j} = \frac{\phi_{n_i n_j}}{\sqrt{\phi_{n_i n_i} \phi_{n_j n_j}}} \quad (4.15)$$

donde $\phi_{n_i n_j}$ es la CPSD entre el ruido de las señales n_i y n_j , y $\phi_{n_i n_i}$, $\phi_{n_j n_j}$ son la Densidad Espectral de Potencia (PSD) del ruido de la señal n_i y la señal n_j , respectivamente. Dicha función de coherencia se encuentra en el intervalo $-1 \leq \Gamma_{ij} \leq 1$. Esta medida puede ser vista como la correlación entre las señales para dos puntos. Al suponer que el ruido entre los sensores se encuentra decorrelacionado, se considera que $\mathbf{\Gamma}_{nn} = \mathbf{I}$, sin embargo, esta

suposición es inválida para arreglos de sensores con espaciado *pequeño*. Si se considera que la PSD del ruido en cada sensor es la misma ($\phi_{n_i n_i} = \phi_{nn}$), entonces $\Gamma_{n_i n_j}$ se puede reescribir como:

$$\Gamma_{n_i n_j} = \frac{\phi_{n_i n_j}}{\phi_{nn}} \quad (4.16)$$

Empleando la ecuación anterior, la Densidad Espectral de Potencia de la señal se puede expresar como:

$$[\Phi_{ss}(\omega)]_{ij} = \frac{\Re\{Y_i(\omega)Y_j^*(\omega)\} - \frac{1}{2}\Re\{\Gamma_{n_i n_j}\}(Y_i(\omega)Y_i^*(\omega) - Y_j(\omega)Y_j^*(\omega))}{1 - \Re\{\Gamma_{n_i n_j}\}} \quad (4.17)$$

Sustituyendo esta expresión en el numerador de la ecuación (4.14), se obtiene:

$$H_{MC}(\omega) = \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M [\Phi_{ss}(\omega)]_{ij}}{\frac{1}{M} \sum_{i=1}^M |Y_i(\omega)|^2} \quad (4.18)$$

El valor de $\Gamma_{n_i n_j}$, se puede calcular a partir de segmentos de la señal donde no se encuentra la señal de voz de interés.

4.1.4 Postfiltro APES

El algoritmo de Extensión de Postfiltrado Adaptable para Formador de Haz Superdirectivo (APES) fue propuesto por Joerg Bitzer, Klaus Simmer y Karl-Dirk Kammeyer como una secuencia de dos postfiltros en cascada. Este sistema emplea un BF superdirectivo (sección 3.3.4). Empleando la estructura de un BF GSC (sección 3.3.7), esto se debe a que un formador de haz MVDR es una extensión de un formador de haz LCMV y, así mismo, éste es una extensión del formador de haz GSC. En la figura 4.4 se muestra el diagrama de bloques de este algoritmo.

Los coeficientes de cada uno de los filtros están dados por las ecuaciones (4.19) y (4.20) [56].

$$H_0(\omega) = \frac{M}{M-1} \frac{P_{UU} - \frac{1}{M^2} \sum_{i=0}^{M-1} P_{X_i X_i}}{P_{UU}} \quad (4.19)$$

$$H_1(\omega) = \frac{P_{ZZ}}{P_{UU}} \quad (4.20)$$

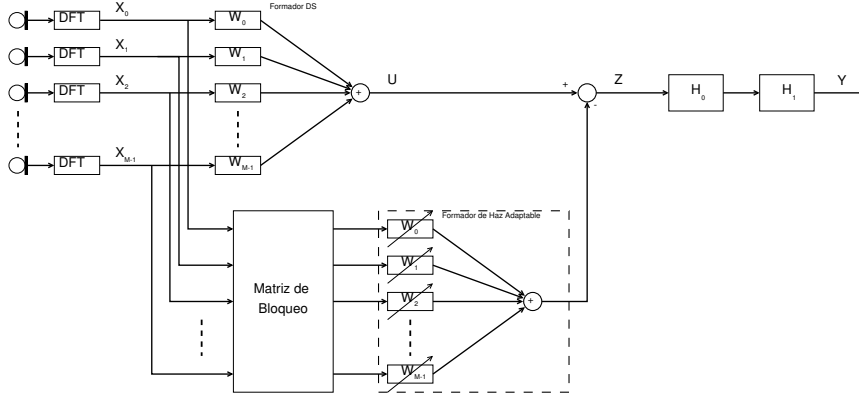


Figura 4.4: Postfiltrado APES.

donde P_{UU} es la PSD de la salida del formador de haz fijo, $P_{X_i X_i}$ es la CPSD de las señales de los micrófonos, y P_{ZZ} es la PSD de la salida del BF superdirectivo. El filtro $H_0(\omega)$ se encarga de suprimir el ruido difuso de alta frecuencia, mientras que el filtro $H_1(\omega)$ suprime el ruido de baja frecuencia, para obtener una mejor calidad en la señal de voz de salida, se recomienda restringir estos filtros de la siguiente manera [56]:

$$0.05 \leq H_0(\omega)H_1(\omega) \leq 1 \quad (4.21)$$

4.1.5 Postfiltro APAB

El Postfiltro Adaptable para Formador de Haz Arbitrario (APAB), propuesto por Simmer [28], emplea una etapa de postfiltrado la cual, se basa en un formador de haz MVDR. Este algoritmo considera que el ruido a la salida del BF ya se encuentra minimizado, de tal manera que los parámetros del postfiltro se determinan como:

$$H_{apab}(\omega) = \frac{\phi_{yy}}{\phi_{xx}} \quad (4.22)$$

donde ϕ_{yy} es la potencia de salida del BF y ϕ_{xx} es la potencia del micrófono más cercano a la señal deseada o alternatively puede ser la potencia promedio a la entrada del Formador de Haz. En la figura 4.5 se muestra el diagrama de bloques de este sistema.

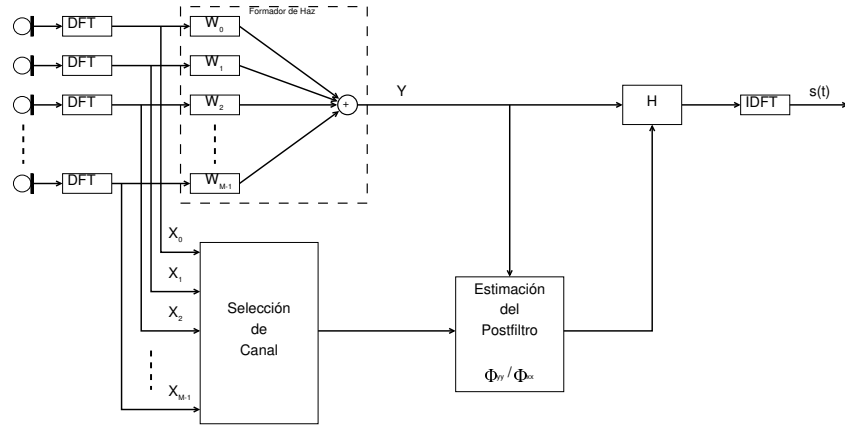


Figura 4.5: Postfiltrado APAB.

4.2 Medidas de Desempeño

Para evaluar el desempeño del sistema es necesario contar con medidas apropiadas, las cuales deben aproximarse a la forma en que el ser humano percibe las señales de voz. Es por eso que en esta sección se describirá una metodología para evaluar el desempeño de los formadores de haz desde el enfoque del filtrado espacial.

Una forma de comparar el desempeño de los formadores de haz es mediante el uso de pruebas subjetivas, las cuales emplean un gran número de personas; sin embargo, si se tiene acceso a las señales originales, se pueden emplear medidas numéricas objetivas, las cuales se basan en medir la “distancia” entre la señal deseada y la señal de salida del filtro espacial [16]. Para que este tipo de medidas sean válidas, deben tener relación con la forma en que los seres humanos percibimos las señales de voz; es por eso que en [57], se propone la descomposición de la señal de salida del BF como la suma de la señal deseada, las señales de interferencia, el ruido y artefactos:

$$\hat{s}_j = S_{target} + e_{interf} + e_{noise} + e_{artif} \quad (4.23)$$

Siendo $\hat{s}_j$ la señal de salida del Formador de Haz cuando éste se encuentra “apuntando” a la dirección de la señal deseada (j -ésima señal). Para lograr la descomposición se emplean distintos operadores ortogonales ($\Pi\{\}$) los cuales proyectan la señal $\hat{s}_j$ en el subespacio de la señal deseada:

$$P_{s_j} = \Pi\{s_j\} \quad (4.24)$$

en el subespacio de las N señales existentes (originales):

$$P_s = \Pi\{(s_{j'})_{1 \leq j' \leq N}\} \quad (4.25)$$

y en el subespacio de las N señales existentes en conjunto con las M señales de ruido:

$$P_{s,\eta} = \Pi\{(s_{j'})_{1 \leq j' \leq N}, (\eta_i)_{1 \leq i \leq M}\} \quad (4.26)$$

Empleando los proyectores de las ecuaciones (4.24), (4.25) y (4.26) se puede descomponer la señal de salida del filtro ($\hat{s}_j$), en las componentes mostradas en la ecuación (4.23), de la siguiente manera [57]:

$$S_{target} = P_{s_j} \hat{s}_j \quad (4.27)$$

$$e_{interf} = P_s \hat{s}_j - P_{s_j} \hat{s}_j \quad (4.28)$$

$$e_{noise} = P_{s,\eta} \hat{s}_j - P_s \hat{s}_j \quad (4.29)$$

$$e_{artif} = \hat{s}_j - P_{s,\eta} \hat{s}_j \quad (4.30)$$

Dado que S_{target} es la proyección de la señal estimada en el subespacio de la señal deseada, se puede emplear el producto interno para su cálculo, obteniéndose:

$$S_{target} = \frac{\langle \hat{s}_j, s_j \rangle s_j}{\|s_j\|^2} \quad (4.31)$$

En muchas ocasiones, la señal estimada y las señales originales no son mutuamente ortogonales, por lo que el cálculo de $P_s \hat{s}_j$ se realiza como :

$$P_s \hat{s}_j = \vec{c}^H \mathbf{S} \quad (4.32)$$

Donde $\vec{c} = \mathbf{R}_{ss}^{-1} \vec{u}^H$, siendo $\mathbf{R}_{ss}$ la matriz de Gram cuyos elementos se definen como $[\mathbf{R}_{ss}]_{j,j'} = \langle s_j, s_{j'} \rangle$, y $\vec{u} = [\langle \hat{s}_j, s_1 \rangle, \langle \hat{s}_j, s_2 \rangle, \dots, \langle \hat{s}_j, s_N \rangle]^H$. Obteniéndose e_{interf} como:

$$e_{interf} = \vec{c}^H \mathbf{S} - S_{target} \quad (4.33)$$

Si se considera que las señales originales son ortogonales a las señales de ruido, entonces $P_{s,\eta}$ se puede obtener como:

$$P_{s,\eta} \hat{s}_j \approx P_{s_j} \hat{s}_j + \sum_{i=1}^M \frac{\langle \hat{s}_j, \eta_i \rangle \eta_i}{\|\eta_i\|^2} \quad (4.34)$$

Obteniéndose:

$$e_{noise} = \sum_{i=1}^M \frac{\langle \hat{s}_j, \eta_i \rangle \eta_i}{\|\eta_i\|^2} \quad (4.35)$$

$$e_{artif} = \hat{s}_j - \vec{c}^H \mathbf{S} - e_{noise} \quad (4.36)$$

Estas medidas son las más populares al evaluar los métodos de separación de audio [16]. Como se mencionó, para emplear dichas medidas se deben conocer las señales deseadas de interferencia y de ruido, lo cual no siempre es posible en un ambiente real; sin embargo, en un ambiente controlado, es posible tener acceso a estas señales.

4.2.1 Relación Señal a Ruido

La Relación Señal a Ruido (SNR) es una medida sumamente empleada en el Procesamiento Digital de Señales, ya que relaciona la potencia de la señal de interés y la potencia del ruido. En [57] se propone el cálculo de la relación fuentes a ruido como:

$$SNR = 10 \log_{10} \frac{\|S_{target} + e_{interf}\|^2}{\|e_{noise}\|^2} \quad (4.37)$$

4.2.2 Relación Señal a Interferencia

En general, pueden existir múltiples fuentes que pueden “competir” con la señal de interés. A este tipo de señales se les conoce como interferencias, y son ejemplos de ellas las señales de voz provenientes de diferentes personas, las cuales se encuentran hablando al mismo tiempo; la música de fondo, etc. Para medir el nivel de filtrado respecto a este tipo de fuentes (las cuales se

encuentran correlacionadas de forma temporal) se emplea la Relación Señal a Interferencia (SIR):

$$SIR = 10\log_{10} \frac{\|S_{target}\|^2}{\|e_{interf}\|^2} \quad (4.38)$$

4.2.3 Relación Señal a Distorsión

La distorsión se define como el cambio en la forma de onda de salida de un sistema con respecto a la forma de onda de entrada y puede ser de tres tipos. *Distorsión en frecuencia*: las magnitudes de las componentes frecuenciales de la señal son alteradas. *Distorsión de retardo (fase)*: se da cuando el retraso de las componentes frecuenciales de la señal no es constante. *Distorsión ali-neal*: se da cuando la salida no es proporcional a la entrada, se manifiesta por la aparición de componentes de frecuencia que no se encuentran presentes en la señal original [6].

La distorsión se puede estimar mediante el error de estimación de la señal de salida del filtro espacial, siendo el error de estimación la suma de los términos de interferencia, ruido y artefactos de la señal de salida. La Relación Señal a Distorsión (SDR) se obtiene como:

$$SDR = 10\log_{10} \frac{\|S_{target}\|^2}{\|e_{interf} + e_{noise} + e_{artif}\|^2} \quad (4.39)$$

4.2.4 Relación Señal a Artefactos

Los artefactos son errores en la percepción o representación de una señal, los cuales son generados por el equipo o el algoritmo empleado, en general, en señales de voz, este tipo de errores puede hacer que la voz suene antinatural. La Relación Señal a Artefactos (SAR) se puede calcular como:

$$SAR = 10\log_{10} \frac{\|S_{target} + e_{interf} + e_{noise}\|^2}{\|e_{artif}\|^2} \quad (4.40)$$

4.3 Simulaciones

En esta sección se analizará el comportamiento de los algoritmos mostrados previamente empleando señales sintéticas. Para esto se realizará la com-

paración entre la señal de arribo al arreglo de micrófonos y la señal de salida del sistema, con el fin de evaluar la reducción de ruido e interferencias obtenida.

4.3.1 Comparación de Formadores de Haz

A continuación, se muestra la comparación entre los algoritmos mencionados en la sección 3.3. Considerando dos señales provenientes de distintos ángulos. La señal deseada es una senoidal con frecuencia de 500 Hz, la cual proviene de la dirección $\theta = 50^\circ$, mientras que la señal de interferencia es una señal tipo diente de sierra de la misma frecuencia con ángulo de arribo de $\theta = 0^\circ$. El arreglo empleado consta de siete sensores espaciados 4 cm y se utilizó una f_s de 8 KHz.

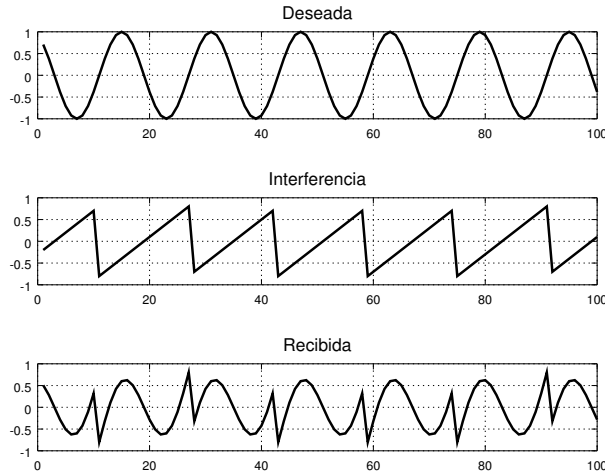


Figura 4.6: Señal de entrada al arreglo.

En la figura 4.6 se muestran las señales deseada, la de interferencia, y la mezcla de éstas para uno de los sensores. Las características de estas señales son²: $SNR_{in} = \infty$, $SIR_{in} = 1.44$, $SDR_{in} = 1.44$ y $SAR_{in} = 314.77$.

²Al no haber presencia de ruido, la Relación Señal a Ruido es infinita; de igual manera, al ser señales simuladas, la presencia de artefactos es mínima

Para evaluar la reducción de ruido debido al BF, se agregó ruido aditivo uniforme a las señales. En la figura 4.7 se muestra dicha señal, cuyas características son: $SNR_{in} = 7.81$, $SIR_{in} = 1.84$, $SDR_{in} = -0.22$ y $SAR_{in} = 20.85$.

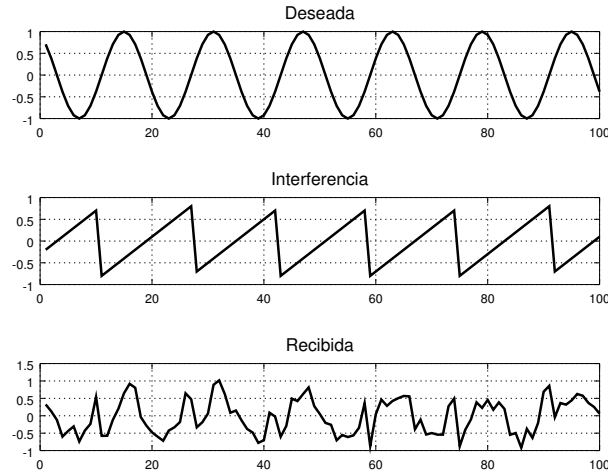


Figura 4.7: Señal de entrada al arreglo con ruido.

Señal sin ruido

En la tabla 4.1 se muestran los resultados a la salida del arreglo, para varios Formadores de Haz, destacando que el mayor nivel de separación se obtiene empleando el BF LCMV, mientras que el menor nivel de distorsión y artefactos se obtuvieron empleando el algoritmo de remuestreo espacial con MVDR.

En la figura 4.8 se muestran las señales de salida para cada uno de los algoritmos. Se puede observar que la forma de onda de la señal obtenida empleando el algoritmo LCMV es la que más se aproxima a la señal deseada, sin embargo, presenta una alta distorsión en fase, por esta razón el nivel de distorsión y de artefactos fue mayor al de los demás algoritmos mostrados.

Algoritmo	SNR_{out}	SIR_{out}	SDR_{out}	SAR_{out}
DS	∞	9.99	-2.43	-1.76
MVDR	∞	15.00	3.46	3.91
MVDR-PHAT	∞	14.62	2.78	3.22
LCMV	∞	20.28	-10.42	-10.37
GSC	∞	12.60	-1.47	-1.06
MVDR-RS	∞	18.64	8.44	8.93

Tabla 4.1: Desempeño de los algoritmos con señales sin ruido.

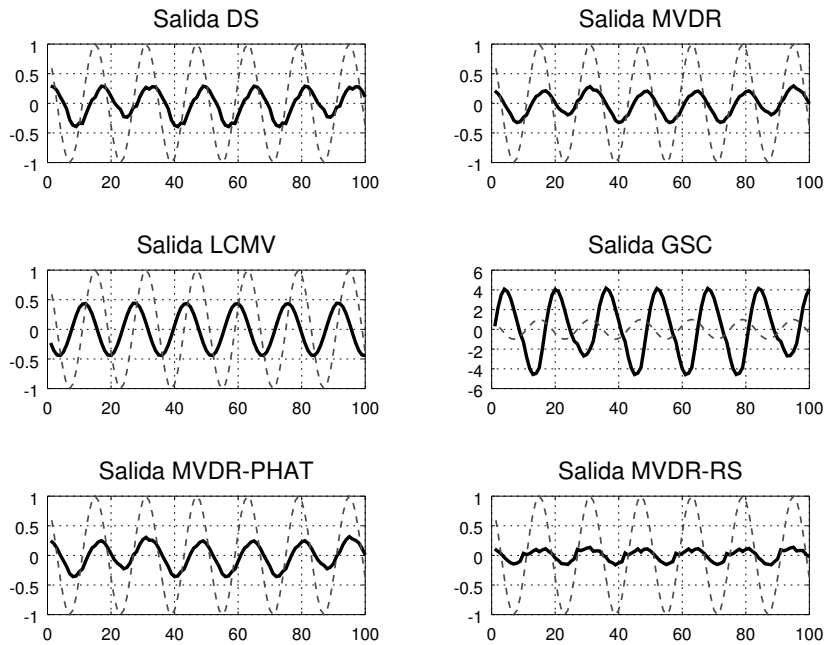


Figura 4.8: Señal de salida de los BF sin ruido.

Señal con ruido

Los resultados empleando las señales con ruido se muestran en la tabla 4.2; en este caso, el algoritmo MVDR con remuestreo espacial fue el que presentó mejores resultados. En la figura 4.9 se muestran las señales de salida para estos algoritmos.

Algoritmo	SNR_{out}	SIR_{out}	SDR_{out}	SAR_{out}
DS	13.17	11.02	-3.24	-2.47
MVDR	13.21	11.87	-3.90	-3.26
MVDR-PHAT	14.56	13.41	-2.95	-2.46
LCMV	3.03	14.42	-17.98	-15.86
GSC	19.49	17.05	0.13	0.40
MVDR-RS	21.42	21.96	2.36	2.55

Tabla 4.2: Desempeño de los algoritmos con señales con ruido.

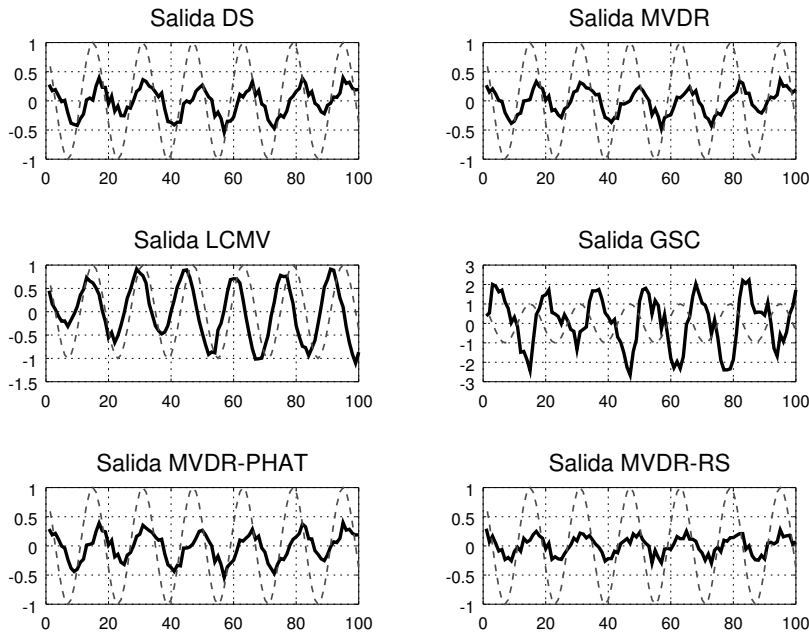


Figura 4.9: Señal de salida de los BF con ruido.

4.3.2 Comparación de Algoritmos de Postfiltrado

De la subsección anterior, se puede concluir que el algoritmo de remuestreo espacial empleando MVDR presenta mejores resultados, por lo que para la evaluación de los algoritmos de postfiltrado, se empleó la señal de salida de este formador de haz.

Señal sin ruido

En la tabla 4.3 se muestran los resultados para diferentes algoritmos de post-filtrado, apreciándose que el algoritmo que muestra un mejor desempeño es el APAB, el cual aumenta de forma significativa el nivel de separación de la señal. En la figura 4.10 se muestran las salidas de los postfiltros.

Algoritmo	SNR_{out}	SIR_{out}	SDR_{out}	SAR_{out}
Zelinski	∞	21.38	6.97	7.16
McCowan	∞	23.42	9.39	9.59
APAB	∞	58.49	12.78	12.78

Tabla 4.3: Desempeño de los algoritmos de postfiltrado con señales sin ruido.

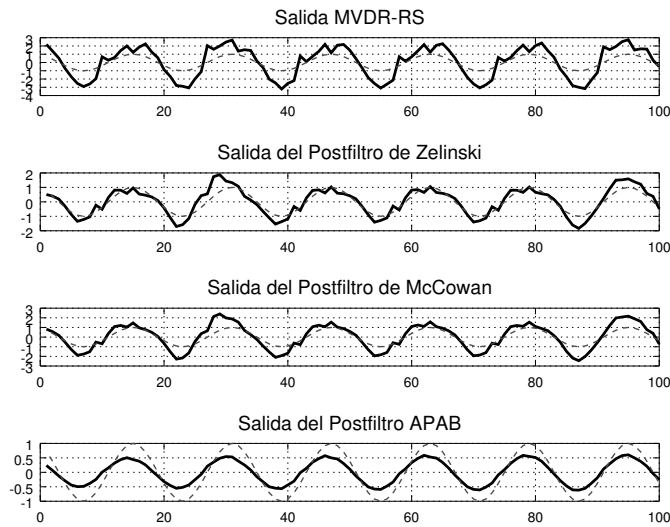


Figura 4.10: Señal de salida de los postilftro sin ruido.

Señal con ruido

La salida de los algoritmos de postfiltrado para señales en presencia de ruido se muestran en la tabla 4.4, se observa que el postfiltro APAB obtiene el

mejor nivel de reducción de ruido, mientras que el postfiltro de McCowan logra un mejor nivel de separación en las señales. Las señales de salida se muestran en la figura 4.11.

Algoritmo	SNR_{out}	SIR_{out}	SDR_{out}	SAR_{out}
Zelinski	30.13	24.33	7.59	7.69
McCowan	21.58	27.72	7.38	7.53
APAB	35.96	24.92	7.97	8.11

Tabla 4.4: Desempeño de los algoritmos de postfiltrado con señales con ruido.

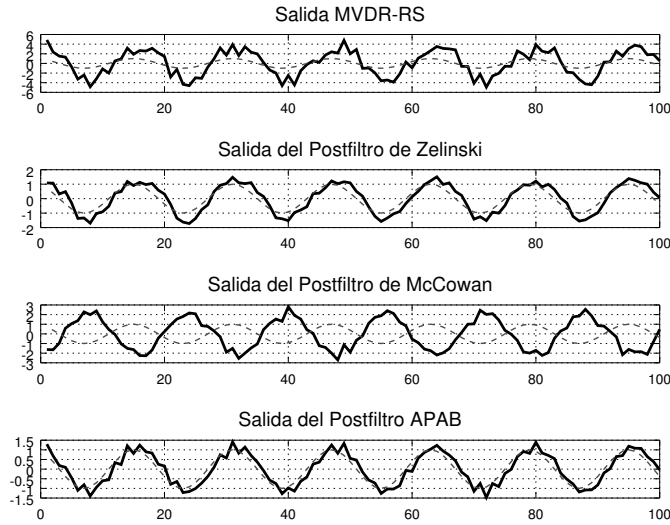


Figura 4.11: Señal de salida de los postilftro con ruido.

De los resultados mostrados, se puede concluir que el algoritmo de Respuesta sin Distorsión de Mínima Varianza empleando Remuestreo Espacial, en conjunto con el Postfiltro Adaptable para Formador de Haz Arbitrario, es el que exhibe la mejor relación de reducción de ruido e interferencias, minimizando la distorsión y artefactos. Otro punto a tomar en cuenta es que, al emplear estos métodos el número de operaciones que se deben realizar es menor que empleando otros algoritmos de filtrado, por lo que es factible emplearlos para filtrar señales en tiempo real.

Resumen

En este capítulo se mostraron algunos algoritmos de postfiltrado que emplean las características espaciales de la señal de voz para minimizar la presencia de ruido e interferencias; así mismo, se presentaron medidas para la evaluación objetiva de las señales con base en la SNR, SIR, SDR y SAR. Los algoritmos se simularon empleando señales sintéticas a las cuales se les agregó ruido uniforme no correlacionado. De las tablas 4.2 y 4.4, se puede concluir que el algoritmo MVDR, empleando remuestreo espacial en conjunto con el post-filtro APAB, es el que presenta mejores resultados, en comparación con los otros algoritmos mencionados.

Capítulo 5

Diseño e Implementación del Sistema

En la sección 4.3 se mostraron diferentes simulaciones para los formadores de haz mencionados en el capítulo 3, de las cuales se puede deducir que el sistema MVDR, empleando remuestreo espacial, en conjunto con el postfiltro APAB, es el que mejor resultados presenta, por lo que en este capítulo se describirá con mayor detalle el diseño de este sistema.

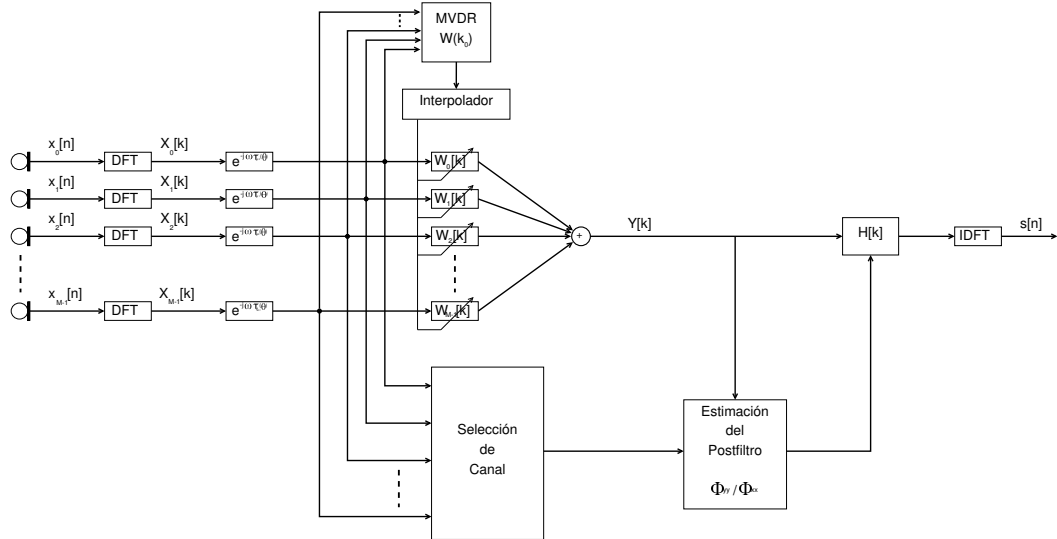


Figura 5.1: Diagrama general del sistema.

En la figura 5.1, se muestra el diagrama general del sistema implementado, cuyos bloques principales son:

- Transformada Discreta de Fourier: para obtener el espectro de la señal de cada micrófono.
- Formador de Haz MVDR: calcula los coeficientes de la frecuencia de referencia.
- Interpolador: obtiene los coeficientes para el resto de las frecuencias.
- BF en frecuencia: aplica los coeficientes interpolados ($W_i[k]$) para cada banda de frecuencia.
- Postfiltro: disminuye el ruido incoherente presente en la señal, se compone del selector de canal, el estimador del postfiltro y el filtrado ($H[k]$).

5.1 Diseño

El sistema se diseñó de tal manera que obtiene señales de un arreglo de micrófonos, colocados en diferentes puntos del espacio, y filtra las señales provenientes de una dirección específica, atenuando las señales de direcciones diferentes a las de interés. El sistema cuenta con dos etapas: Formador de Haz y Postfiltro, cuyo diseño se describe a continuación.

5.1.1 Diseño del Formador de Haz

Como se mencionó previamente, se empleó un Formador de Haz de Respuesta sin Distorsión de Mínima Varianza ya que este BF minimiza la potencia de las señales de interferencia sin modificar la señal de interés, sección 3.3.4. Sin embargo, la implementación de este algoritmo para señales de banda ancha, como es el caso de la señal de voz, conlleva un alto costo computacional, ya que se debe obtener la inversa de la matriz de correlación para cada una de las bandas de frecuencias a emplear, ecuación (5.1).

$$\vec{W}(\omega, \theta) = \frac{\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}(\omega) \vec{a}(\omega, \theta)}{\vec{a}^H(\omega, \theta) \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}(\omega) \vec{a}(\omega, \theta)} \quad (5.1)$$

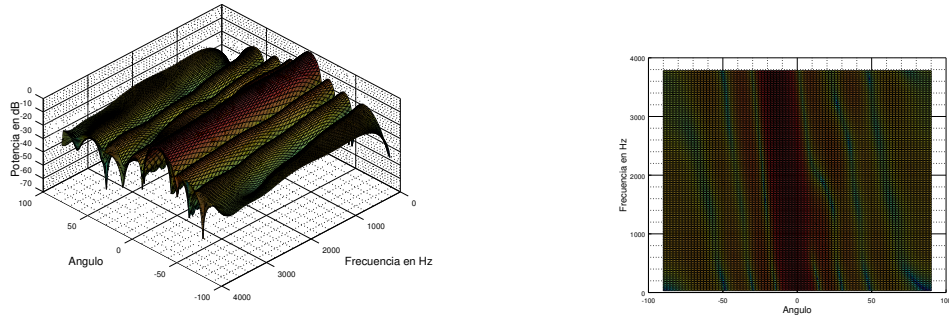
Por esta razón, se utilizó modelado numérico para calcular los coeficientes del formador de haz de banda ancha a partir de los coeficientes obtenidos en una

frecuencia de referencia. El algoritmo empleado es el de *remuestreo espacial*, mencionado en la sección 3.4.3. Este algoritmo requiere que se calculen los coeficientes $\vec{W}(\omega_0, \theta)$, empleando la ecuación (5.1), y mediante la ecuación (5.2), se obtienen los coeficientes interpolados para el resto de las frecuencias.

$$W_m(\omega, \theta) = \frac{d_0}{d_\omega} \sum_{i=1}^M W_i(\omega_0) \frac{\sin[\pi((m-1)d_0 - (i-1)d_\omega)/d_\omega]}{\pi((m-1)d_0 - (i-1)d_\omega)/d_\omega} \quad (5.2)$$

donde d_0 es la distancia entre los micrófonos y d_ω es la distancia óptima entre los micrófonos para la frecuencia ω , la cual se obtiene empleando la ecuación (3.9).

Debido a que el remuestreo espacial se basa en la interpolación de los coeficientes del BF y solo toma en cuenta el vector de dirección ($\vec{a}$) para los coeficientes de la frecuencia de referencia, se presenta un efecto de “desvío” en la dirección del haz. En la figura 5.2 se aprecia que al disminuir la frecuencia de la señal, por debajo de los 2 KHz, la dirección a la que “apunta” el BF tiende a reorientarse al ángulo $\theta = 0^\circ$; este comportamiento puede compensarse empleando retardos en la señal de entrada, similar al BF GSC de la sección 3.3.7, para de esta manera “alineal” las señales en la dirección deseada y obtener un comportamiento como el mostrado en la figura 3.24.



(a) Respuesta en frecuencia (isométrico). (b) Respuesta en frecuencia (vista superior).

Figura 5.2: SDP MVDR de banda ancha empleando remuestreo espacial.

Se empleó un Arreglo Lineal Uniforme (ULA) de siete micrófonos y, para evitar la aparición de lóbulos gratinados debido al aliasing espacial, sección 3.2.5, los micrófonos deben estar espaciados $d = \lambda_{min}/2$, donde $\lambda_{min} = c/f_{max}$. Como se mencionó en el capítulo 2, la señal de voz tiene un ancho de banda

de 4 KHz, por lo que la distancia entre micrófonos debe ser $d \leq 4.25$ cm. por esta razón la distancia empleada en el arreglo de micrófonos fue de 4 cm.

5.1.2 Diseño del Postfiltro

En el capítulo 4 se mostraron diferentes etapas de postfiltrado multicanal, las cuales emplean información espacial de las señales de voz para maximizar la SNR, donde el Postfiltro Adaptable para Formador de Haz Arbitrario (APAB), sección 4.1.5, mostró obtener mejores resultados en las simulaciones de la sección 4.3. Este postilftro se define como:

$$H_{apab}(\omega) = \frac{\phi_{yy}}{\phi_{xx}} \quad (5.3)$$

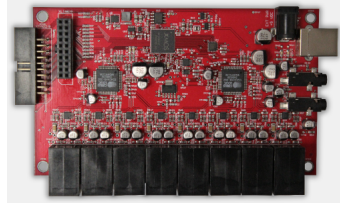
donde ϕ_{yy} y ϕ_{xx} son la Densidad Espectral de Potencia de la salida del BF y de uno de los canales de entrada, las cuales se calculan empleando el método de Welch [58]. La etapa de selección de canal requiere tomar como referencia el micrófono que se encuentre más cercano a la fuente, debido a que se tiene una etapa de retrasos, la cual alinea las señales de los micrófonos para que el BF se encuentre orientado en la dirección del broadside, se empleó la PSD promedio de los micrófonos para el cálculo de ϕ_{xx} .

5.2 Implementación

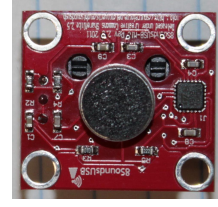
Para implementar el sistema, primero se creó un banco de señales multicanal, con el cual se pudiera probar y evaluar el funcionamiento del sistema, posteriormente se programaron los algoritmos mencionados en la sección anterior en diferentes lenguajes de programación.

5.2.1 Adquisición de las Señales

Para probar los algoritmos fuera de línea, se creó un *corpus* de señales provenientes de distintas direcciones, empleando la tarjeta *8soundsUSB*, figura 5.3, la cual cuenta con ocho entradas analógicas diferenciales [7] y un puerto USB de alta velocidad (*USB High-Speed*), mediante el cual se puede enviar audio multicanal de acuerdo con la especificación USB [59].



(a) Tarjeta de adquisición.



(b) Micrófonos empleados.

Figura 5.3: Tarjeta 8SoundsUSB [7].

Las señales de los micrófonos se almacenaron en diferentes archivos de audio con formato *.wav*, empleándose una frecuencia de muestreo de 48 KHz y un ancho de palabra de 16 bits.

Como se mencionó, se empleó un arreglo ULA de siete micrófonos con espaciamiento de cuatro centímetros, por lo que el tamaño del arreglo fue de 24 cm. Para considerar que la señal de voz tiene un frente de onda plano (sección 2.1.1), se debe cumplir con la condición mencionada en la ecuación (2.7), por lo que, para una frecuencia de 4KHz, la distancia entre la fuente y el arreglo de micrófonos debe ser $r > 2D^2/\lambda = 2(0.24)^2/0.085 = 1.3$ m. Las grabaciones se realizaron a una distancia $r \simeq 2$ m.

Este corpus se realizó empleando grabaciones estáticas y en movimiento, así como una grabación con dos personas hablando simultáneamente. Los ángulos empleados para la creación del corpus fueron:

- Estáticas: $\theta = -30^\circ, 0^\circ, 20^\circ, 35^\circ$ y 40° .
- Móviles: $\theta = 30^\circ$ a -40° .
- Simultánea: $\theta = -40^\circ$ y 30° .

Configuración del Equipo

Para realizar las grabaciones del corpus se montó el arreglo de micrófonos en un tripie con un disco transportador, para de esta manera, mantener fija la posición del arreglo de micrófonos y de la bocina, rotando el arreglo para modificar el DOA. En la figura 5.4 se muestra la configuración del arreglo.



(a) Señal arribando desde el Broadside. (b) Señal arribando desde la dirección θ .

Figura 5.4: Movimiento angular del arreglo [8].

El arreglo de micrófonos se conectó a la tarjeta de adquisición (figura 5.3a) y ésta, a su vez, se conectó a una computadora para la grabación de los canales de audio, figura 5.5.



Figura 5.5: Sistema de adquisición.

Las grabaciones se realizaron en el Laboratorio de Computación del Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS) de la UNAM, con la finalidad de obtener resultados en un ambiente ruidoso y reverberativo, de tal manera que se aproximen a ambiente reales.

5.2.2 Software

La implementación de los algoritmos previamente mencionados se realizó en programas que funcionan tanto fuera de línea como en tiempo real. La implementación de los algoritmos fuera de línea se realizó empleando *GNU Octave*, el cual es un lenguaje de alto nivel destinado a cálculos numéricos [60]. Para la implementación en tiempo real, se usó *JACK Audio Connection Kit*, esta es una API que provee la infraestructura básica para la creación de aplicaciones para procesamiento de audio en tiempo real [61].

Como se mencionó en la sección 2.3, la señal de voz puede ser considerada estacionaria cuando se emplean segmentos de 10 a 30 ms, por esta razón se emplea la técnica de Traslape y Suma (OLA), con tramas de 1024 datos, que equivalen a $\simeq 21$ ms, las cuales se encuentran superpuestas un 50%, figura 5.6, así mismo se empleó una ventana Hanning:

$$w[n] = 0.5 - 0.5\cos\left(\frac{2\pi n}{N-1}\right) \quad (5.4)$$

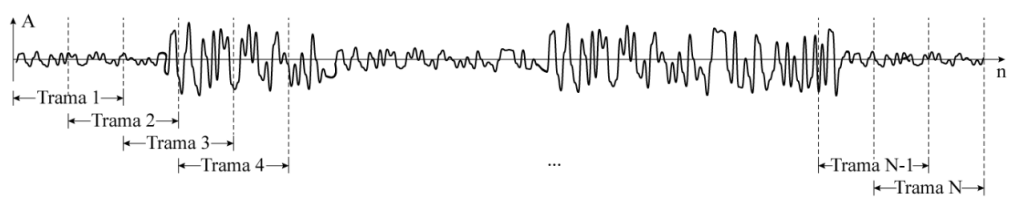


Figura 5.6: Ejemplo de OLA de una señal con traslape de 50% [9].

Esta ventana tiene las siguientes características [62]:

- Ancho del lóbulo principal: $8\pi/N$.
- Diferencia entre el lóbulo principal y el secundario: -41 dB.
- A_s mínima: -53 db

Los programas implementados realizan las siguientes tareas:

1. Lectura de los archivos de audio.

2. Traslape y Suma.
3. Transformada Discreta de Fourier de la ventana de cada canal.
4. Aplicación del vector de dirección para cada frecuencia discreta.
5. Cálculo de los coeficientes de referencia.
6. Cálculo de los coeficientes interpolados.
7. Aplicación de los coeficientes interpolados a la señal (BF).
8. Cálculo de la Densidad Espectral de Potencia de la entrada y salida.
9. Estimación del postfiltro (H).
10. Cálculo de la Transformada Discreta de Fourier Inversa.

Para el cálculo de los coeficientes del BF, se requieren los coeficientes de referencia, estos se calculan de acuerdo a los coeficientes MVDR estimados en una frecuencia previamente determinada. En este trabajo se probaron con diferentes frecuencias y se encontró que cuando se toma como referencia $f_0 = 500$ Hz se obtienen los mejores niveles de reducción de ruido y de separación, por lo que el sistema se diseñó empleando esta frecuencia como referencia.

Resumen

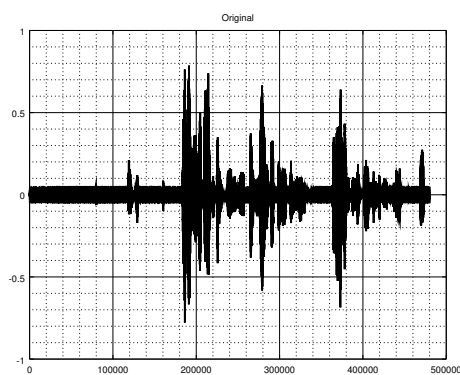
En este capítulo se describió el proceso de diseño del sistema, así como la implementación del mismo. El sistema cuenta con dos etapas, las cuales emplean la información espacial de la señal para realizar el filtrado. La primera de estas etapas es el Formador de Haz de banda ancha, en el cual se empleó modelado numérico para obtener un ancho de haz constante en frecuencia, la segunda etapa es un postfiltro, que emplea la PSD de la salida del BF en conjunto con la PSD de uno de los canales de entrada, el más cercano a la señal deseada, para la estimación de los parámetros del filtro.

Capítulo 6

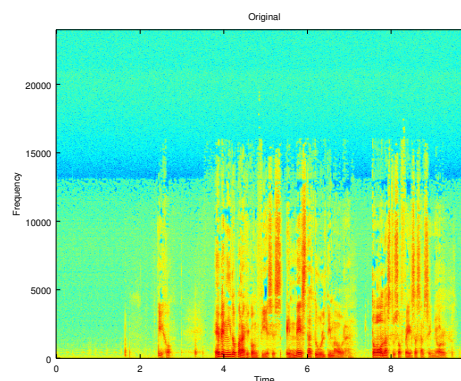
Pruebas y Resultados

En este capítulo se muestran las pruebas realizadas con el sistema implementado, así como los resultados obtenidos evaluando el nivel de reducción de ruido y la disminución de interferencias. Para obtener el nivel de desempeño del sistema se realizaron pruebas con una, dos y tres fuentes, las cuales consistieron en mezclar las grabaciones del corpus para generar distintos escenarios buscando filtrar la señal proveniente del broadside, en todos los casos se emplearon señales con una duración de 10 segundos.

En la figura 6.1 se muestra la señal de interés, la cual se encuentra en la dirección normal al arreglo.



(a) Señal deseada.



(b) Espectrograma de la señal deseada.

Figura 6.1: Señal original.

6.1 Resultados con Una Fuente

La primer prueba del sistema, descrito en el capítulo 5, consistió en la medición del nivel de reducción de ruido, es decir la Ganancia del Arreglo (AG), sección 3.2.6, en la figura 6.2 se muestra la señal recibida por uno de los micrófonos, mientras que en la tabla 6.1 se muestran los resultados obtenidos.

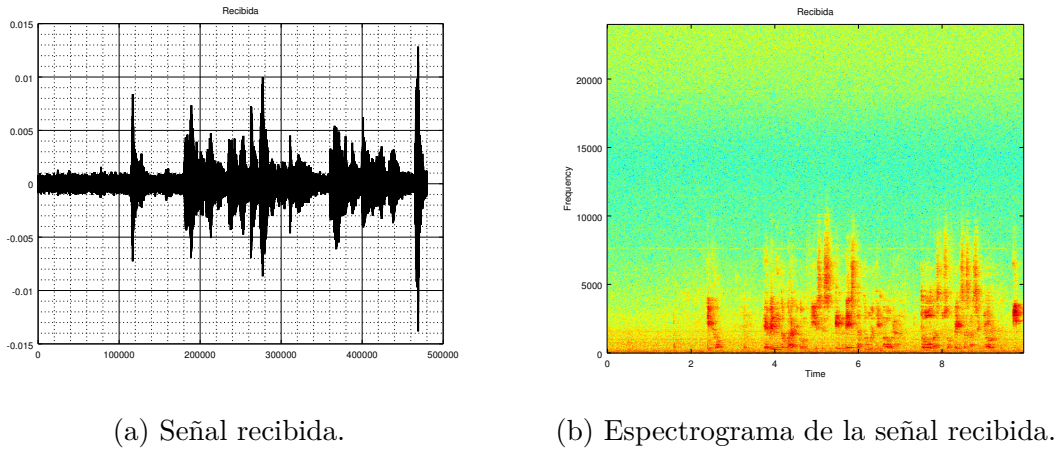
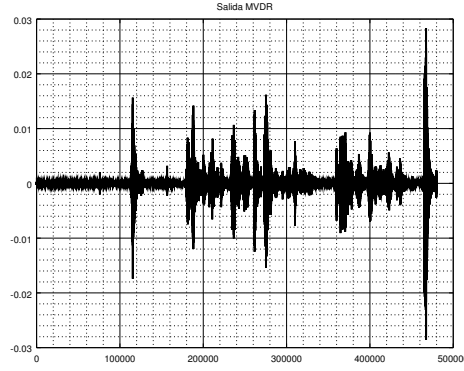


Figura 6.2: Señal captada sin interferencia.

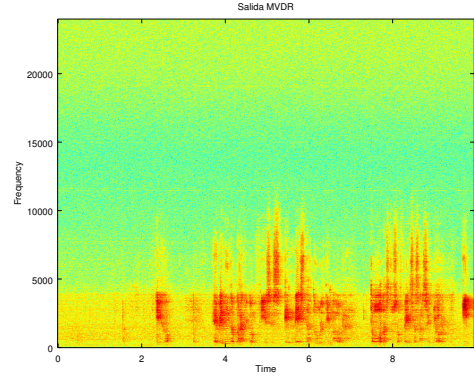
	Entrada	Salida del BF	Salida del Postfiltro
SNR	2.99	15.71	23.39
SDR	-34.17	-48.82	-51.14
SAR	-32.39	-48.71	-51.12

Tabla 6.1: Resultados señal sin interferencia.

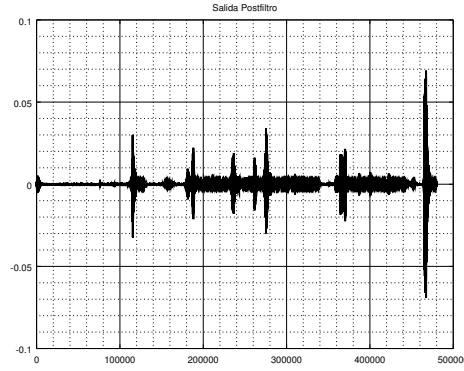
De la tabla se puede observar que el BF presenta una AG de mas de 12 dB, la cual aumenta hasta casi 20 dB cuando se emplea el postfiltro, sin embargo, el nivel de distorsión y artefactos aumenta considerablemente, 14 y 16 dB, respectivamente, cuando se emplea el BF y aumentan a 17 y 19 dB con el postfiltro. En la figura 6.3 se muestra la señal de salida tanto para el BF como para el postfiltro, así como sus espectrogramas.



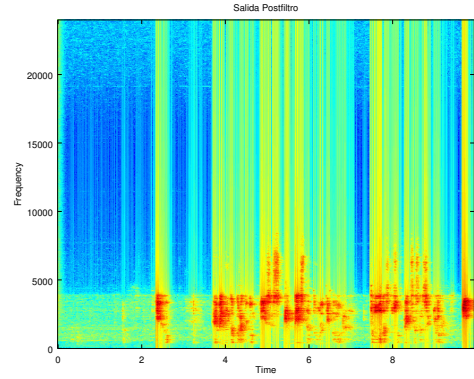
(a) Salida MVDR.



(b) Espectrograma de la salida MVDR.



(c) Salida Postfiltro.



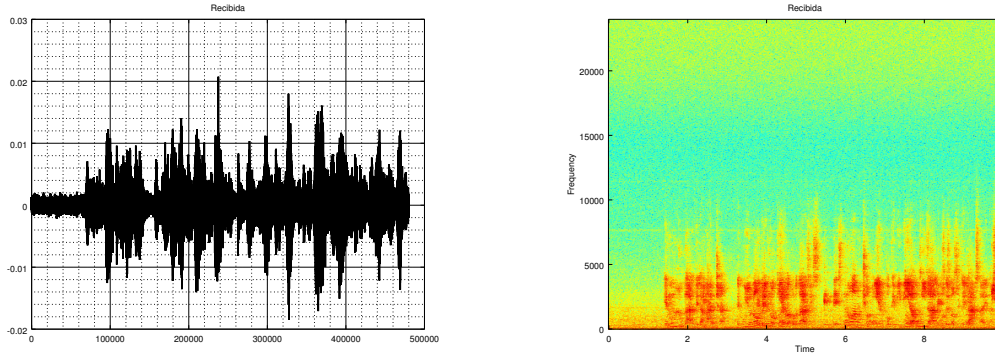
(d) Espectrograma de la salida del Postfiltro.

Figura 6.3: Salida del sistema sin interferencia.

6.2 Resultados con Dos Fuentes

Esta prueba consistió en mezclar la señal deseada con una señal proveniente de otra dirección, en la figura 6.4 se muestra la señal captada por uno de los micrófonos cuando existe una interferencia en la dirección $\theta = 20^\circ$.

De igual manera se midió la energía de la señal deseada (E_d), de la interferencia (E_i), así como la energía de la señal a la entrada del micrófono y a la salida tanto del BF (E_{MVDR}) como del postfiltro (E_{Post}):



(a) Señal recibida.

(b) Espectrograma de la señal recibida.

Figura 6.4: Señal captada con interferencia en 20° .

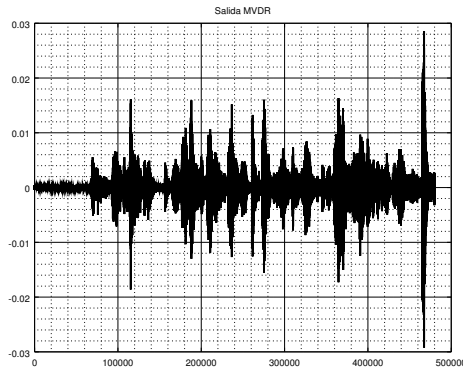
- $E_d = -4.90$ dB.
- $E_i = 1.40$ dB.
- $E_{in} = 2.30$ dB.
- $E_{MVD R} = 0.63$ dB.
- $E_{Post} = 4.74$ dB.

Los resultados obtenidos para estas condiciones se muestran en la tabla 6.2, donde se observa una ganancia de 4 dB de la señal deseada respecto a la señal interferencia debida al BF, mientras que al emplear el postfiltro se obtiene una ganancia de 32 dB, así mismo, se obtienen niveles de reducción de ruido de 6 dB y 30 dB para el BF y el postfiltro, respectivamente.

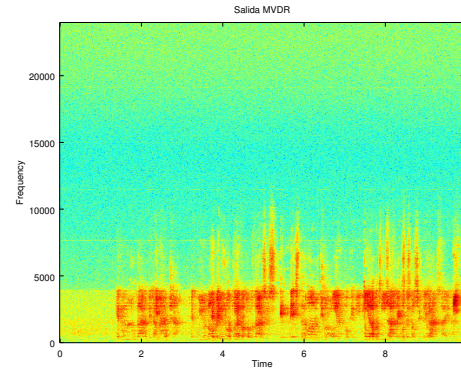
	Entrada	Salida del BF	Salida del Postfiltro
SNR	8.19	14.60	38.17
SIR	-7.39	-3.18	25.40
SDR	-44.03	-48.09	-48.03
SAR	-35.30	-43.05	-48.02

Tabla 6.2: Resultados con interferencia en 20° .

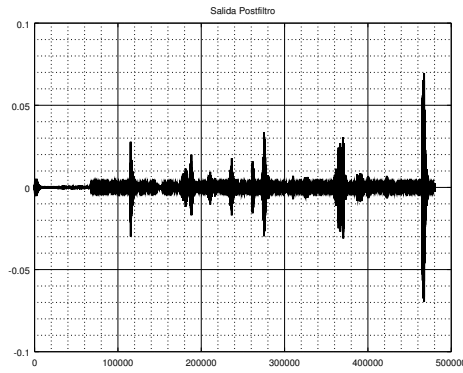
Cabe destacar que la señal de interferencia tiene una mayor energía que la señal deseada, por esta razón la SIR mostrada en la tabla 6.2 es negativa a la entrada del sistema. En la figura 6.5 se muestran las señales obtenidas, junto con sus espectrogramas.



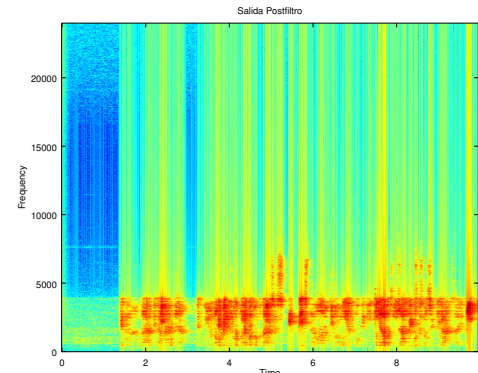
(a) Salida MVDR.



(b) Espectrograma de la salida MVDR.



(c) Salida Postfiltro.

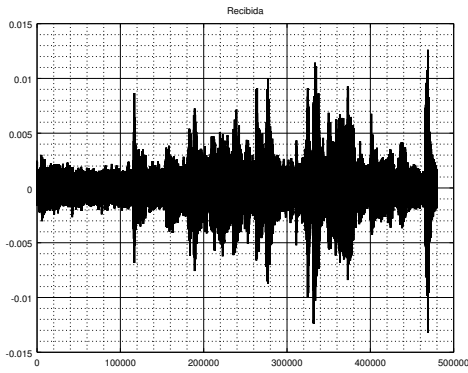


(d) Espectrograma de la salida Postfiltro.

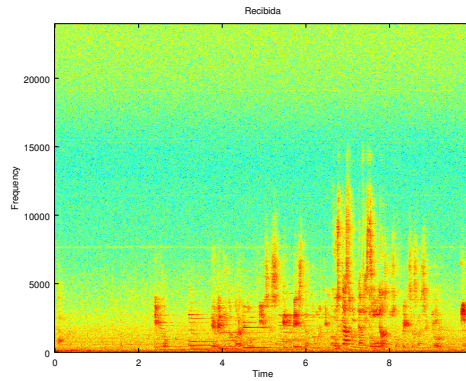
Figura 6.5: Salida del sistema con interferencia en 20° .

Así mismo se probó el sistema empleando una señal de interferencia en $\theta = 35^\circ$, en la figura 6.6 se muestra la señal obtenida por uno de los micrófonos, mientras que las energías de las señales fueron:

- $E_d = -4.90$ dB.
- $E_i = -4.29$ dB.
- $E_{in} = -1.51$ dB.
- $E_{MVDR} = -2.08$ dB.
- $E_{Post} = 3.80$ dB.



(a) Señal recibida.



(b) Espectrograma de la señal recibida.

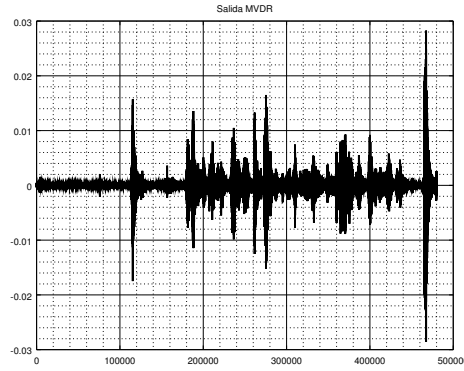
Figura 6.6: Señal captada con interferencia en 35° .

En la tabla 6.3 se muestran los resultados obtenidos, mientras que en la figura 6.7 se muestran las gráficas de la señal de salida.

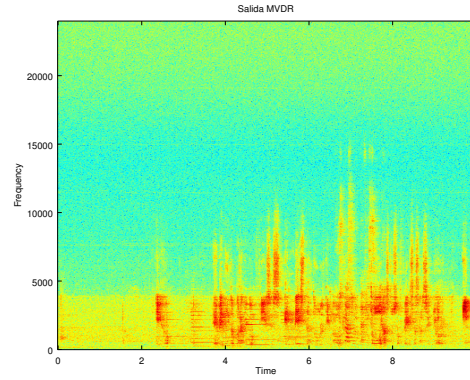
	Entrada	Salida del BF	Salida del Postfiltro
SNR	4.22	13.63	22.49
SIR	-6.42	-0.17	14.36
SDR	-37.21	-45.78	-49.86
SAR	-28.48	-42.50	-49.68

Tabla 6.3: Resultados con interferencia en 35° .

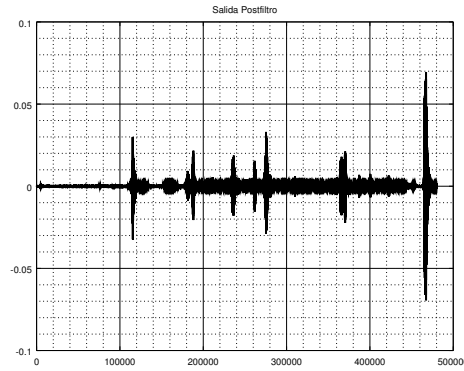
De la tabla 6.3 se aprecia una ganancia de 6 dB en la Relación Señal a Interferencia y de 9 dB en la Relación Señal a Ruido cuando se emplea el Formador de Haz, mientras que cuando se emplea en conjunto con el postfiltro la SNR aumenta en 18 dB y la SIR aumenta en 20 dB.



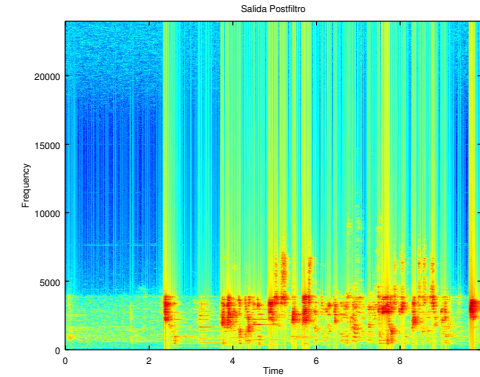
(a) Salida MVDR.



(b) Espectrograma de la salida MVDR.



(c) Salida Postfiltro.



(d) Espectrograma de la salida Postfiltro.

Figura 6.7: Salida del sistema con interferencia en 35° .

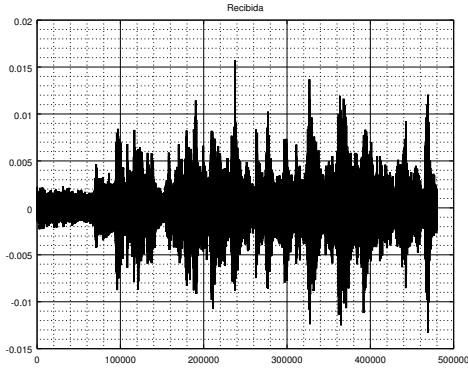
6.3 Resultados con Tres Fuentes

Se probó el sistema empleando dos señales interferencia, las cuales se encontraban en 20° y 35° , cuya relación de energías fue la siguiente:

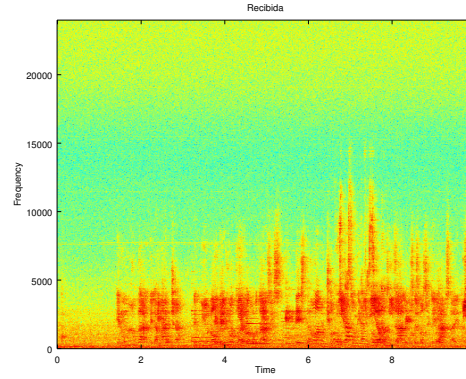
- $E_d = -4.90$ dB.
- $E_{i20} = -2.11$ dB.
- $E_{i35} = -1.08$ dB.
- $E_{in} = 0.42$ dB.

- $E_{MVDR} = -2.79$ dB.
- $E_{Post} = -0.14$ dB.

En la figura 6.8 se muestra la señal captada por uno de los micrófonos.



(a) Señal recibida.



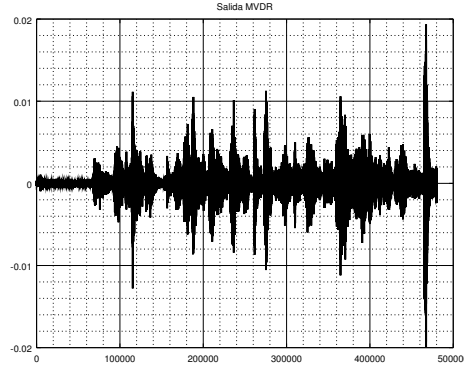
(b) Espectrograma de la señal recibida.

Figura 6.8: Señal captada con interferencia en 20° y 35° .

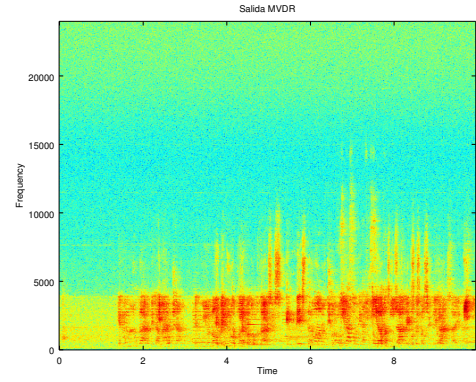
Los resultados obtenidos se muestran en la tabla 6.4, e la cual se observa que la Relación Señal a Interferencia es negativa a la entrada del sistema, por lo que la energía de las señales de interferencia es mayor que la energía de la señal deseada, sin embargo, el sistema es capaz de atenuar las señales indeseadas, aumentando la SIR en 22 dB, así mismo disminuye el ruido en 37 dB. En la figura 6.9 se muestran las señales obtenidas por el BF y el postfiltro.

	Entrada	Salida del BF	Salida del Postfiltro
SNR	4.81	11.84	41.20
SIR	-7.35	-3.17	15.07
SDR	-40.89	-46.83	-48.05
SAR	-31.55	-41.68	-47.92

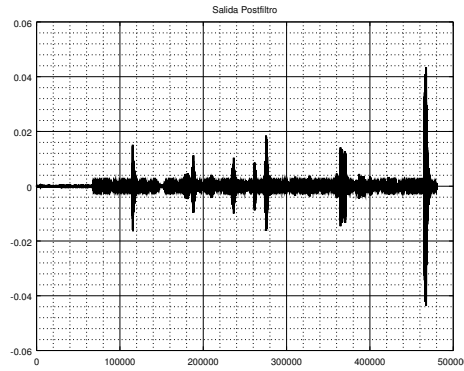
Tabla 6.4: Resultados con interferencias en 20° y 35° .



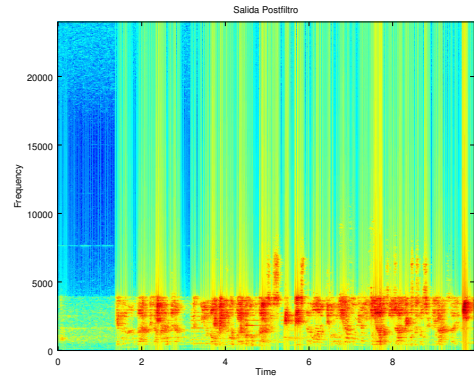
(a) Salida MVDR.



(b) Espectrograma de la salida MVDR.



(c) Salida Postfiltro.



(d) Espectrograma de la salida Postfiltro.

Figura 6.9: Salida del sistema con interferencias en 20° y 35° .

6.4 Evaluación del Formador de Haz Implementado

El sistema se evaluó con algunos de los algoritmos mostrados en el capítulo 3, esto con el fin de mostrar una comparación experimental de los formadores de haz en ambientes reales, cabe destacar que la implementación de estos algoritmos se realizó fuera de línea empleando *GNU Octave*. Las condiciones empleadas fueron las mismas que las empleadas en la sección 6.2 con la señal indeseada en la dirección $\theta = 35^\circ$. En la tabla 6.5 se muestran los resultados obtenidos:

	Entrada	DS	MVDR	GSC	MVDR-RS
SNR	4.22	1.40	0.92	0.92	14.60
SIR	-6.42	-15.00	-15.53	-15.53	-3.18
SDR	-37.21	-44.49	-45.16	-45.16	-48.09
SAR	-28.48	-26.97	-26.92	-26.92	-43.05

Tabla 6.5: Resultados para diferentes BF.

De la tabla anterior, se observa que el algoritmo implementado supera a los otros formadores de haz tanto, en nivel de separación como en reducción de ruido, figura 6.10, esto se debe a que el remuestreo espacial mejora la respuesta en frecuencia del BF, sin embargo, la cantidad de artefactos y de distorsión aumenta. Cabe destacar que el resultado obtenido empleando MVDR y GSC es el mismo, esto se debe a que, como se mencionó en el capítulo 3, el Formador de Haz GSC es una formulación alternativa del LCMV, el cual, a su vez, es una generalización del MVDR cuando el vector de respuesta tiene una sola restricción.

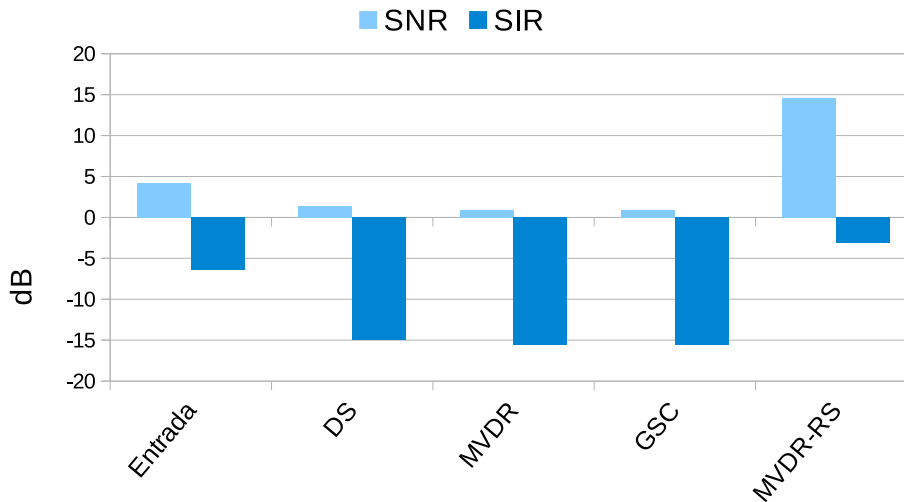


Figura 6.10: Comparación de SNR y SIR para diferentes BF.

Otra de las ventajas de emplear modelado numérico para el calculo de los coeficientes del BF es que la carga computacional disminuye en comparación con los algoritmos MVDR¹ y GSC. A continuación se muestran los tiempos promedio de ejecución del sistema implementado:

- Formador de Haz: 4.167 ms.
- Formador de Haz y Postilftro: 4.180 ms.

De los tiempo anteriores, se observa que la etapa de postfiltrado no aumenta de manera significativa el tiempo de proceso del sistema, sin embargo, mejora la respuesta de éste; así mismo, se observa que el tiempo empleado para procesar los siete canales de audio es menor al tiempo necesario para obtener una trama de la señal (21 ms.), sección 5.2.2. El sistema se probó en una computadora marca *Ghia* con las siguientes características:

- Procesador: Intel Core i7-4790.
- Velocidad: 3.6 GHz.
- Memoria RAM: 8 GB.
- Procesador Gráfico: Intel Haswell Desktop.
- Sistema Operativo: Ubuntu 16.04 LTS.

Resumen

En este capítulo se mostraron las pruebas realizadas al sistema, así como los resultados obtenidos para diferentes escenarios, los cuales variaron en el número de señales indeseadas. Se demostró que el sistema implementado mejora tanto la Relación Señal a Ruido como la Relación Señal a Interferencia, a costa de aumentar el nivel de artefactos y distorsión. De igual manera, se midió el tiempo requerido para procesar las señales, el cual fue menor al tiempo necesario para su adquisición, por lo que se comprueba que el sistema puede operar en línea. Así mismo, se realizó la comparación con algunos de los algoritmos comúnmente empleados (DS, MVDR y GSC), mostrando que

¹Este algoritmo no se pudo implementar en tiempo real debido que se requiere invertir una matriz de 7x7 por cada banda de frecuencia.

el sistema propuesto supera a dichos algoritmos; también se observa que el sistema responde de forma sobresaliente en condiciones adversas de ruido e interferencias.

Capítulo 7

Conclusiones

En este trabajo se presentó el diseño e implementación de un sistema de filtrado direccional, el cual es capaz de atenuar las señales provenientes de direcciones diferentes a la de interés y así mismo, disminuye el ruido no correlacionado con la señal deseada. Este sistema se puede redireccionar para filtrar señales que arriben desde distintos ángulos.

Debido a que la respuesta en frecuencia de los Formadores de Haz no es constante y tiende a ser pobre para frecuencias bajas, se empleó modelado numérico, en particular *remuestreo espacial*, para obtener una respuesta en frecuencia constante en el ancho de banda de la señal de voz.

El sistema se compone de dos bloques principales, el Formador de Haz y una etapa de Postfiltrado multicanal. El primero de estos bloques se encarga de realizar el filtrado espacial de las señales, mientras que el segundo maximiza la Relación Señal a Ruido de la señal. Cabe destacar que ambos bloques calculan sus coeficientes de acuerdo a la información estadística de las señales recibidas, por lo que estos se van adaptando de forma continua; para esto se adquieren bloques de 1024 datos de cada uno de los micrófonos con los cuales se obtienen la Matriz de Correlación y la Densidad Espectral de Potencia.

La matriz de correlación se emplea para calcular los coeficientes óptimos para una banda de frecuencia, los cuales se interpolan para obtener los coeficientes empleados para el resto de las frecuencias. La PSD se emplea para el cálculo de los coeficientes del postfiltro. Estos procesos se realizan en el dominio de la frecuencia, por lo que, la primer tarea que realiza el sistema, después de

la adquisición de los datos, es la Transformada Rápida de Fourier por cada canal.

El sistema se probó bajo condiciones adversas de ruido e interferencias, como se puede observar en los resultados del capítulo 6, donde se emularon diferentes escenarios, los cuales consistieron en:

1. Señal deseada en presencia de ruido.
2. Señal deseada y señal interferencia en presencia de ruido.
3. Señal deseada y dos señales interferencia en presencia de ruido.

De dichas pruebas, el sistema demostró ser capaz de mejorar la SNR en hasta 37 dB; así mismo aumentó la SIR en hasta 32 dB. De igual manera, se comparó el algoritmo implementado con diferentes BF convencionales, DS, MVDR y GSC, donde se demostró que el Formador de Haz empleando remuestreo espacial mejora las características de la señal de salida respecto a dichos algoritmos.

El sistema se implementó en una computadora con un procesador Intel Core i7-4790, obteniéndose tiempos de ejecución mucho menores al tiempo necesario para la obtención de los datos, requiriendo 4.180 ms. para realizar todo el proceso.

El sistema demostró ser capaz de filtrar señales de acuerdo a la dirección de arribo de estas al arreglo de micrófonos, sin embargo, en los resultados se observa que la señal de salida presenta distorsión y artefactos, por lo que en un futuro se pueden buscar algoritmos que disminuyan el nivel de distorsión.

Bibliografía

- [1] Donald E. Hall. *Basic acoustics*. Malabar, Florida : Krieger, 1993.
- [2] Agustín Silva Lang. Localización de una fuente de voz con arreglo de micrófonos. Master's thesis, Facultad de Ingeniería, UNAM, 2013.
- [3] Erwin Meyer, John M. Taylor, and Erwin Meyer. *Physical and applied acoustics : an introduction*. New York : Academic Press, 1972.
- [4] Wolfgang Herbordt. *Sound capture for human/machine interfaces : practical aspects of microphone array signal processing*. Lecture notes in control and information sciences: 315. Berlin: Springer, 2005.
- [5] Chris Rowden. *Speech processing*. McGraw-Hill, 1992.
- [6] William Littell Everitt and George E. Anner. *Ingeniería de comunicaciones*. Arbo, 1961.
- [7] IntRoLab. 8SoundsUSB main page. https://sourceforge.net/p/eightsoundsusb/wiki/Main_Page. Accesado: 2017-05-09.
- [8] Agustín Silva Lang. Detector de ángulo de arribo de una señal en un arreglo de micrófonos, 2010.
- [9] Mauricio Michel Olvera Zambrano. Sistema de identificación automática del lenguaje hablado en archivos multimedia de voz, 2017.
- [10] Seung Kwon Beack, Byunghwa Lee, Minsoo Hahn, and Seung Hyon Nam. Blind source separation and kalman filter-based speech enhancement in a car environment. *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 520–523, 2004.

- [11] Hamid Krim and Mats Viberg. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, (13):67–94, 1996.
- [12] Thomas D. Rossing. *Springer handbook of acoustics*. New York, New York : Springer, 2014.
- [13] Mingsian Bai, Jeong-Guon Ih, and Jacob Benesty. *Acoustic array systems, theory, implementation and application*. John Wiley & Sons, 2013.
- [14] Allan D. Pierce. *Acoustics: An introduction to its physical principles and applications*. Acoustical Society of America, 1994.
- [15] Robert J. Mailloux. *Phased array antenna handbook*. Boston : Artech, c1994, 1994.
- [16] Mohammad Dmour. *Mixture of Beamformers for Speech Separation and Extraction*. PhD thesis, The University of Edinburgh, 2010.
- [17] Simon Doclo. *Multi-microphone noise reduction for speech applications*. PhD thesis, Katholieke Universiteit Leuven, 2003.
- [18] Michel Bruneau. *Fundamentals of Acoustics*. ISTE Ltd., 2006.
- [19] Stephen T. Neely and Jont B. Allen. Invertibility of a room impulse response. *Journal of the Acoustical Society of America*, (6):165–169, 1979.
- [20] Luis Alberto Álvarez Fernández. Sistema de alarma residencial empleando voz sintética vía telefónica, 2014.
- [21] Samuel Salvador Vázquez Sánchez. Reducción de ruido en señales de voz mediante un formador de haz, 2012.
- [22] Edgar Tello Paleta. Diseño e implementación de un codificador decodificador perceptual de audio digital. Master’s thesis, Facultad de Ingeniería, UNAM, 2016.
- [23] Simon Haykin and James Justice. *Array signal processing*. Prentice Hall, 1985.
- [24] Barry Van Veen and Buckley Kevin. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, (5):4–24, 1998.

- [25] Jacob Benesty, Jingdong Chen, and Yiteng Huang. *Microphone array signal processing*. Springer, 2008.
- [26] Jeffrey Krolik and David Swingler. Focused wide-band array processing by spatial resampling. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38:356–360, 1990.
- [27] Stefania Repetto and Andrea Trucco. Designing superdirective microphone arrays with a frequency-invariant beam pattern. *IEEE Sensors Journal*, (6):737–747, 2006.
- [28] Michael Brandstein and Darren Ward. *Microphone arrays : signal processing techniques and applications*. Digital signal processing. Berlin: Springer Verlag, 2001.
- [29] Reuven Berkun, Israel Cohen, and Jacob Benesty. Combined beamformers for robust broadband regularized superdirective beamforming. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, (23):877–886, 2015.
- [30] Don H. Johnson and Dan E. Dudgeon. *Array signal processing : concepts and techniques*. Prentice-Hall signal processing series. Englewood Cliffs, New Jersey : Prentice Hall PTR, c1993, n.d.
- [31] Jack Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57:1408–1418, 1969.
- [32] Do Hoang and Harvey Silverman. A robust sound-source separation algorithm for an adverse environment that combines mvdr-phat with the casa framework. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011.
- [33] Charles Knapp and G. Clifford Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24:320–327, 1976.
- [34] Maurizio Omologo and Piergiorgio Svaizer. Use of the crosspower-spectrum phase in acoustic event localization. *Acoustics, Speech, and Signal Processing (ICASSP)*, pages II/273–II/276, 1997.

- [35] Wei Liu and Stephan Weiss. *Wideband Beamforming: Concepts and Techniques*. Wiley Blackwell, 2010.
- [36] Meng Er and Antonio Cantoni. Derivative constraints for broad-band element space antenna array processors. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31:1378–1393, 1983.
- [37] Kevin Buckley. Spatial/spectral filtering with linearly constrained minimum variance beamformers. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35:249–266, 1987.
- [38] Mahdi Parchami, Wei-Ping Zhu, Bemoit Champagne, and Eric Plourde. Recent developments in speech enhancement in the short-time fourier transform domain. *IEEE Circuits and Systems Magazine*, 16:45–77, 2016.
- [39] Jian Yang, Xu-qi Xi, and Yong Yu. A implementation method to constant beamwidth and adaptive beamforming of broadband signal. *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2014.
- [40] Bogdad Roberto Carlos Espinosa vargas. Filtrado espacial de señales acústicas. Master’s thesis, Facultad de Ingeniería, UNAM, 2013.
- [41] Darren B. Ward, Rodney A. Kennedy, and Robert C. Williamson. Fir filter design for frequency invariant beamformers. *IEEE Signal Processing Letters*, 3:69–71, 1996.
- [42] Marco Crocco and Andrea Trucco. A computationally efficient procedure for the design of robust broadband beamformers. *IEEE Transactions on Signal Processing*, (58):5420–5424, 2010.
- [43] Andrea Trucco and Stefania Repetto. Frequency invariant beamforming in very short arrays. *OCEANS ’04. MTTs/IEEE TECHNO-OCEAN ’04*, (2):635–640, 2004.
- [44] Hong Wang and Mostafa Kaveh. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33:823–831, 1985.

- [45] Henri Clergeot and Michel Olivier. New simple implementation of the coherent signal subspace method for wide band direction of arrival estimation. *Acoustics, Speech and Signal Processing (ICASSP)*, 1989.
- [46] Georges Bienvenu, Georges Vezzosi, and Laurent Kopp. Coherent wide band high resolution processing for linear array. *Acoustics, Speech and Signal Processing (ICASSP)*, 1989.
- [47] Jeffrey Krolik and David Swingler. The performance of minimax spatial resampling filters for focussing wide-band arrays. *IEEE Transactions on Signal Processing*, 39:1899–1903, 1991.
- [48] Jacob Benesty, M. Mohan Sondhi, and Yiteng Arden Huang. *Springer handbook of speech processing*. Berlin : Springer Science+Business Media, 2008.
- [49] Rainer Zelinski. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1988.
- [50] Joachim Thiemann and Emmanuel Vincent. An experimental comparison of source separation and beamforming techniques for microphone array signal enhancement. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013.
- [51] Israel Cohen. Multichannel post-filtering in nonstationary noise environments. *IEEE Transactions on Signal Processing*, 52:1149–1160, 2004.
- [52] Klauss Simmer and Alexander Wasiljeff. Adaptive microphone arrays for noise suppression in the frequency domain. *Workshop on Adaptive Algorithms in Communications*, 1992.
- [53] Claude Marro, Yannick Mahieux, and Klauss Simmer. Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering. *IEEE Transactions on Speech and Audio Processing*, 6:240–259, 1998.
- [54] Huajun Yu and Tim Fingscheidt. A post-filter for wideband speech beamforming in automotive applications. *International Symposium on Consumer Electronics (ISCE)*, 2010.

- [55] Iain McCowan and Hervé Bourslard. Microphone array post-filter based on noise field coherence. *IEEE Transactions on Speech and Audio Processing*, 11:709–716, 2003.
- [56] Joerg Bitzer, Klauss Simmer, and Karl-Dirk Kammeyer. Multi-microphone noise reduction by post-filter and superdirective beam-former. In *International Workshop on Acoustic Echo and Noise Control (IWAENC 99)*, pages 100–103, Pocono Manor, USA, Sep 1999.
- [57] Emmanuel Vincent, Rémi Gribonval, and Févotte Cédric. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 14:1462–1469, 2006.
- [58] Peter Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, (15):70–73, 1967.
- [59] USB Implementers Forum, Inc. *Universal Serial Bus Device Class Definition for Audio Devices*, 1998.
- [60] John W. Eaton. GNU Octave main page. <https://www.gnu.org/software/octave/>. Accesado: 2017-05-11.
- [61] Jason Costello. JACK Audio Connection Kit main page. <http://www.jackaudio.org/>. Accesado: 2017-05-11.
- [62] Larry Hipólito Escobar Salguero. *Diseño de filtros digitales*. Facultad de Ingeniería, UNAM, 2006.

Lista de símbolos

$\vec{a}$ Vector $\mathbf{a}$.

$\mathbf{A}$ Matriz $\mathbf{A}$.

$\mathbf{A}^T$ Transpuesta de la matriz $\mathbf{A}$.

$\mathbf{A}^H$ Hermitiano de la matriz $\mathbf{A}$.

$[\mathbf{A}]_{ij}$ Elemento (i,j) de la matriz $\mathbf{A}$.

$\mathbf{R}_{xx}$ Matriz de autocorrelación de la señal $\mathbf{x}$.

$\mathbf{R}_{xy}$ Matriz de correlación cruzada entre la señal $\mathbf{x}$ y la señal $\mathbf{y}$.

$\mathbf{\Gamma}_{vv}$ Matriz de coherencia del ruido.

$\mathbf{I}$ Matriz identidad.

$\vec{r}_{xx}$ Vector de autocorrelación de la señal $\mathbf{x}$.

$\vec{r}_{xy}$ Vector de correlación entre la señal $\mathbf{x}$ y la señal $\mathbf{y}$.

j Número imaginario equivalente a $\sqrt{-1}$.

f_s Frecuencia de muestreo.

T_s Periodo de muestreo.

c Velocidad del sonido ($\sim 340 \frac{m}{s}$ en aire a 22°C).

ω Frecuencia Angular.

$\vec{k}$ Vector de onda.

k Número de onda.

λ Longitud de onda.

$\Re$ Parte real de un número complejo.

dB Decibel.

I_A Intensidad Acústica.

I_{A0} Intensidad Acústica de referencia $10^{-12} W/m^2$.

∇ Operador diferencial en coordenadas cartesianas se expresa como $\nabla = \hat{x} \frac{\partial}{\partial x} + \hat{y} \frac{\partial}{\partial y} + \hat{z} \frac{\partial}{\partial z}$.

$\Pi\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k\}$ Proyector ortogonal en el subespacio de los vectores $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$.

T_{60} Tiempo de reverberación.

Acrónimos

AG Ganancia del Arreglo.

APAB Postfiltro Adaptable para Formador de Haz Arbitrario.

APES Extensión de Postfiltrado Adaptable para Formador de Haz Superdirectivo.

BF Formador de Haz.

BSS Separación Ciega de Fuentes.

CPSD Densidad Espectral de Potencia Cruzada.

CSD Densidad Espectral Cruzada.

DFT Transformada Discreta de Fourier.

DOA Dirección de Arribo.

DS Retraso y Suma.

DSP Procesador Digital de Señales.

FFT Transformada Rápida de Fourier.

FIBP Formador de Haz Invariante en Frecuencia.

FIR Respuesta Finita al Impulso.

FT Transformada de Fourier.

GSC Cancelador de Lóbulos Laterales Generalizado.

GSVD Descomposición de Valores Singulares Generalizada.

ICA Análisis de Componentes Independientes.

IDFT Transformada Discreta de Fourier Inversa.

IFFT Transformada Rápida de Fourier Inversa.

IIR Respuesta Infinita al Impulso.

LCMV Mínima Varianza Linealmente Restringido.

MMSE Error Cuadrático Medio Mínimo.

MVDR Respuesta sin Distorsión de Mínima Varianza.

MVDR-RS Respuesta sin Distorsión de Mínima Varianza con Remuestreo Espacial.

OLA Traslape y Suma.

PCA Análisis de Componentes Principales.

PDS Procesamiento Digital de Señales.

PHAT Transformada de Fase.

PSD Densidad Espectral de Potencia.

SAR Relación Señal a Artefactos.

SDP Patrón de Dirección Espacial.

SDR Relación Señal a Distorsión.

SIR Relación Señal a Interferencia.

SNR Relación Señal a Ruido.

SPL Nivel de Presión de Sonido.

SRMV Mínima Varianza Remuestreado Espacialmente.

SRP Patrón de Respuesta Direccional.

ULA Arreglo Lineal Uniforme.

UPA Arreglo Planar Uniforme.

WNG Ganancia de Ruido Blanco.

WSS Estacionario en Sentido Amplio.