
# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA ELÉCTRICA – PROCESAMIENTO DIGITAL DE SEÑALES

## ESTIMACIÓN DE LA DIRECCIÓN DEL ÁNGULO DE ARRIBO DE UNA FUENTE DE VOZ.

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

PRESENTA:
MIGUEL ANGEL FLORES GÓMEZ

TUTOR PRINCIPAL:
M.I. LARRY HIPÓLITO ESCOBAR SALGUERO

CDMX, SEPTIEMBRE 2017.

**JURADO ASIGNADO:**

| | |
|---|---|
| Presidente: | Dr. Caleb Rascón Estebané |
| Secretario: | Dr. Pablo Roberto Pérez Alcázar |
| Vocal: | M.I. Larry Hipólito Escobar Salguero |
| 1er. Suplente: | Dr. Jesús Savage Carmona |
| 2do. Suplente: | Dra. Lucía Medina Gómez |

Lugar o lugares donde se realizó la tesis:

Laboratorio de Procesamiento Digital de Señales, Facultad de Ingeniería, UNAM.

**TUTOR DE TESIS:**

M.I. Larry Hipólito Escobar Salguero

---

**FIRMA**

*"Aquel que tiene un porqué para vivir se puede enfrentar a todos los cómos"*

Friederich Nietzsche.

# Dedicatoria

El presente trabajo se lo dedico con todo mi amor y respeto a las personas que han sido mis compañeros y maestros de vida, mis padres Ramón Flores Ávalos y Geogina Gómez Guíza, mis hermanos Ramón Flores Gómez y Marco Antonio Flores Gómez, mi sobrino Emiliano Flores García y mi cuñada Alba García Peña, ya que le dan sentido y alegría a cada instante de la misma.

# Agradecimientos

Agradezco a Dios por brindarme la oportunidad de seguirme preparando y concluir las metas de vida que me he propuesto.

A mi padre Ramón Flores Ávalos por su amor, apoyo incondicional, consejos, ser sensato en sus comentarios y enseñarme que todo es posible desde el amor.

A mi madre Georgina Gómez Guíza, que siempre está en mi corazón, por haberme impulsado el tiempo que estuvo a mi lado y enseñarme a enfrentarme a la vida.

A mis hermanos Ramón Flores y Marco Antonio Flores por su apoyo incondicional, por ser mis compañeros en este viaje y mi ejemplo a seguir.

A mi cuñada Alba Peña y mi sobrino Emiliano Flores por su amor, alegría y risas.

Al maestro Larry H. Escobar Salguero por haberme abierto las puertas de su laboratorio como lugar de trabajo, su apoyo incondicional en el desarrollo del presente trabajo, por su paciencia y asesorías.

Al doctor Caleb Rascón Estebané por haberme apoyado con asesorías y facilitarme la infraestructura de su laboratorio en donde se realizó la adquisición de las señales.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico que me brindaron durante mis estudios de la maestría.

Al proyecto PAPIME PE100616 con el nombre ”Servidor para prácticas de procesamiento digital de señales en tiempo real” por el apoyo en cuanto al equipo utilizado en el laboratorio de procesamiento digital de señales del posgrado de ingeniería, el cuál se utilizó en el presente trabajo.

A mis compañeros del laboratorio de procesamiento digital de señales Luis Álvarez, Yoloxóchil Jiménez, Michel Olvera, Carlos Ignacio García e Iván Menendez por su apoyo, compañía, alegría y recomendaciones durante el tiempo que he estado en el mismo.

A los Frecuencia: Luis Enrique Galván, Armando Vázquez, Jorge Cervantes y Adalberto Gó-

mez por su apoyo, sus risas y locuras, pero sobre todo por caminar conmigo en el mismo sueño.

A las personas que han estado conmigo en las buenas y en las malas ofreciéndome su mano como apoyo: Paulina de Leija, Alberto Fernández, Jeison Méndez, Maribel Venegas, Wendelyn Curvelo, Beatriz Ordoñes, Erika González, Cristóbal Alfonso Santos, Vanesa Sánchez, Liliana Madrid, Ricardo Zúñiga, Edgar Eloy García, Gustavo López, Genaro García, Oscar Pilloni, Laura Oropeza, Erick Omar Morales, Xóchitl Oropeza, Ricardo Ávila, Elizabeth Valencia y Thaell Ontiveros. Simplemente gracias por ser y estar.

# Resumen

En el presente trabajo de tesis se expone el análisis teórico e implementación de un sistema, el cual estima el ángulo de la dirección de arribo (DOA) de las señales emitidas por una fuente de voz ya sea estática o con desplazamiento inmersa en un ambiente con condiciones reales (ambiente acústico no controlado). La implementación de dicho sistema se realizó por medio de la tarjeta de adquisición de señales de voz *8-SoundUSB* en conjunto con el software de audio *Jack Audio Connection kit* que controla el flujo de datos de entrada y salida de las señales adquiridas entre la tarjeta de adquisición y del programa en lenguaje C. El sistema puede efectuar la estimación del DOA ya sea de las señales adquiridas por la tarjeta o por medio de un banco de señales previamente adquiridas con dicha tarjeta, en ambos casos la ejecución del sistema se lleva a cabo en línea con un tiempo de respuesta mínimo de $t = 348$ milisegundos en cada estimación.

Se implementaron dos tipos de arreglos de micrófonos diferentes, un arreglo lineal uniforme y un arreglo circular equidistante, ambos con seis micrófonos, comparando los resultados obtenidos con dichos arreglos.

La estimación del ángulo de arribo se calculó realizando la eigendescomposición de la matriz de covarianza y construyendo la respuesta de dirección promedio a partir de la teoría del formador de haz fijo, utilizando los eigenvectores relacionados con las señales en la entrada de la estructura del formador de haz, de tal manera que se pueda estimar el ángulo azimutal de arribo dentro del intervalo $-90 \leq \theta \leq 90$ grados para el arreglo lineal de micrófonos, y $0 \leq \theta < 359$ grados para el arreglo circular de micrófonos.

Se implementó un clasificador estadístico para asignar los ángulos de arribo, obtenidos a la salida de la respuesta de dirección promedio, a las fuentes de voz correspondientes que fueron seguidas en el tiempo anterior. Dicho clasificador puede generar una nueva clase (fuente seguida) cuando el ángulo de arribo no corresponde a ninguna de las fuentes previamente seguidas.

Finalmente se implementó un Filtro de Kalman para seguir la trayectoria de las fuentes de voz que sufren un desplazamiento durante la emisión de sus señales, describiendo la trayectoria de dicha fuente de voz por medio de las gráficas de seguimiento (ángulo de arribo vs tiempo), histograma y velocidades instantánea y promedio.

En los experimentos realizados se estima la dirección de arribo de las señales provenien-

tes de fuentes de voz, utilizando dos tipos de arreglos de micrófonos, el arreglo lineal uniforme de micrófonos y el arreglo circular uniforme de micrófonos. Durante dichos experimentos, las fuentes de voz tuvieron un comportamiento estático y con desplazamiento, además se utilizaron fuentes de voz ya sean únicas o simultáneas.

El presente trabajo se elaboró en el laboratorio digital de señales de la Facultad de Ingeniería de la UNAM y las pruebas de adquisición de señales se realizaron en el laboratorio de robótica del cuarto piso del IIMAS (Instituto de investigaciones en matemáticas aplicadas y en sistemas), donde se reprodujeron audiolibros por medio de unas bocinas (A225 Dell) localizadas en una dirección de arribo específica en cada experimento, mientras que la adquisición de señales por medio del arreglo de micrófonos conectado a la tarjeta de adquisición de señales de voz mencionada anteriormente.

# Índice general

# Índice de figuras

# Capítulo 1.

# Introducción

El objetivo del Procesamiento digital de señales (PDS) de arreglos de sensores está centrado en la fusión de datos adquiridos por medio de diferentes sensores posicionados en lugares específicos, para extraer características útiles y realizar una tarea de estimación en tiempo y espacio [8], [9].

El procesamiento de señales adquiridas por arreglos de sensores surge a partir de la teoría de arreglos de antenas en el área de comunicaciones para aplicaciones militares, como el radar y el sonar. Posteriormente se comenzaron a implementar las técnicas de procesamiento de señales acústicas por medio de arreglos de micrófonos para desarrollar aplicaciones en las áreas de procesamiento de voz, robótica y navegación.

El estudio de los arreglos de micrófonos se basa principalmente en el procesamiento por medio de los retardos y variaciones de amplitudes de los frentes de onda adquiridos en cada uno de los elementos del arreglo. Los retardos de tiempo de arribo entre micrófonos van a depender del tipo de geometría del arreglo utilizado, la distancia entre micrófonos y la distancia entre la respectiva fuente de señales y el centro del arreglo de micrófonos.

Las aplicaciones más comunes en el uso de arreglos de micrófonos se centran en la reducción de ruido, localización de fuentes de sonido y filtrado espacial, entre otras. La localización de fuentes de sonido (específicamente fuentes de voz) surge a través de la estimación del retardo de tiempo entre un par de micrófonos o un arreglo de los mismos. Este problema es complejo, principalmente porque las señales de voz son no estacionarias, además de la reverberación generada por las características del lugar [10].

Se han desarrollado diferentes algoritmos para la estimación de la dirección de arribo (DOA) de fuentes de voz con distintas técnicas. El método de correlación cruzada generalizada (GCC) es el más común para obtener los retardos de tiempo del arribo de señales entre un par de micrófonos, sin embargo, su respuesta es pobre cuando se trata de un recinto con reverberación [10].

Algunos algoritmos se basan en las características de la matriz de covarianza de las señales adquiridas a través de los micrófonos del arreglo por medio de la eigendescomposición, como es el caso del algoritmo llamado "Clasificación de múltiples señales (MUSIC)" [4]. Este algoritmo es capaz de estimar el número de señales y su respectivo DOA, con la restricción de ser para señales de banda angosta y estimar hasta $M-1$ señales para un arreglo de $M$ micrófonos.

Una metodología muy utilizada es la del formador de haz (BF), la cuál se basa en realizar el filtrado espacial de fuentes de señales, es decir, adquiere la señal proveniente de una dirección mientras que atenúa las señales provenientes desde otras direcciones. La dirección de la señal de interés debe estar dentro del intervalo de apertura de adquisición del arreglo de micrófonos. Dicho intervalo depende del tipo de arreglo de micrófonos utilizado. De manera similar, el formador de haz también se puede utilizar para estimar el ángulo de arribo de una señal. Para efectuar la estimación de DOA utilizando la técnica del formador de haz, se construye el patrón de radiación del formador de haz, el cual muestra la distribución de la energía adquirida por un arreglo de micrófonos en forma espacial, es decir, en un intervalo de apertura de observación [5]. La técnica del formador de haz se centra en señales de banda angosta, lo cual significa que para señales con un ancho de banda amplio se debe ejecutar el algoritmo para cada una de las frecuencias dentro del ancho de banda, lo que dificulta su ejecución para aplicaciones en tiempo real.

## 1.1. Objetivo general

Diseñar e implementar un sistema capaz de estimar en tiempo real el ángulo de arribo de una señal de voz proveniente de una fuente inmersa en ruido ambiental, utilizando un arreglo de micrófonos.

### 1.1.1. Objetivos específicos

1. Estudiar, analizar y simular los diferentes algoritmos de estimación del ángulo de dirección de arribo DOA, para una o múltiples fuentes de voz.

2. Evaluación de los algoritmos de estimación del ángulo de arribo de señales para su implementación en tiempo real.

3. Análisis matemático de arreglos de micrófonos e implementación.

4. Analizar, simular e implementar algoritmos de detección de actividad de voz para arreglos de micrófonos en tiempo real.

5. Implementación del sistema en tiempo real.

## 1.2. Descripción del problema

La audición es un sentido de primordial importancia para los seres humanos porque es una herramienta que funciona como alerta de peligro y comunicación sin necesidad de tener un contacto visual. Además, es capaz de detectar la dirección en la que se localizan múltiples fuentes de sonido, al mismo tiempo, en un intervalo de 0-360 grados dentro de un recinto con ambiente acústico no controlado.

El sistema de audición humano es de tipo binaural, lo que significa que cuando un sonido es adquirido por un escucha, los dos oídos reciben casi la misma información auditiva con diferencias en el tiempo de llegada e intensidad que dependen del plano en donde se localice la fuente sonora. Esto permite que la localización de fuentes de sonido ses estimada a través de la diferencia de tiempo de llegada e intensidades entre las señales capturadas.

Actualmente la estimación de la dirección de arribo de fuentes sonoras ha sido de gran interés en el área de procesamiento de señales, particularmente en procesamiento de voz para aplicaciones de robótica, videoconferencias, seguridad y navegación.

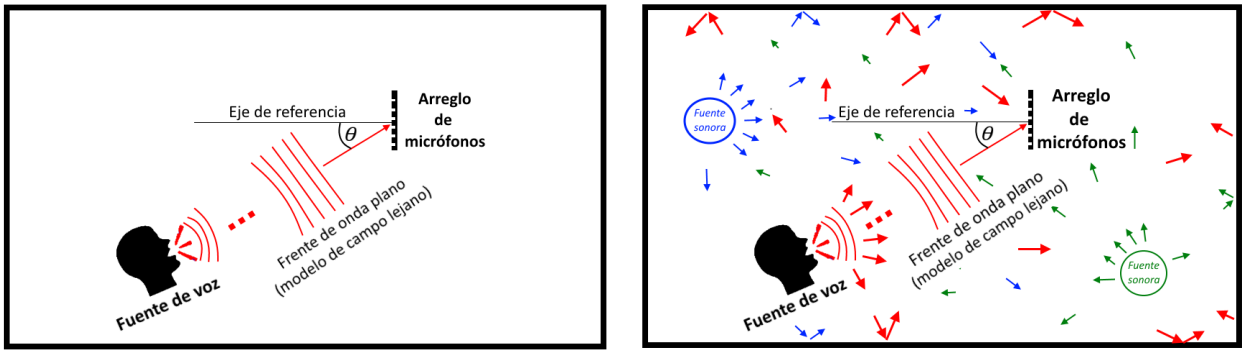
La localización de una fuente sonora, específicamente una fuente de voz, se estima con base a la dirección de arribo de las señales adquiridas por un arreglo de micrófonos con respecto a un eje de referencia. Para realizar la estimación de DOA, se requieren al menos dos micrófonos posicionados de manera estratégica, que detecten el retardo de las señales en cada micrófono.

La localización de una fuente de voz dentro de un recinto se puede llevar a cabo como se observa en la Figura 1.1(a), en donde se muestra que a partir de un arreglo de micrófonos es posible estimar la localización de la fuente de voz a través del ángulo $\theta$, formado entre el eje de referencia del arreglo de micrófonos y la recta normal al frente de onda de la fuente de voz.

La localización de la fuente de voz mostrada en la Figura 1.1(a) es realizada en un recinto de ambiente acústico controlado, como es en el caso de una cámara anecoica, la cual contiene aislamiento acústico, es decir, que no permite que penetren señales acústicas desde el exterior del recinto, además de que existe absorción de la energía radiada por la fuente de voz en el interior.

En el caso de un recinto de ambiente acústico no controlado, como se muestra en la Figura 1.1(b), la estimación de la dirección del ángulo de arribo de una fuente de voz no es tan sencilla,

porque no solamente existe la señal emitida por la fuente de voz, sino que también se encuentran presentes los rebotes de la misma señal, es decir, las señales reflejadas de menor intensidad y con retardo producidas en las paredes del recinto, fenómeno conocido como reverberación. Además, la señal puede ser corrompida por ruido aditivo proveniente de diferentes fuentes de sonido como pueden ser algún instrumento musical, la suma de un número de instrumentos musicales en un ensamble, o simplemente ruido emitido por algún objeto.




(a) DOA de una fuente de voz en un ambiente acústico controlado

(b) DOA de una fuente de voz inmersa en ruido y reverberción

Figura 1.1: Estimación de DOA en un recinto.

En un ambiente acústico no controlado, también pueden estar presentes más de una fuente de voz localizadas en diferentes posiciones del recinto y estar comunicándose entre sí o hablar al mismo tiempo, de tal manera que dificulta la localización de cada una de ellas, a este fenómeno se le conoce como fiesta de *coktail*.

La disyuntiva de los métodos de localización de fuentes de voz es que para frecuencias bajas, la apertura del lóbulo principal en el patrón de radiación del formador de haz es muy amplio, lo que dificulta la exactitud de la localización y al mismo tiempo limita el número de fuentes sonoras que pueden ser localizadas. Además, los métodos de localización de fuentes como lo son el formador de haz y métodos por eigendescomposición de la matriz de correlaciones son importados de la teoría de arreglos de antenas, de tal manera que se enfocan en señales de banda angosta. Es decir, para señales de banda ancha como es el caso de voz, que tiene un ancho de banda de *4 KHz*, se requieren aplicar los métodos para cada una de las frecuencias, produciendo un gran número de operaciones en la ejecución y dificultando su implementación en tiempo real.

De tal manera que la problemática del presente trabajo se centra en estimar el ángulo de arribo en tiempo real de una o múltiples fuentes de voz inmersas en un ambiente acústico no controlado.

## 1.3. Antecedentes y estado del arte

La estimación de la dirección del ángulo de arribo de una fuente de voz surge a partir del área de telecomunicaciones por medio de arreglos de antenas y estudio de señales electromagnéticas. En esta área se han desarrollado diferentes métodos para la estimación de fuentes en el caso de señales de banda angosta con base en los estudios realizados respecto al radar y sonar.

El método de Clasificación de múltiples señales (MUSIC) es un método desarrollado en 1986 por Ralph 0. Schmidt [4], en el cual se realiza la eigendescomposición de la matriz de covarianza de las señales de entrada de banda angosta y, por medio de la selección de los eigenvectores referidos al ruido, verifica la ortogonalidad en cada una de las direcciones posibles del intervalo de apertura de adquisición. Los picos generados en el espectro MUSIC se consideran como posibles direcciones de incidencia de la señal. A partir de la propuesta de este método, surgieron nuevas investigaciones con la idea de expandir el algoritmo para señales de banda ancha, en donde repiten el algoritmo para cada una de las frecuencias y de esta manera aplicarlo a la localización de fuentes de voz [10], [11], [12] lo que propicia que el número de operaciones sea muy grande y un problema para ejecutarlo en tiempo real.

Otras técnicas, como la diferencia del tiempo de arribo (TDOA), se basan en calcular el ángulo de arribo de la señal en relación al retardo entre las señales adquiridas por dos micrófonos, calculado a través de la correlación cruzada [10], [13], [14]. La extensión a tres micrófonos ubicados en forma triangular se aplica para realizar la localización en un intervalo de 0 a 360 grados en [15].

Las técnicas basadas en TDOA son estimadas por variaciones de la correlación cruzada, como la correlación cruzada generalizada (GCC) y en conjunto con la transformada de fase (GCC-PHAT) como es el caso de las investigaciones en [16],[17]. Muchos de los trabajos realizados han sido desarrollados para su implementación en audición robótica para robots de servicios [12], [13], [15], [18] y en sistemas de videoconferencias [19].

Un modelo importante para la estimación del ángulo de arribo de señales a través de arreglos de sensores es el llamado formador de haz (BF), en el cual se calcula la energía direccionando el ángulo del formador de haz. El BF surge con la idea de amplificar la señal focalizada en una dirección de interés mientras atenúa la energía proveniente desde otras direcciones [10], [5], [20]. La extensión del formador de haz para señales de banda ancha se representa tanto en el dominio del espacio y tiempo, como en el dominio de la frecuencia [10], [5], [21].

Por otro lado, existen los métodos adaptables que actualizan los coeficientes asignados a las señales de entrada con base en la diferencia entre la señal deseada y la calculada por medio del algoritmo *least-mean-square* (LMS) [22]. Un método adaptable importante es el llamado respuesta sin distorsión de mínima varianza (MVDR) propuesto por Capon [10], [23]. También se han realizado investigaciones sobre el formador de haz adaptable utilizando la técnica cancelador de lóbulos laterales (GSC) [24], [25].

## 1.4. Justificación

La determinación del ángulo de arribo de una señal de voz es de gran importancia en particular para las áreas de navegación, comunicaciones, videoconferencias, procesamiento de voz y robótica. En el estudio de señales acústicas inmersas en condiciones ambientales, es de gran importancia conocer la procedencia de las señales emitidas por las fuentes, ya que se puede realizar posteriormente una separación de fuentes o un filtrado espacial de una fuente de interés con base en la información proporcionada por el sistema de estimación del ángulo de arribo.

Al hablar de localización de fuentes de voz, implícitamente estamos hablando del análisis de señales por medio de arreglos de micrófonos, por lo que es necesario implementar una geometría adecuada para la detección de retardos entre las señales adquiridas por los micrófonos [26].

Los algoritmos para la estimación del ángulo de arribo son adecuados para la detección de fuentes en ambientes ruidosos, pero con la disyuntiva de que no son robustos para recintos en donde existe una elevada reverberación.

Este proyecto se lleva a cabo con el objetivo de analizar y estudiar el proceso de la dirección de ángulo de arribo y una posible mejora en los métodos de DOA para el caso de señales con un ancho de banda.

## Organización del trabajo

En el Capítulo 2 se explica el desarrollo matemático para la obtención de la ecuación de onda y el modelo de la señal utilizado a lo largo del trabajo. Además, se muestran los diferentes arreglos de micrófonos estudiados y sus correspondientes vectores de dirección. Finalmente, se exponen algunos algoritmos de detección de actividad de voz por medio de la energía de la señal adquirida.

El Capítulo 3 se enfoca en los métodos de estimación de la dirección de arribo de señales de banda angosta y banda ancha. Los métodos de estimación se dividen en tres grupos: los métodos de diferencia del tiempo de arribo (TDOA) por medio de correlaciones y GCC-PHAT, las técnicas basadas en la eigendescomposición de la matriz de covarianza y el formador de haz.

El diseño e implementación del sistema se expone en el Capítulo 4, el cual se basa en la eigendescomposición de la matriz de covarianza formada por las señales de entrada y posteriormente construye la respuesta de dirección con base en la teoría del formador de haz. Posteriormente, se explica la implementación del clasificador estadístico utilizado y el seguimiento de la fuente de voz por medio de un filtro de Kalman.

En el Capítulo 5 se muestran los resultados obtenidos utilizando un arreglo lineal uniforme de seis micrófonos y un arreglo circular de seis micrófonos. En ambos casos se realizaron experi-

mentos tanto con una fuente de voz estática, como con múltiples fuentes simultáneas, además de la estimación de la dirección de arribo de una fuente de voz en movimiento, analizando su desplazamiento con base en sus velocidades instantánea y promedio, además de su histograma.

Finalmente, en el Capítulo 6 se presentan las conclusiones a partir de los resultados obtenidos, explicando los objetivos logrados, observaciones finales y el trabajo a futuro.

# Capítulo 2.

# Arreglos de micrófonos y detector de actividad de voz

El objetivo de las aplicaciones del procesamiento digital de señales se centra en la extracción de los parámetros de las señales presentes en el entorno, tales como señales ambientales (temperatura, presión o humedad), sísmicas, electromagnéticas o acústicas. Una clase importante de dichas señales son las señales acústicas que son producidas por vibraciones de un cuerpo en un intervalo de frecuencia, y que se propagan a través de un medio. Gracias a la percepción del sonido, los seres humanos son capaces de detectar peligro a su alrededor y comunicarse entre sí. Sin embargo, antes de realizar la adquisición de dichas señales por medio de transductores acústicos, es necesario comprender la propagación del sonido desde la fuente emisora hasta el micrófono o arreglo de micrófonos, con el objetivo de determinar el modelo matemático a utilizar; dicho modelo va a variar en función de la posición de los micrófonos del arreglo y la distancia definida entre la fuente emisora y la etapa de adquisición.

El presente capítulo se divide en las siguientes secciones: la primera se centra en el desarrollo matemático de la ecuación de onda a partir de la generación del sonido por una fuente, además se explica el modelo de la señal utilizado y su importancia en el proyecto. De la misma

manera, en la Sección 2.1.3 se muestra el modelo de campo lejano de la señal y su respectiva condición matemática para poder considerarlo. Posteriormente, en la Sección 2.2 se muestran los tipos de arreglos de micrófonos estudiados en el presente trabajo y su respectivo desarrollo matemático para obtener el vector de direcciones que caracteriza a dichos arreglos. Finalmente, en la Sección 2.3 se exponen la detección de actividad de voz y los algoritmos considerados para efectuar la detección de voz por medio de la energía de la señal.

## 2.1. El sonido y su modelo matemático

En general, la definición de sonido es utilizada tanto para describir las ondas físicas que viajan a través del aire, como para explicar la sensación psicológica que causan dentro de nuestro cerebro. Haciendo hincapié en la descripción de ondas mecánicas dentro del área de física, el sonido es definido como vibraciones periódicas o aperiódicas emitidas por un cuerpo, las cuales se propagan a través de un medio elástico, es decir, cada pequeño paquete de partículas de aire vibra de alguna manera y éste transfiere la perturbación a sus partículas vecinas, pero mientras la perturbación se transfiere disminuyendo la energía a distancia, cada paquete de partículas permanece en la vecindad de su posición original.

El intervalo de frecuencias audibles para los seres humanos está comprendido entre los $20Hz$ y $20kHz$, sin embargo, fuera de dicho intervalo también existen señales acústicas utilizadas por otros seres vivos para comunicarse. Estas vibraciones se clasifican como longitudinales, es decir, las partículas se desplazan en la dirección de propagación de las ondas sonoras. El desplazamiento relativo de las partículas en las ondas sonoras generan un pequeño cambio de presión y densidad. Además del aire, todos los gases, líquidos y sólidos son medios elásticos que pueden conducir tipos de vibraciones similares percibidas como sonido, ya sea por un acoplamiento directo al oído humano o tomando el aire como medio de acoplamiento, lo que significa que en el vacío no existe la propagación del sonido.

El sonido se genera cuando el medio se perturba dramáticamente, tal perturbación genera cambios de presión, densidad, velocidad de la partícula y temperatura [27]. En la siguiente sección se obtendrá la ecuación de onda a partir de las variables previamente mencionadas para deducir el modelo de la señal utilizado en el presente trabajo (Sección 2.1.2).

### 2.1.1. Ecuación de onda del sonido

Ya que la naturaleza del sonido requiere regiones de compresión y rarefacción[1], es importante considerar los cambios de densidad en masa, presión, temperatura y la energía interna del fluido. La densidad de masa se puede modelar como se muestra en la Ecuación (2.1),

$$\rho = \rho_0 + \rho_1(x,t) \tag{2.1}$$

donde $\rho_0$ es la constante de densidad en el ambiente en condiciones de equilibrio y $\rho_1$ representa el cambio de densidad provocado por la onda. De manera similar se modelan las variables

[1] Proceso por el que un cuerpo o sustancia se hace menos denso y se contrapone al fenómeno de compresión

de presión, energía y temperatura como se muestran en las Ecuaciones (2.2),(2.3) y (2.4) respectivamente.

$$p = p_0 + p_1(x,t) \tag{2.2}$$

$$e = e_0 + e_1(x,t) \tag{2.3}$$

$$T = T_0 + T_1(x,t) \tag{2.4}$$

Para determinar el modelo matemático de la ecuación de onda, vamos a considerar un cuadrado como una muestra de fluido como el que se muestra en la Figura 2.1, en donde se observa que tiene una posición inicial $x$ y posteriormente sufre un desplazamiento ($\xi(x,t)$).

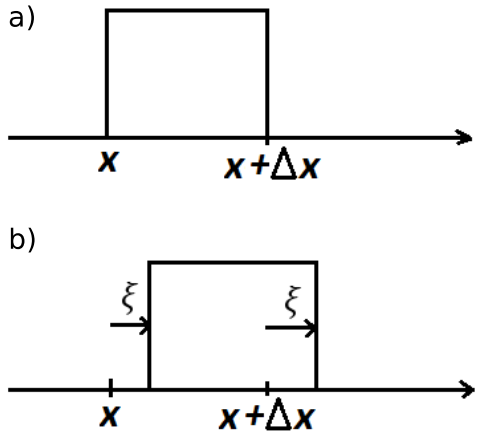



Figura 2.1: Cuadrado de fluido que sufre un cambio de densidad provocado por la onda de sonido, sufriendo un desplazamiento $\xi(x)$ por la cara izquierda del cubo y $\xi(x+\Delta x)$ el desplazamiento por la cara derecha del mismo [1].

Primeramente, obtenemos la relación que existe entre el desplazamiento y la densidad. Es necesario tomar en cuenta que si cada elemento del fluido tiene el mismo desplazamiento, no debería de haber cambio en la densidad. Si consideramos que el movimiento es únicamente sobre el eje x, la cara izquierda del cubo de la Figura 2.1 se desplaza bajo la influencia de la onda de sonido desde la posición $x$ hasta $x+\xi(x,t)$, mientras que la cara derecha del cubo se mueve de $x+\Delta x$ a $x+\Delta x+\xi(x+\Delta x,t)$, como se observa en la Figura 2.1. El ancho del cubo cambia su longitud desde un valor inicial $\Delta x$ hasta $\Delta x+\xi(x+\Delta x,t)-\xi(x,t)$, de tal manera que el volumen del cubo se multiplica por el siguiente factor:

$$1+\frac{\xi(x+\Delta x,t)-\xi(x,t)}{\Delta x} \tag{2.5}$$

sin embargo en el límite de un $\Delta x$ pequeño, éste es simplemente $1+\frac{\partial \xi}{\partial x}$ [1], [27].

Sabemos que el sonido se propaga en todas las direcciones y no solamente en el eje x como se explicó anteriormente, de tal manera que ahora vamos a considerar que la posición x está definida por un vector de tres dimensiones $(x,y,z)$, y su respectivo desplazamiento $\xi$, de la misma

manera, será un vector con componentes $\xi x$, $\xi y$ y $\xi z$, es decir que el cuadrado de la Figura 2.1 se debe de expandir a tres dimensiones, de manera que el volumen del cubo se puede expresar como se muestra en las Ecuaciones (2.6) y (2.7).

$$\Delta x\Delta y\Delta z\left(1+\frac{\partial \xi_x}{\partial x}\right)\left(1+\frac{\partial \xi_y}{\partial y}\right)\left(1+\frac{\partial \xi_z}{\partial z}\right) \simeq \Delta x\Delta y\Delta z\left(1+\frac{\partial \xi_x}{\partial x}+\frac{\partial \xi_y}{\partial y}+\frac{\partial \xi_z}{\partial z}\right) \tag{2.6}$$

$$\Delta x\Delta y\Delta z\left(1+\frac{\partial \xi_x}{\partial x}+\frac{\partial \xi_y}{\partial y}+\frac{\partial \xi_z}{\partial z}\right) = \Delta x\Delta y\Delta z\ (1+\bigtriangledown\cdot\xi) \tag{2.7}$$

Para ondas de amplitud suficientemente pequeñas, los términos cruzados de la Ecuación (2.6), como por ejemplo $\left(\frac{\partial \xi_x}{\partial x}\right)\left(\frac{\partial \xi_y}{\partial y}\right)$ o $\left(\frac{\partial \xi_y}{\partial y}\right)\left(\frac{\partial \xi_z}{\partial z}\right)$, pueden ser descartados si todas las derivadas de $\xi$ son suficientemente pequeñas comparadas con la unidad [1]. A pesar de que las dimensiones del cubo han cambiado, la masa es la misma, es decir que la densidad debió de haber decrementado con la misma relación que el volumen incrementó, esto es:

$$\rho_0+\rho_1=\frac{\rho_0}{1+\bigtriangledown\cdot\xi} \tag{2.8}$$

Finalmente, se puede obtener la Ecuación (2.9) tomando en cuenta que cuando una cantidad $\varsigma$ es muy pequeña comparada con 1, la expresión $\frac{1}{1+\varsigma}$ se puede aproximar por $1-\varsigma$ utilizando una expansión binomial

$$\rho_1(x,t)\cong -\rho_0\bigtriangledown\cdot\xi \tag{2.9}$$

La segunda ley de Newton explica que cuando una fuerza se aplica a un cuerpo, ya sea en reposo o en movimiento, la fuerza cambia el estado de movimiento del cuerpo experimentando una aceleración en dirección a la fuerza aplicada, es decir, $\mathbf{F}=m\mathbf{a}$. Aplicando la teoría de la segunda ley de Newton a una fracción de fluido como en el cubo de la Figura 2.1, el fluido del lado izquierdo empuja al derecho en dirección x con presión $p_0+p_1(x+\xi_x(x,t),t)$ sobre una superficie de área $\Delta y\Delta z$, mientras que el fluido derecho presiona en sentido contrario con una presión ligeramente diferente $p_0+p_1(x+\Delta x+\xi_x(x+\Delta x,t))$. Por lo que la fuerza neta aplicada en la dirección $x$ está dada por la Ecuación (2.10)

$$F_x=\Delta y\Delta z\left[p_1(x,t)-p_1(x+\Delta x,t)\right] \tag{2.10}$$

Teniendo en cuenta que la definición de la derivada es descrita por el cociente de Newton $\frac{f(x+\Delta x)-f(x)}{\Delta x}$, la Ecuación (2.10) se puede expresar como se muestra en (2.11)

$$F_x=\Delta y\Delta z\left(-\frac{\partial p_1}{\partial x}\right)\Delta x\,, \tag{2.11}$$

la fuerza en las componentes $y$ y $z$ se pueden expresar de manera similar a la Ecuación (2.11), por lo que el vector de fuerza neta aplicada es:

$$\mathbf{F}=-\Delta x\Delta y\Delta z\bigtriangledown p_1 \tag{2.12}$$

Finalmente, la segunda ley de Newton la podemos expresar por medio de la ecuación (2.13), tomando en cuenta que la masa de la porción del fluido es $\rho_0\Delta x\Delta y\Delta z$.

$$-\nabla p_1 = \rho_0 \frac{\partial^2 \xi}{\partial^2 t} \tag{2.13}$$

La presión $p$ se determina únicamente como una función de la densidad $\rho$. Entonces la derivada $\left(\frac{dp}{d\rho}\right)_{ad}$ es una derivada ordinaria y bien definida que al ser evaluada en $\rho = \rho_0$, el resultado es simplemente un número constante. Por lo que una expansión de la serie de Taylor nos indica que las pequeñas desviaciones del estado de equilibrio va a obedecer la Ecuación (2.14) [1], [27]

$$p \simeq p\left(\rho_0\right) + \left(\rho - \rho_0\right) \frac{dp}{d\rho}\,, \tag{2.14}$$

además $p = p_0 + p_1$, entonces:

$$p1 \simeq \left(\frac{dp}{d\rho}\right)_{ad} \rho_1 \tag{2.15}$$

Sustituyendo la Ecuación (2.9) en (2.15) se obtiene la Ecuación (2.16).

$$p_1 = -p_0 \left(\frac{dp}{d\rho}\right)_a \nabla \cdot \xi \tag{2.16}$$

Además, calculando la divergencia de la Ecuación (2.13) y la segunda derivada con respecto al tiempo de la Ecuación (2.16), se obtienen las siguientes expresiones:

$$\nabla^2 p_1 = -\rho_0 \frac{\partial^2 \left(\nabla \cdot \xi\right)}{\partial^2 t} \tag{2.17}$$

$$\frac{\frac{\partial^2 p_1}{\partial^2 t}}{\left(\frac{dp}{d\rho}\right)_{ad}} = -\rho_0 \frac{\partial^2 \left(\nabla \cdot \xi\right)}{\partial^2 t} \tag{2.18}$$

Igualando las ecuaciones (2.17) y (2.18), se obtiene la Ecuación (2.19).

$$\nabla^2 p_1 = \frac{\frac{\partial^2 p_1}{\partial^2 t}}{\left(\frac{dp}{d\rho}\right)_{ad}} \tag{2.19}$$

Considerando que $c^2 = \left(\frac{dp}{d\rho}\right)_{ad}$, la ecuación de onda se expresa en la Ecuación (2.20)

$$\frac{\partial^2 p}{\partial^2 t} = c^2 \nabla^2 p\,, \tag{2.20}$$

donde $c$ se conoce como la velocidad de propagación del sonido. Se puede observar de la Ecuación (2.20) que se ha eliminado el subíndice $p_1$ para la parte acústica de la presión, esto es por simplicidad de notación y además para cualquier tiempo $t$ nos podremos referir a la presión total (Ecuación (2.2)).

Para tener una mejor representación de la posición de una fuente de sonido puntual, el vector de posición $x = \left(x, y, z\right)$ lo vamos a representar en coordenadas polares en lugar de cartesianas como se muestra en la Ecuación (2.21).

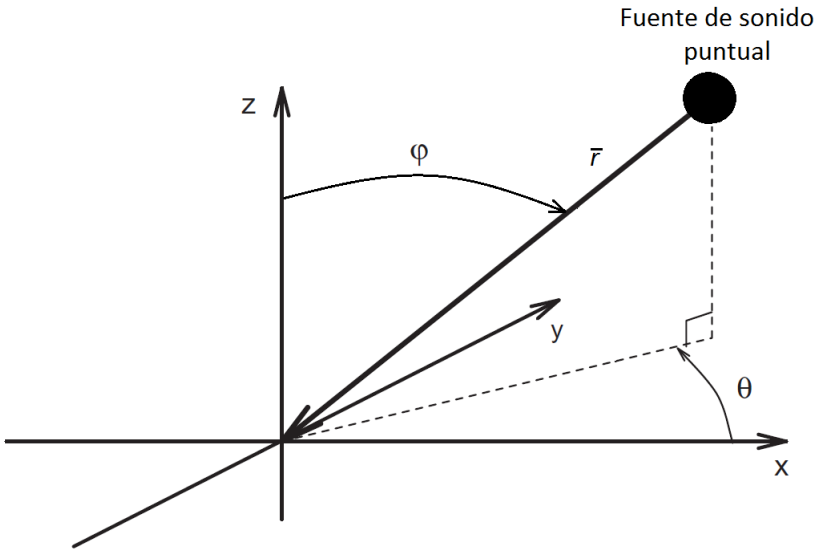



Figura 2.2: Posición de una fuente de sonido puntual [2].

$$\mathbf{r} = \left(r\cos\phi\sin\theta, r\sin\phi\sin\theta, r\cos\phi\right). \tag{2.21}$$

donde los ángulos $\theta$ y $\phi$ son el ángulo de azimuth y de elevación respectivamente y $r$ es la distancia entre la fuente en la posición $\mathbf{r}$ y el punto de observación como se muestra en la Figura 2.2, tomando en cuenta que se considera una fuente puntual.

El operador Laplaciano de la Ecuación (2.20), se puede expresar en coordenadas polares como se muestra en (2.22)

$$\bigtriangledown^2 p = \left(\frac{1}{r^2}\right)\left[\frac{\partial\left(r^2\left(\frac{\partial p}{\partial r}\right)\right)}{\partial r} + \left(\frac{1}{\sin\phi}\right)\frac{\partial\left(\sin\phi\left(\frac{\partial p}{\partial \phi}\right)\right)}{\partial \phi} + \left(\frac{1}{\sin^2\phi}\right)\frac{\partial^2 p}{\partial \theta^2}\right] \tag{2.22}$$

Las soluciones de la ecuación de onda dependen solamente de la distancia al emisor y no de la dirección. Para considerar la dirección de la fuente vamos a utilizar el Laplaciano de la Ecuación (2.22) en la ecuación de onda (2.20) obteniendo la llamada ecuación de onda esférica que se muestra en la Ecuación (2.23) [1], [2]

$$\frac{\partial^2 p}{\partial t^2} = \left(\frac{c^2}{r^2}\right)\frac{\partial\left(r^2\partial p/\partial r\right)}{\partial r} \tag{2.23}$$

Suponiendo un pequeño arreglo de sensores situados en una región lejana a la fuente, éstos pueden adquirir la información de la señal de la forma:

$$p\left(\mathbf{r}, t\right) = s\left(t - r/c\right) \tag{2.24}$$

por lo que vamos a definir a $s\left(t\right)$ como la señal propagada situada en el arreglo de micrófonos.

Un caso particular y de interés es una señal monocromática, es decir que contiene una frecuencia pura como se muestra a continuación:

$$p(\mathbf{r}, t) = Ae^{j\omega(t-\mathbf{r}\cdot\mathbf{u_r}/c)} = Ae^{j(\omega t-\mathbf{r}\cdot\mathbf{k})} \tag{2.25}$$

donde $\mathbf{u_r}$ es un vector unitario que apunta en la dirección de propagación de la fuente, $\mathbf{r}$ es la posición de observación y $\mathbf{k} = k\mathbf{u_r}$ es conocido como el vector de onda definido por la Ecuación (2.26)

$$\mathbf{k} = \begin{bmatrix} k_x \\ k_y \\ kz \end{bmatrix} = -k \begin{bmatrix} \sin\phi\cos\theta \\ \sin\phi\sin\theta \\ \cos\phi \end{bmatrix} \tag{2.26}$$

y $k$ es el número de onda definido como

$$k = \frac{\omega}{c}\,. \tag{2.27}$$

### 2.1.2. Modelo de la señal

Suponiendo que la señal de banda angosta con frecuencia $\omega$ de la Ecuación (2.25) se adquiere en diferentes posiciones de interés $\mathbf{r_m}$, para $m = 1, 2, \ldots, M$, y las amplitudes $A(t - r_m/c) \simeq A(t - r/c)$, la representación de la señal en la m-ésima posición está dada por:

$$p(\mathbf{r_m}, t) = A(t - r/c)\, e^{j(\omega t-\mathbf{k}\cdot\mathbf{r_m})} \tag{2.28}$$

donde $r$ es la distancia entre la fuente y la posición de medición, $c$ la velocidad de propagación del sonido ($340\, m/s$) y $k$ el número de onda. Si ahora en cada posición de medición colocamos un micrófono con su respectivo sistema de adquisición, la señal adquirida por el m-ésimo micrófono estará dada por

$$x_m = e^{-j\mathbf{k}\cdot\mathbf{r_m}} s(t) \tag{2.29}$$

donde $s(t) = A(t - r/c)$ define la amplitud de la señal. De esta manera en el caso de banda angosta, las salidas de los sensores son coherentes y solamente difieren en el cambio de fase [2]. La señal modelada en (2.29) puede ser representada a través de un vector de dimensión M, denominando la salida del arreglo de micrófonos en la Ecuación (2.30)

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ \cdot \\ x_M(t) \end{bmatrix} = \mathbf{a}(\theta, \phi)\, s(t) \tag{2.30}$$

donde $\mathbf{a}(\theta,\phi)$ se denomina como el vector de direcciones (*steering vector*) o vector de respuesta del arreglo en $\theta$ y $\phi$ (ángulos azimut y de elevación respectivamente), definido como:

$$\mathbf{a}\left(\theta,\phi\right) = \begin{bmatrix} e^{-jk(\theta,\phi)r_1} \\ e^{-jk(\theta,\phi)r_2} \\ \cdot \\ \cdot \\ \cdot \\ e^{-jk(\theta,\phi)r_M} \end{bmatrix} \tag{2.31}$$

El vector de direcciones modela los desplazamientos de fase relativos de una señal en las diversas posiciones de los elementos de un arreglo de micrófonos. En algunos casos de arreglos de micrófonos, como el arreglo lineal, solamente se utiliza el modelo con el ángulo de azimut ($\theta$), es decir, que $\phi = 90$, por lo que es común utilizar la notación para el vector de direcciones como $\mathbf{a}(\theta)$.

Para el caso de $P$ señales simultáneas, podemos aplicar superposición, esto es:

$$\mathbf{x}(t) = \sum_{p=1}^{P} \mathbf{a}\left(\theta_p\right) s_p(t) + \mathbf{n}(t) \tag{2.32}$$

donde $\theta_p$ y $s_p(t)$ representan el ángulo de arribo (DOA) y la envolvente compleja de la señal $p$ respectivamente, y $\mathbf{n}(t)$ el vector de ruido aditivo. La expresión (2.32) se puede modelar en forma vectorial a partir de la matriz $\mathbf{A}(\theta) = \begin{bmatrix} \mathbf{a}(\theta_1) & \mathbf{a}(\theta_2) & \cdots & \mathbf{a}(\theta_P) \end{bmatrix}$, con $\theta = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_P \end{bmatrix}^T$ y el vector de señales $s(t) = \begin{bmatrix} s_1(t) & s_2(t) & \cdots & s_P(t) \end{bmatrix}^T$ como:

$$\mathbf{x}(t) = \mathbf{A}(\theta)\mathbf{s}(t) + \mathbf{n}(t) \tag{2.33}$$

La estimación se basa en un número finito de muestras $N$ de $x(t)$ tomado por cada tiempo $t_s$ determinado por la frecuencia de Nyquist, por lo que la representación en tiempo discreto de $x(t)$ esta dada como:

$$\mathbf{x}(n) = \mathbf{A}(\theta)\mathbf{s}(n) + \mathbf{n}(n), \quad para \quad n = 1,2...N. \tag{2.34}$$

La matriz de covarianza de la salida del arreglo de micrófonos se define por la Ecuación (2.35)

$$\mathbf{R_{xx}} = E\left[\mathbf{x}(n)\mathbf{x}^H(n)\right] \tag{2.35}$$

donde la expresión $E(\cdot)$ define la esperanza matemática y $(\cdot)^H$ el complejo conjugado transpuesto, conocido como hermitiano. A partir de las ecuaciones (2.34) y (2.35) se observa que es posible realizar la estimación de las matrices de covarianza de las señales emitidas por la fuente y el ruido como se observa en las ecuaciones (2.36) y (2.37) respectivamente

$$\mathbf{R_{ss}} = E[\mathbf{s}(n)\mathbf{s}^H(n)] \tag{2.36}$$

$$\mathbf{R_{nn}} = E[\mathbf{n}(n)\mathbf{n}^H(n)] \tag{2.37}$$

El ruido usualmente es modelado como blanco, de tal manera que $\mathbf{R_{nn}} = \sigma_n^2\mathbf{I}$, donde $\sigma_n^2$ es la potencia del ruido y $\mathbf{I}$ es una matriz identidad de tamaño $M\ X\ M$. Finalmente la Ecuación (2.35) se puede expresar en términos de (2.36) y (2.37) de la siguiente manera:

$$\mathbf{R_{xx}} = \mathbf{A}(\theta)\mathbf{R_{ss}}\mathbf{A}(\theta) + \mathbf{R_{nn}} \tag{2.38}$$

En la práctica, la matriz de covarianza $\mathbf{R_{xx}}$ se puede estimar por medio de la Ecuación (2.39)

$$\hat{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n) \tag{2.39}$$

de tal manera que cuando $N \longrightarrow \infty$, $\hat{\mathbf{R}}_{\mathbf{xx}} \longrightarrow \mathbf{R_{xx}}$.

### 2.1.3. Modelo de campo lejano

Las señales emitidas por fuentes de sonido tienen un frente de onda de tipo esférico con el objetivo de propagarse en cualquier dirección, es decir, cuando el arreglo de micrófonos se localiza a una distancia corta con respecto a la posición de la fuente, los retardos de las señales en cada uno de los micrófonos no son equidistantes, de la misma manera ocurre con las amplitudes en cada una de las señales adquiridas por los micrófonos, por lo tanto, se debe considerar en el modelo matemático de cada micrófono.

En el presente trabajo se tiene la consideración de que las señales emitidas por la fuente de voz presentan un frente de onda plano cuando llegan al arreglo de micrófonos, es decir, que se considera el modelo de campo lejano. En este modelo, la fuente de voz se localiza a una distancia suficientemente lejos comparada con la longitud del arreglo, de tal manera que el frente de onda que incide con el arreglo es una región local del frente de onda que puede considerarse de forma plana. Esto implica que las variaciones de las amplitudes de las señales en los micrófonos sean aproximadamente nulas y que los retardos de las mismas sean equidistantes [21]. En la Figura 2.3 se muestra un ejemplo del modelo de campo lejano con un arreglo lineal de micrófonos.

El modelo de campo lejano se cumple considerando la Ecuación (2.40), la cual describe la distancia de Fraunhofer [28], [29].

$$r > \frac{2d_{total}^2}{\lambda}, \tag{2.40}$$

donde $r$ es la distancia radial entre el centro de la fuente y el arreglo de micrófonos, $d_{total}$ es la longitud total del arreglo y $\lambda$ la longitud de onda. Sin embargo, dicha distancia no es completamente aplicable para arreglos de micrófonos ya que el modelo matemático describe dicho fenómeno con señales electromagnéticas, de tal manera que $d_{total}$ considera el tamaño de la antena [28]. Para considerar el modelo de campo lejano dentro del entorno de arreglos de micrófonos, los requisitos más aceptables son los siguientes:

$$r >> d_{total}$$
$$r >> \lambda_{max}$$

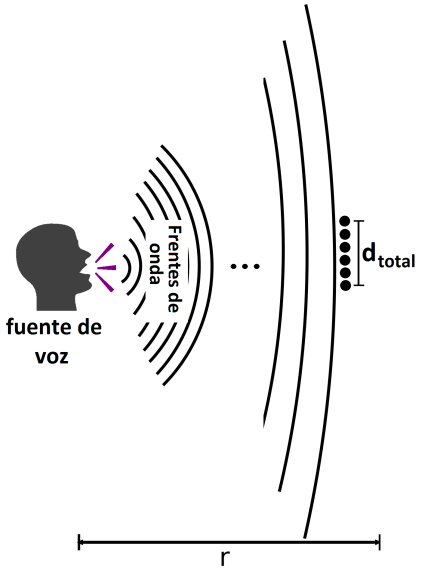



Figura 2.3: Modelo de campo lejano.

donde $\lambda_{max}$ es la longitud de onda mayor que existe en la señal de origen. De tal manera que la distancia $r$ debe ser al menos 10 veces mayor a la longitud del arreglo [28].

La longitud de onda está definida como :

$$\lambda = \frac{c}{f} \tag{2.41}$$

donde $f$ es la frecuencia de oscilación y $c$ la velocidad del sonido, misma que se puede calcular como:

$$c = 331.57 + 0.607T \tag{2.42}$$

donde $T$ es la temperatura del medio de propagación del sonido [14]. La velocidad de propagación del sonido en este trabajo se considera constante a una razón de 340 $m/s$.

## 2.2. Arreglos de micrófonos

Las señales de voz adquiridas por un micrófono pueden ser fácilmente corrompidas por ruido aditivo y reverberación producida por el reflejo de las señales en las paredes del recinto. Un método para reducir las señales de distorsión y enfatizar la calidad de la señal de interés es utilizar múltiples micrófonos localizados en el espacio de forma particular [8].

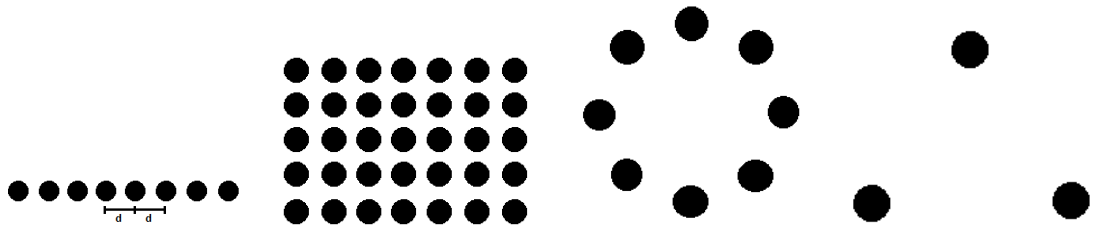
Un arreglo de sensores es un conjunto de transductores localizados en el espacio de manera específica, en donde la localización de los mismos depende de la aplicación de interés. El procesamiento de un arreglo de micrófonos tiene como objetivo la extracción de parámetros de las señales adquiridas o la extracción de las señales de interés (filtrado espacial). Por ejemplo, se pueden detectar los retardos y variaciones de amplitudes existentes entre las señales adquiridas por cada uno de los sensores y por medio de éstos, es posible construir el patrón de radiación,

herramienta que sirve para modelar la distribución de la energía adquirida por el arreglo de micrófonos en un intervalo de apertura espacial para un instante de tiempo *t*, y manipular la orientación del arreglo en forma electrónica.

Existen diferentes áreas de desarrollo en las que su base de investigación surge a partir de un arreglo de micrófonos, las principales son [10]:

1. Reducción de ruido
2. Reverberación
3. Localización de una fuente de sonido
4. Estimación del número de fuentes
5. Localización de múltiples fuentes de sonido
6. Separación de fuentes
7. Fiesta de cóctel

Los arreglos de micrófonos se clasifican con base en la geometría formada por sus elementos. Existen tres grupos, el arreglo de una dimensión (arreglo lineal), arreglos de dos dimensiones (arreglo planar) y arreglos de tres dimensiones [26]. En la Figura 2.4 se muestran algunos ejemplos de arreglos de micrófonos de una y dos dimensiones.




(a) Arreglo lineal (b) Arreglo rectangular (c) Arreglo circular (d) Arreglo triangular

Figura 2.4: Arreglos de micrófonos.

La geometría del arreglo de micrófonos depende de la aplicación, ya que algunas veces las geometrías regulares simplifican el problema de estimación, mientras que la cantidad de micrófonos utilizados va a depender de la capacidad del sistema de adquisición, es decir, se requiere un dispositivo que sea capaz de adquirir las señales provenientes de todos los sensores con su respectiva etapa de conversión analógica a digital (ADC). Además, entre mayor es el número de micrófonos, mayor es el número de operaciones en la etapa de procesamiento lo que dificulta su implementación para tiempo real. En las Figuras 2.4(a) y 2.4(b) se observa que el arreglo rectangular es una extensión de un arreglo lineal de micrófonos, sin embargo con el arreglo rectangular es posible determinar la posición de una fuente en función del ángulo azimutal $\theta$ y el

ángulo de elevación $\phi$ [26].

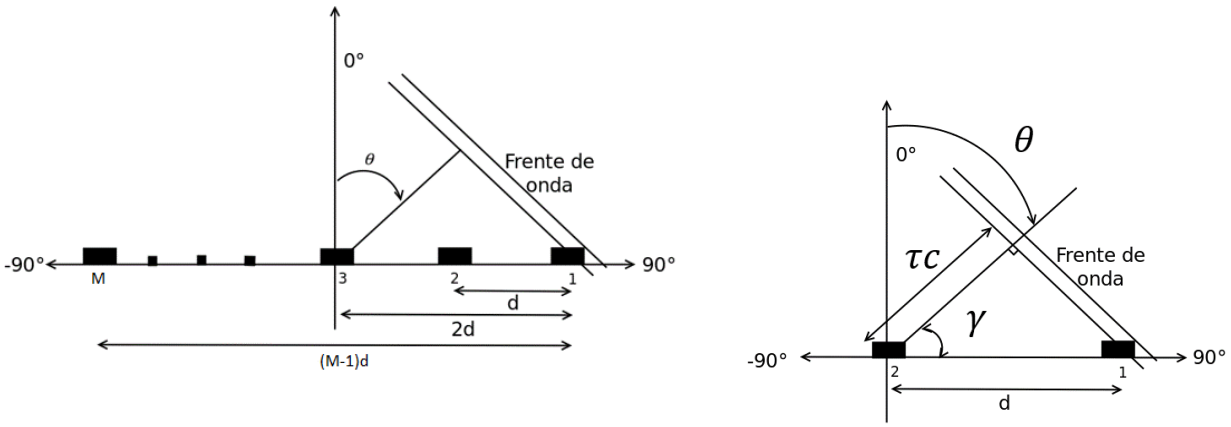
La importancia de cada geometría se centra en los retardos de tiempo de llegada de la señal en cada micrófono, lo que significa que se debe realizar un modelo matemático para cada elemento del arreglo y representarlo por medio del vector de direcciones previamente definido en la Ecuación (2.31). Cabe destacar que los arreglos de micrófonos presentados anteriormente se pueden colocar tanto de forma horizontal como vertical, sin embargo, en cada caso el modelo matemático es diferente.

En los apartados 2.2.1 y 2.2.2 se explican los modelos matemáticos de los arreglos de micrófonos estudiados en el presente trabajo.

### 2.2.1. Arreglo lineal uniforme de micrófonos

Un arreglo lineal de micrófonos se caracteriza porque sus $M$ elementos se localizan sobre una línea recta, separados entre sí con distancia $d$, como se observa en la Figura 2.5(a).

La característica principal del arreglo lineal uniforme de micrófonos, considerando el modelo de campo lejano, es que los retardos de las señales adquiridas por los micrófonos son equidistantes; además, todos los retardos son relacionados con respecto al micrófono de referencia, como se muestra en la Figura 2.5(a). Se observa que el ángulo de arribo $\theta$, es el ángulo que se forma entre el eje de referencia en cero grados y la recta de dirección normal al frente de onda.




(a) Arreglo lineal uniforme de M micrófonos

(b) Retardo de tiempo y DOA entre dos micrófonos

Figura 2.5: Arreglo lineal uniforme de micrófonos.

Para calcular el tiempo de retardo $\tau$ de la señal entre dos micrófonos, vamos a considerar un arreglo de dos micrófonos como el que se muestra en la Figura 2.5(b), donde existe un triángulo rectángulo formado a través del cateto opuesto del ángulo $\gamma$, definido como el frente de onda de la señal, el cateto adyacente que es la distancia definida desde el frente de onda cuando la señal llega al micrófono 1 hasta el micrófono 2, y su hipotenusa que está definida por la distancia

entre el par de micrófonos. A partir de estos elementos se puede obtener la Ecuación (2.43)

$$\cos\gamma = \frac{\tau c}{d} \tag{2.43}$$

Además sabemos que la relación de los ángulos $\theta$ y $\gamma$ es $\theta = 90 - \gamma$, por lo que al expresar la Ecuación (2.43) en función del ángulo $\theta$ y en términos de la función seno, el retardo entre un par de micrófonos se expresa a continuación

$$\tau = \frac{d\sin\theta}{c} \tag{2.44}$$

Finalmente, al extender la Ecuación (2.44), para obtener el retardo en cualquier elemento $m$ de un arreglo lineal uniforme de $M$ micrófonos, se obtiene la Ecuación (2.45).

$$\tau_m = \frac{(m-1)d\sin\theta}{c}, \quad para \quad m = 1,2,3,...,M. \tag{2.45}$$

El arreglo lineal uniforme se caracteriza por realizar la estimación únicamente con base en el ángulo de azimuth $\theta$, esto es debido a que los micrófonos se ubican espacialmente en una sola dirección, lo que significa que el ángulo de elevación es $\phi = 90$ (ver Figura 2.2) en el vector de direcciones de la Ecuación (2.31), es decir, se considera que la fuente de voz se encuentra en el mismo plano que el arreglo de micrófonos. Tomando en cuenta lo anterior, el vector de direcciones que representa a un arreglo lineal uniforme de micrófonos se muestra en la Ecuación (2.46)

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{-ik\tau_1(\theta)} & e^{-ik\tau_2(\theta)} & e^{-ik\tau_3(\theta)} & \cdots & e^{-ik\tau_M(\theta)} \end{bmatrix}^T \tag{2.46}$$

donde $k$ es el número de onda definido en la Ecuación (2.27) y el respectivo retardo de cada micrófono $\tau_m(\theta)$ en función del ángulo de direccionamiento. De tal manera que al sustituir la Ecuación (2.45) en (2.46), el vector de direcciones para un arreglo lineal uniforme de micrófonos se representa en la Ecuación (2.47)

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{-ikdsin(\theta)} & e^{-ik2dsin(\theta)} & e^{-ik3dsin(\theta)} & \cdots & e^{-ik(M-1)dsin(\theta)} \end{bmatrix}^T. \tag{2.47}$$

### 2.2.2. Arreglo planar de micrófonos

Un arreglo planar de micrófonos contiene sus elementos distribuidos sobre un plano. Por lo regular los arreglos de micrófonos planares suelen tener una distancia equidistante $d$ entre sus elementos. Los arreglos planares pueden tener diferentes geometrías y el tiempo de retardo de las señales adquiridas va a depender de la distribución de sus elementos, por lo que ya no serán uniformes en todos los micrófonos como en el caso del arreglo lineal.

Las geometrías más utilizadas en arreglos planares son:

1. Arreglo rectangular
2. Arreglo triangular equidistante
3. Arreglo circular uniforme

y la selección de la geometría del arreglo planar depende del tipo de estimación que se realice. El arreglo rectangular, como se mencionó anteriormente, es una extensión del arreglo lineal de micrófonos, la diferencia es que con el arreglo rectangular también se puede estimar el ángulo de elevación $\phi$. Sin embargo, en el presente trabajo se desea estimar únicamente la estimación del ángulo azimutal $\theta$, además que para mantener el lóbulo principal del patrón de radiación lo más angosto posible se requiere adquirir las señales de un mayor número de micrófonos con respecto al arreglo lineal lo que implica mayor costo computacional dificultando su implementación en tiempo real.

**Arreglo triangular equidistante de micrófonos**

Un arreglo triangular equidistante tiene tres micrófonos localizados en los extremos de un triángulo equilátero separados a una distancia $d$ como se muestra en la Figura 2.6.

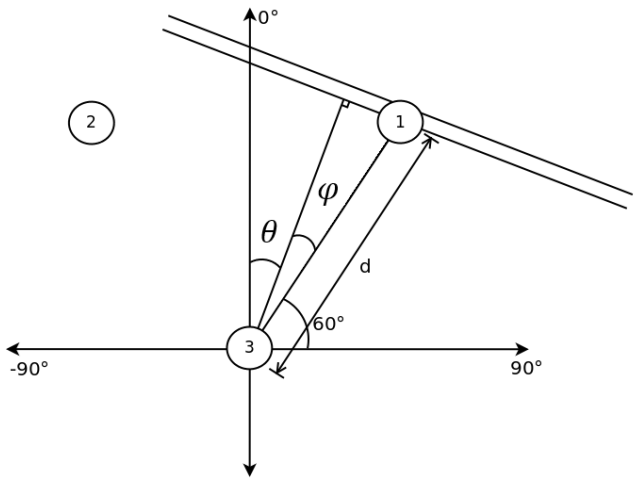



Figura 2.6: Arreglo triangular equidistante de micrófonos.

En el arreglo triangular se pueden considerar dos retardos de tiempo de arribo del frente de onda de la señal $\tau_1$ y $\tau_2$, que son los respectivos retardos en los micrófonos dos y tres, comparados con el micrófono de referencia (micrófono uno). Si tomamos el marco de referencia de la Figura 2.6, el retardo $\tau_1$ (retardo entre los micrófonos uno y dos), se calcula de la misma manera que en la Ecuación (2.44), mientras que para el retardo de tiempo $\tau_2$ (retardo entre los micrófonos uno y tres) se calcula por:

$$\cos\varphi = \frac{\tau_2 c}{d}\,. \tag{2.48}$$

De la misma manera que en el retardo de tiempo $\tau_1$, existe una relación entre los ángulos $\varphi$ y $\theta$ de $\varphi = 30-\theta$, por lo que al sustituir la relación de ángulos en (2.48) y expresando en términos de la función seno, el retardo de tiempo del micrófono tres se expresa en la Ecuación (2.49).

$$\tau_2 = \frac{d\sin(60+\theta)}{c} \tag{2.49}$$

Una característica de este tipo de arreglo de micrófonos es que se puede estimar el DOA en un intervalo de $0 \leqslant \theta < 360$ grados, mientras que en un arreglo lineal uniforme de micrófonos el intervalo de apertura para estimación es de $-90 \leqslant \theta \leqslant 90$ grados. La desventaja del arreglo triangular es que el lóbulo principal del patrón de radiación es ancho y no permite realizar la estimación de DOA de manera fina utilizando el formador de haz.

El vector de direcciones de un arreglo triangular equidistante de tres micrófonos como el de la Figura 2.6, se calcula como:

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{-ikd\sin(\theta)} & e^{-ikd\sin(60+\theta)} \end{bmatrix}^T . \tag{2.50}$$

El retardo entre los micrófonos dos y tres se muestra en la Ecuación (2.51)

$$\tau_3 = \frac{d\sin(\theta - 30)}{c} \tag{2.51}$$

**Arreglo circular uniforme**

Los $M$ elementos de un arreglo circular de micrófonos se localizan sobre un plano cartesiano de la forma $(r\cos\theta_m, r\sin\theta_m)$, donde $\theta_m = \frac{2\pi(m-1)}{M}$ es la posición angular de cada micrófono, mientras que el centro del arreglo de micrófonos se encuentra en el origen del marco de referencia como se muestra en la Figura 2.7 [8], [30].

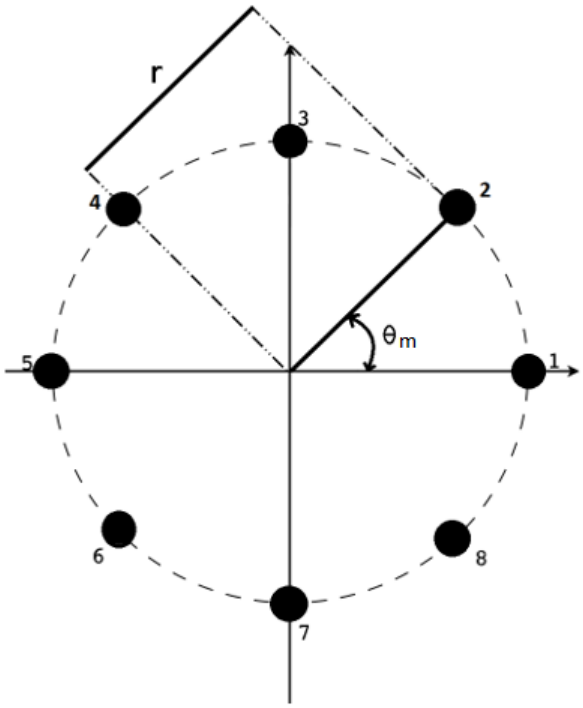



Figura 2.7: Arreglo circular de micrófonos.

En la Figura 2.7 se muestra un arreglo circular de 8 micrófonos posicionados de manera equidistante. A diferencia del arreglo lineal uniforme de micrófonos, los retardos de tiempo de

cada elemento dependen de su posición específica, es decir que cada uno tendrá una relación de retardos diferente que depende de la posición en que se encuentren. Sin embargo, cuando el número de micrófonos es par, los valores del vector de direcciones en los micrófonos posicionados de manera opuesta (180 grados) son complejos conjugados.

La posición de cada micrófono dependerá de la longitud del radio de curvatura $r_c$ del arreglo circular y del ángulo de su posición $\theta_m$ con respecto al eje de referencia o cero grados. El radio de curvatura del arreglo circular de micrófonos se calcula por medio de la Ecuación (2.52)

$$r_c = \frac{\lambda_s}{4\sin\left(\frac{\pi}{M}\right)}, \tag{2.52}$$

donde $\lambda_s$ es la longitud de onda más pequeña del ancho de banda de las señales que se van a adquirir con el arreglo de micrófonos [26], [30]. Si el radio de la circunferencia es mayor al valor calculado en la Ecuación (2.52), se presenta un fenómeno conocido como *aliasing espacial* en el patrón de radiación que se explica más adelante en la Sección 3.3.

Tomando en cuenta que la estimación del ángulo de arribo se realiza en el plano horizontal de los micrófonos, es decir, el ángulo de azimuth $\theta$, el ángulo de elevación se declara como constante en $\phi = 90$, por lo que los retardos de tiempo de cada micrófono $\tau_m$ respecto al centro del arreglo de micrófonos se calculan por medio de la Ecuación (2.53)

$$\tau_m = r_c \pi \cos(\theta - \theta_m) \tag{2.53}$$

donde $\theta_m$ es el ángulo de posición de cada micrófono.

El vector de direcciones para el caso de un arreglo de micrófonos circular se puede observar en la Ecuación (2.54) cuando el ángulo de elevación $\phi = 90$ grados.

$$\mathbf{a}(\theta) = \left[ \; e^{jkr_c\cos(\theta-\theta_1)} \quad e^{jkr_c\cos(\theta-\theta_2)} \quad e^{jkr_c\cos(\theta-\theta_3)} \quad \ldots \quad e^{jkr_c\cos(\theta-\theta_M)} \; \right] \tag{2.54}$$

El vector de direcciones para un arreglo circular con un número $M$ par de micrófonos tiene la característica de que la segunda mitad de sus elementos corresponden al conjugado complejo de los elementos de la primera, esto se debe a la función coseno utilizada en la Ecuación (2.54), por ejemplo, suponiendo que se tiene un arreglo circular de 8 micrófonos, los micrófonos van a estar localizados sobre la circunferencia cada 45 grados, además de que las señales adquiridas tienen un ancho de banda de $200Hz$ a $4kHz$, por lo que el radio de curvatura del arreglo circular se considera de $r = 0.055m$. El vector de direcciones para este caso se muestra a continuación:

$$\begin{aligned}\mathbf{a}(\theta) = [ \; & e^{jkr_c\cos(\theta)} \quad e^{jkr_c\cos(\theta-45)} \quad e^{jkr_c\cos(\theta-90)} \quad e^{jkr_c\cos(\theta-135)} \\ & e^{jkr_c\cos(\theta-180)} \quad e^{jkr_c\cos(\theta-225)} \quad e^{jkr_c\cos(\theta-270)} \quad e^{jkr_c\cos(\theta-315)} \; ]\end{aligned} \tag{2.55}$$

sin embargo, se puede calcular la mitad del vector de direcciones de la Ecuación (2.55) considerando que:

$$\begin{aligned} e^{jkr_c\cos(\theta-180)} &= e^{-jkr_c\cos(\theta)} \\ e^{jkr_c\cos(\theta-225)} &= e^{-jkr_c\cos(\theta-45)} \\ e^{jkr_c\cos(\theta-270)} &= e^{-jkr_c\cos(\theta-90)} \\ e^{jkr_c\cos(\theta-315)} &= e^{-jkr_c\cos(\theta-135)} \end{aligned}$$

## 2.3. Detección de actividad de voz (VAD)

Muchos de los algoritmos de procesamiento de voz, ya sea estimación de dirección de arribo de señales provenientes de fuentes, filtrado espacial y reconocimiento, consideran que la presencia de voz existe en todo momento, pero en realidad la presencia de voz en un discurso consta de lapsos de tiempo en forma aleatoria, por lo que en un frame con un número finito de muestras es importante detectar si existe la presencia de voz o si las señales adquiridas son únicamente ruido del ambiente.
Diseñar una etapa de detección de actividad de voz en un ambiente real no es una tarea sencilla porque se desconocen los siguientes elementos:

1. Número de fuentes de voz.

2. Características de las fuentes de voz (posición, componentes frecuenciales, intensidad).

3. Características del ruido acústico presente en el recinto.

Utilizando el modelo de la señal empleado en la Ecuación (2.33), se puede estimar la presencia de señal de voz o ruido en una serie de datos a través de la siguiente hipótesis [2]:

$$H_0: \quad \mathbf{x}(n) = \mathbf{A}(\theta)\mathbf{s}(n) + \mathbf{n}(n) \tag{2.56}$$

$$H_1: \quad \mathbf{x}(n) = \mathbf{n}(n)$$

La función básica de un algoritmo VAD es obtener algunas características de la señal de entrada y compararlas con un umbral [31]; comúnmente la energía de la señal se utiliza como medida de comparación como se muestra en la Figura 2.8.

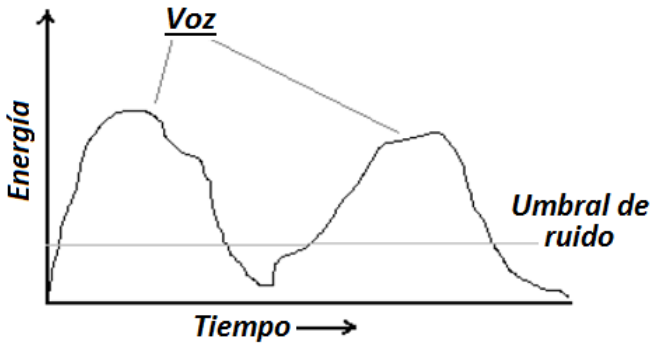



Figura 2.8: VAD basado en un umbral de energía fijo [3].

Los regiones en donde existe información de las señales de voz se conocen como ”voz activa” mientras que los segmentos de pausa entre un discurso se llaman ”voz inactiva” o ”silencio”. La precisión de un algoritmo de detección de actividad de voz depende en gran medida de los umbrales de decisión.

### 2.3.1. VAD con umbral de energía fijo

Las señales se pueden relacionar con cantidades físicas que representan potencia y energía. Cuando una señal periódica tiene energía finita en un intervalo de tiempo $T_p$, la potencia de la misma se define por la Ecuación (2.57) [32], [33].

$$P_x = \frac{1}{T_p} \int_{T_p} |x(t)|^2 \, dt \tag{2.57}$$

De manera similar, para una señal de tiempo discreto $x(n)$ definida en el intervalo $n_1 < n < n_2$, la energía de la señal se calcula por medio de la Ecuación (2.58)

$$E_x = \frac{1}{N} \sum_{n=n_1}^{n_2} |x(n)|^2 \tag{2.58}$$

por lo tanto, si definimos que el intervalo de una señal discreta completa está definido como $0 < n < N$ y además se le aplican ventanas cuadradas de longitud $L$, la energía para cada ventana $l$ se calcula por medio de la Ecuación (2.59)

$$E_x(l) = \frac{1}{NL} \sum_{l=1}^{NL} \sum_{n=1}^{L} |x((l-1)L + n)|^2 \tag{2.59}$$

donde $NL$ es el número total de ventanas.

La energía de cada ventana adquirida $E_x(l)$ de la señal $x(n)$, se compara con un umbral de valor propuesto $T_H$. Si la energía rebasa el valor del umbral, la ventana $l$ de señal adquirida se considera como segmento activo (VAD=1), es decir que se detecta la actividad de voz, de otra manera el segmento es inactivo (VAD=0) y el algoritmo clasificará la ventana como ruido. Lo explicado anteriormente se puede observar en el diagrama de flujo de la Figura 2.9.

#### Valor inicial del umbral

Para calcular el valor del umbral $T_H$, se consideran las características del ruido del recinto en donde se esté realizando la estimación. Cuando existe actividad de voz, la energía de la señal calculada en la salida de un micrófono contendrá la suma de la energía del ruido y de la señal, como se observa en la Ecuación (2.60)

$$E_{total} = E_s + E_\eta \tag{2.60}$$

por lo tanto, es importante determinar la energía relacionada al ruido $E_\eta$ y con base en éste valor, determinar el umbral de energía para la detección de la actividad de voz.

Para calcular el umbral nos basaremos en la hipótesis $H_1$ de la Ecuación (2.56), es decir, se puede suponer que los primeros instantes de la adquisición de las señales son únicamente ruido acústico del ambiente, por lo que se calcula el promedio de la energía del ruido de $I$ ventanas iniciales como se muestra en la Ecuación (2.61)

$$E_{\eta_{prom}} = \frac{1}{I} \sum_{i=0}^{I} E_i \tag{2.61}$$

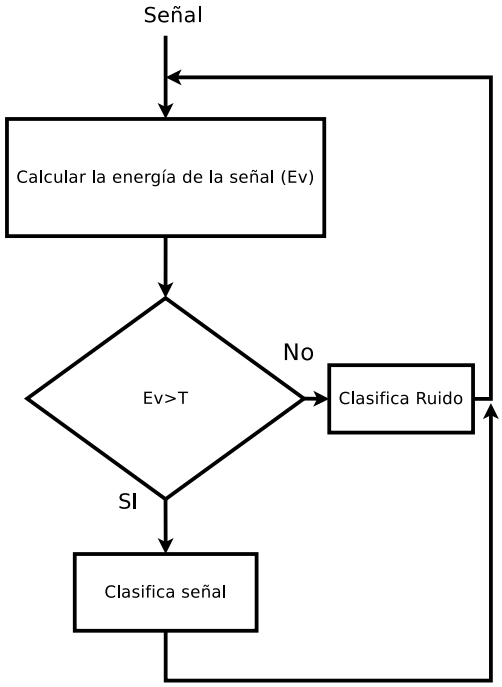



Figura 2.9: Algoritmo VAD para un umbral de energía fijo.

Sustituyendo la Ecuación (2.59) en (2.61), se puede obtener la energía promedio en el intervalo de tiempo inicial como:

$$E_{\eta_{prom}} = \frac{1}{LI} \sum_{l=1}^{I} \sum_{n=1}^{L} |x((l-1)L+n)|^2 \tag{2.62}$$

Finalmente, el umbral de ruido inicial se puede calcular por medio de la Ecuación (2.63)

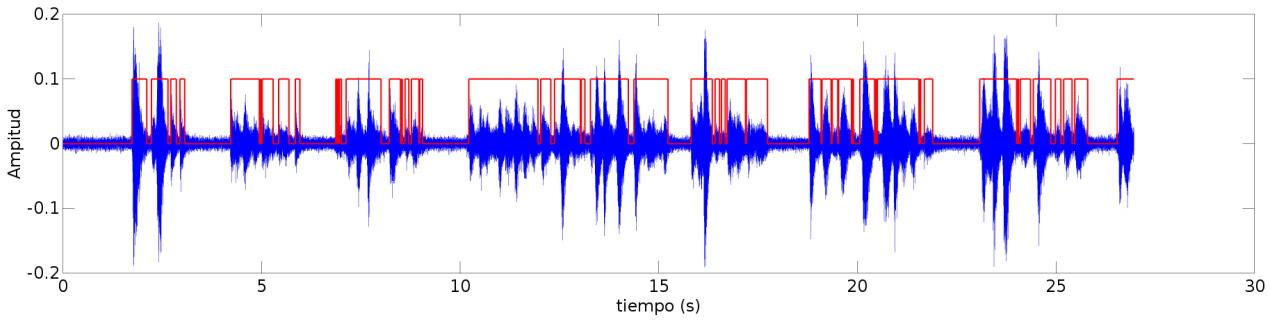
$$T_{inicial} = E_{\eta_{prom}} + k_{VAD} E_{\eta_{prom}} \tag{2.63}$$

donde $k$ es un factor de escala que define la tolerancia de la energía de paso, lo que significa que entre mayor sea el valor de $k_{VAD}$, la señal de voz deberá contener mayor energía $E_s$ para superar el umbral de decisión.

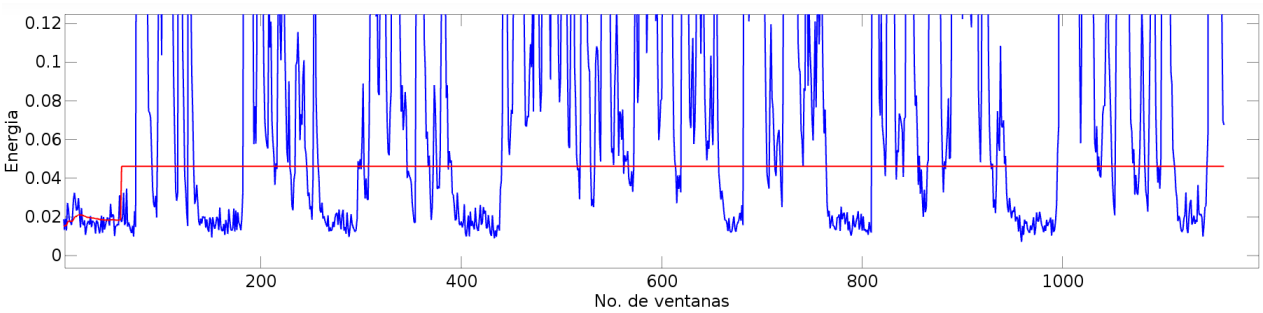
En la Figura 2.10 se muestra un ejemplo de un VAD fijo con una señal de voz adquirida con una frecuencia de muestreo de $f_s = 44100\,Hz$ y un valor de $k_{VAD} = 1.5$. En la Figura 2.10(a) se puede observar la señal de voz (color azul) comparada con los fragmentos detectados como activos en forma de funciones escalón (color rojo), mietras que en la Figura 2.10(b) se exhibe el valor del umbral fijo comparado con la energía de las ventanas de la señal adquirida.

### 2.3.2. VAD adaptable

Cuando el ruido adquirido por los micrófonos no es estacionario, se requiere que el valor del umbral de energía se actualice de manera proporcional a los cambios del ruido en el dominio del tiempo. Los algoritmos de detección de actividad de voz adaptables se basan en calcular




(a) Actividad de voz con umbral fijo.




(b) Comparación del umbral con la energía de la señal.

Figura 2.10: VAD fijo.

un umbral de energía en forma iterativa, calculando un nuevo valor de energía con base en los anteriores y la variación del ruido. Este valor es usualmente calculado en segmentos de tiempo cuando no existe actividad de voz.

Un algoritmo de detección de actividad de voz robusto debe cumplir con las siguientes características [34]:

1. Implementar una buena regla de decisión que explote las características de la señal de voz para clasificar los segmentos de la señal como activos o inactivos.

2. Adaptación al ruido acústico no estacionario.

3. Baja tasa de error en la clasificación de voz como ruido.

Se han estudiado diferentes algoritmos basados en calcular un umbral que sea capaz de seguir la variación del ruido e incluso en el momento en que una ventana se clasifica como activa. Los siguientes métodos se basan en actualizar el valor del umbral con base al valor anterior y la variación de la energía del ruido en las señales adquiridas.

**Detector basado en la energía lineal (LED I)**

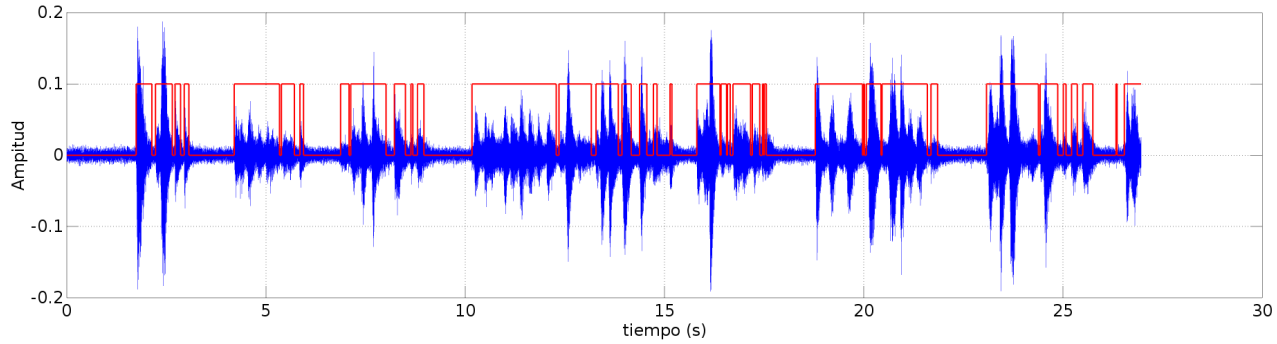
Este método surge a partir de que un valor de umbral fijo no es capaz de identificar las variaciones del ruido. En este método se aprovechan las ventanas inactivas para actualizar el valor del umbral, de tal manera que sea capaz de seguir el comportamiento del ruido.

El valor de la energía del ruido se actualiza por medio de la Ecuación (2.64)

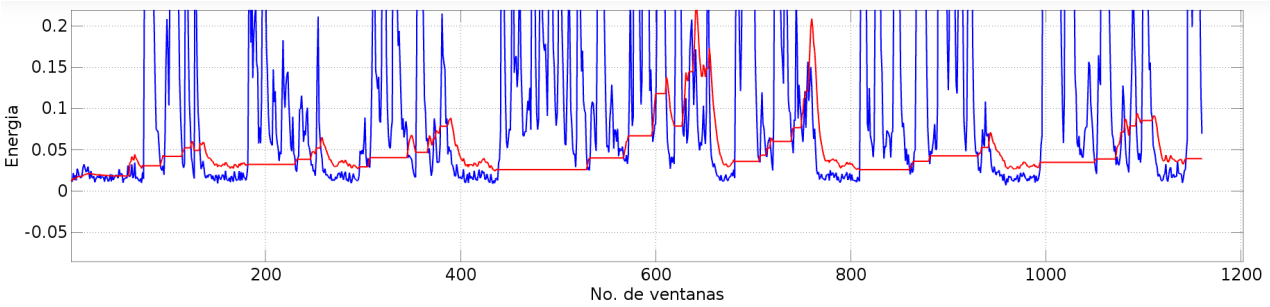
$$E_{rnuevo} = \left(1 - p_{VAV}\right) E_{ranterior} + p_{VAD} E_{\eta} \tag{2.64}$$

donde $E_{rnuevo}$ es el valor actualizado del umbral, $E_{ranterior}$ el valor del umbral de la ventana inactiva anterior y $E_{\eta}$ es el valor de energía calculada de la ventana inactiva más reciente. El valor de $p_{VAD}$ se selecciona considerando la respuesta al impulso de la Ecuación (2.64) y se encuentra en el intervalo de $(0 < p_{VAD} < 1)$ [34].

En la Figura 2.11(a) se muestra un ejemplo de la respuesta del detector de actividad de voz de una señal de voz inmersa en ruido ambiental con una relación señal a ruido de 7.79 dB. La señal que se muestra en la Figura 2.11(a) es una señal muestreada a $f_s = 44100 Hz$. En la Figura 2.11(b) se muestra el comportamiento del umbral de energía comparada con la energía de la señal utilizando la constante $p_{VAD} = 0.25$.




(a) VAD.




(b) Comportamiento del umbral de energía.

Figura 2.11: VAD adaptable con LED I.

**Detector basado en energía lineal con doble umbral (LED II)**

Al considerar que el ruido presente en el recinto es de tipo no estacionario, la energía del ruido va a estar variando en el instante que exista presencia de voz y sin ella, por lo que un mejor enfoque para el cálculo del umbral es estimar la energía del ruido en ambos casos.

En este método se consideran dos umbrales, uno se calcula en ventanas activas, es decir, que contiene presencia de voz, y el otro umbral cuando no existe actividad de voz. El algorit-

mo consta primero en calcular el nivel de energía de ruido por medio de la Ecuación (2.63). Posteriormente, el valor del umbral se calcula con la Ecuación (2.65).

$$E_{r_{nuevo}} = \lambda_1 E_{r_{anterior}} + (1-\lambda_1)\, E_j \tag{2.65}$$

para segmentos activos, mientras que para segmentos inactivos por medio de la Ecuación (2.66).

$$E_{r_{nuevo}} = \lambda_2 E_{r_{anterior}} + (1-\lambda_2)\, E_j \tag{2.66}$$

donde $\lambda_1$ y $\lambda_2$ son los factores de adaptación que están definidos en los intervalos [0.85,0.95] y [0.98,0.999], respectivamente [35]. Los valores $\lambda_1$ y $\lambda_2$ definen un filtro paso bajas en la energía de la señal. El valor de decaimiento definido por $\lambda_1$ es fijo con la restricción de ser pequeño para que siga el comportamiento del ruido, pero mayor a la variación de la señal de voz para evitar que la siga. De la misma manera para el caso del valor de $\lambda_2$, debe de ser suficientemente grande para evitar seguir el comportamiento de la energía de la señal de voz pero lo suficientemente pequeño para adaptarse en las variaciones del ruido [34], [35].

Los umbrales en los segmento activos e inactivos se calculan por medio de la Ecuación (2.67)

$$T_{voz_{nuevo}} = E_{r_{nuevo}} + \delta_{voz} \tag{2.67}$$

$$T_{silencio_{nuevo}} = E_{r_{nuevo}} + \delta_{silencio}$$

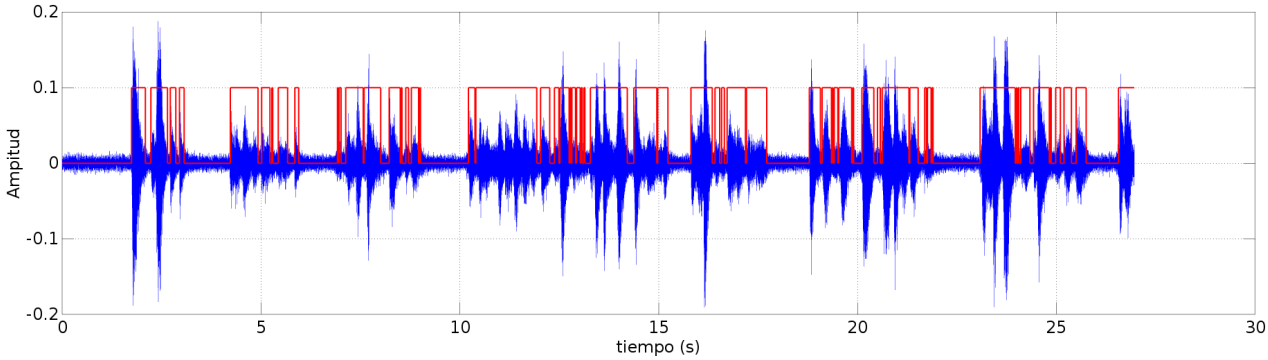
donde $\delta_{silencio}$ y $\delta_{voz}$ son constantes utilizadas para determinar el umbral. Por lo regular se seleccionan en los intervalos $[0.1, 0.4]$ para $\delta_{silencio}$ y $[0.5, 0.8]$ para $\delta_{voz}$, pero pueden cambiarse dependiendo de cada caso en particular.

En la Figura 2.12(a) se muestra la respuesta del detector de actividad de voz utilizando la misma señal que en el caso de la Figura 2.11(a). Los valores de las constantes fueron $\lambda_1 = 1$ y $\lambda_2 = 0.999$, además de utilizar los valores de $\delta_{voz} = 0.35$ y $\delta_{silencio} = 0.3$ para el caso del umbral de ruido y con actividad de voz. En la Figura 2.12(b) se muestra el comportamiento del umbral comparado con la energía de la señal.

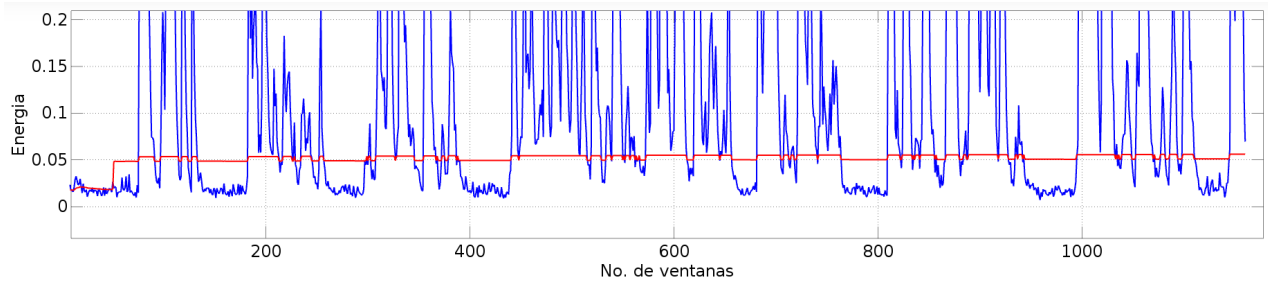
### 2.3.3. VAD en el Dominio de la Frecuencia

Otra forma de realizar la detección de actividad de voz se puede hacer en el dominio de la frecuencia de una manera parecida a los métodos mencionados anteriormente. La idea del método se basa en calcular la energía del espectro de la señal en un ancho de banda determinado para compararlo con el valor de un umbral previamente determinado. El ancho de banda de una señal en el dominio de la frecuencia $X(f) = F[x(n)]$ se divide en $NB$ bandas, con el objetivo de calcular un umbral para cada una de ellas.

Los valores iniciales de los umbrales para cada una de las bandas se pueden adquirir de manera parecida a como se realizó en el caso de la Sección 2.3.1, es decir, la energía promedio calculada en la banda $W$ para un número NL de frames se obtiene por medio de la Ecuación (2.68)




(a) VAD.




(b) Comportamiento del umbral de energía.

Figura 2.12: VAD adaptable con LED II.

$$E_{prom_{frec}}(W) = \frac{1}{NL}\sum_{l}^{NL}\sum_{f=1}^{AB}\left|X\left((l-1)\,W+f\right)\right|^2 \tag{2.68}$$

con $W = 1,2,\ldots,NB$, donde $AB$ es el tamaño de la banda de trabajo. El valor inicial del umbral para la banda $W$ se calcula por la Ecuación (2.69).

$$T_{inicial}(W) = E_{prom_{frec}}(W) + k_{VAD}E_{prom_{frec}}(W) \tag{2.69}$$

En la Figura 2.13 se muestra la detección de actividad de voz cuando el ancho de banda de la señal $x(n)$ se divide en cuatro bandas equidistantes, con los siguientes intervalos: $0-1\,kHz$, $1-2\,kHz$, $2-3\,kHz$ y $3-4\,kHz$.

La mayor parte de la energía en una señal de voz se concentra en las frecuencias dentro del intervalo $0-1\,kHz$, es por eso que se considera la banda principal a la de menor frecuencia [3]. Se puede observar de la Figura 2.13 que un segmento está activo cuando la energía de al menos dos bandas en conjunto con la principal, superan los umbrales de dichas bandas.

Finalmente, el valor de los umbrales se puede actualizar en los momentos cuando se detectan segmentos inactivos de voz. De manera similar que en la Sección 2.3.2, el umbral se actualiza por medio de la Ecuación (2.70).

$$E_{rnuevo}(W) = \left(1-p_{VAD}\right)E_{ranterior}(W) + p_{VAD}E_{silencio}(W) \tag{2.70}$$

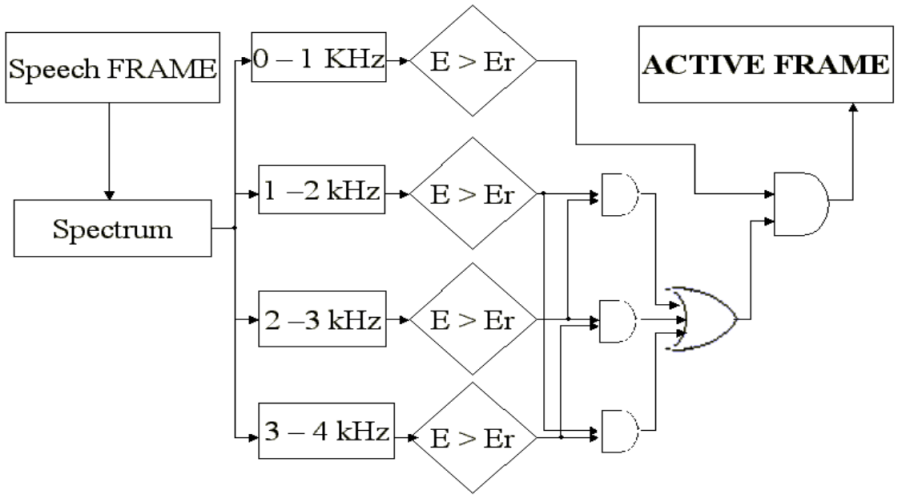



Figura 2.13: VAD en el dominio de la frecuencia con cuatro bandas de frecuencia [3]

En la Figura 2.14 se muestra la respuesta de un detector de actividad de voz con una partición del ancho de banda en cuatro bandas en los intervalos: $0-1\,kHz$, $1-2\,kHz$, $2-3\,kHz$ y $3-4\,kHz$. La señal de voz mostrada en color azul de la Figura 2.14 se adquirió con una frecuencia de muestreo de 44100 $Hz$, utilizando un valor constante de $p_{VAD} = 0.25$ y la curva de color rojo muestra los fragmentos de voz activos que fueron detectados por el algoritmo, dichos fragmentos activos se representan por medio de funciones escalón, donde el nivel alto de energía muestra la detección de un fragmento de voz activo mientras que los inactivos se representan en cero.

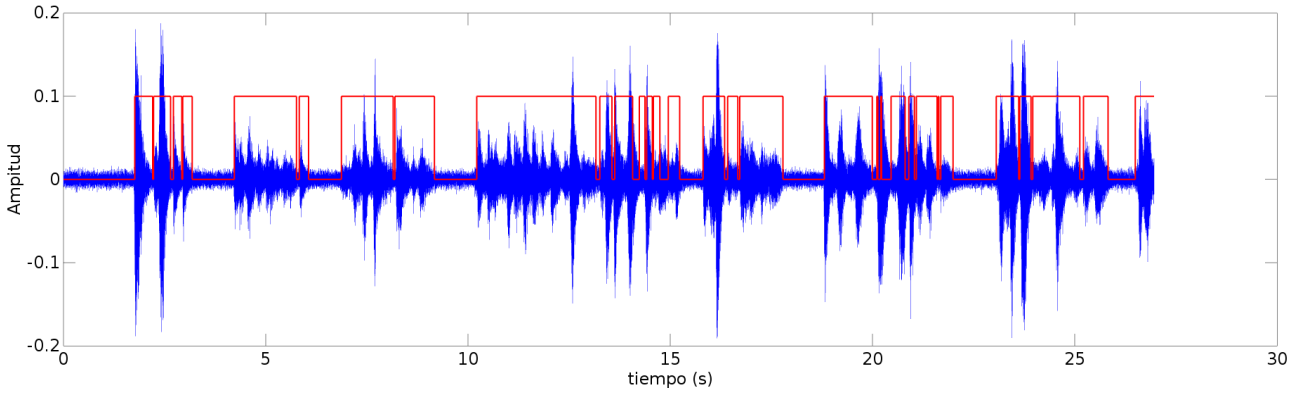



Figura 2.14: Respuesta en tiempo de un VAD de cuatro bandas.

La detección de actividad de voz en el dominio de la frecuencia ha mostrado una mejor detección utilizando cuatro bandas comparado con el VAD calculado en el dominio del tiempo. Dicha comparación se puede observar a través de las Figuras 2.11 y 2.14, donde el VAD en el dominio del tiempo ha detectado pequeños fragmentos de voz los cuales representan el ruido de fondo cuando la potencia del mismo tuvo un incremento inesperado, mientras que el VAD en el dominio de la frecuencia mantiene su detección cuando la voz se encuentra activa.

## 2.4. Resumen

En este capítulo se han estudiado las bases de propagación del sonido con su respectivo desarrollo matemático, para deducir el modelo de la señal utilizado en el presente trabajo. Cabe aclarar que la acústica es un área de investigación extensa, sin embargo en el presente capítulo se muestran únicamente las bases necesarias para entender el comportamiento de la propagación de la señal hacía un arreglo de micrófonos. Se realizó el estudio de los arreglos de micrófonos de interés (arreglo lineal uniforme, arreglo circular uniforme y arreglo triangular equidistante), sin embargo, es necesario tomar en cuenta que existen otros tipos de geometrías para arreglos de micrófonos, no obstante, se eligieron éstas principalmente porque se puede utilizar un número pequeño de micrófonos en el arreglo y de esta manera la carga computacional en el procesamiento digital de señales será menor y facilitará la implementación del sistema en tiempo real. Por otro lado, la detección de actividad de voz es una etapa fundamental para algunas de las aplicaciones que se basan en la extracción de parámetros de las señales, tales aplicaciones son la identificación de idiomas, la estimación de DOA y reconocimiento de voz, entre otras. Comparando las respuestas de los VADs por medio de las figuras 2.10, 2.11, 2.12 y 2.14, se puede observar que el VAD en el dominio de la frecuencia tiene una mejor respuesta que los demás, por tal motivo, en el presente trabajo se implementó un VAD en el dominio de la frecuencia con cuatro bandas como se explicó en la Sección 4.2.

# Capítulo 3.

# Métodos de estimación de la dirección de arribo

La estimación de dirección de arribo de una señal ha sido un área de investigación muy concurrida en las últimas décadas, principalmente por los campos de aplicación en el procesamiento de voz como lo son videoconferencia, seguridad y robótica entre otras.

El objetivo de la estimación de dirección de arribo es realizar un procesamiento específico para determinar el ángulo de procedencia de las señales emitidas por una fuente y adquiridas por un arreglo de micrófonos; para fines de este trabajo las señales procesadas son emitidas por una fuente de voz.

Las señales de voz tienen un ancho de banda en el intervalo de $250\,Hz - 4\,kHz$, además de tener un frente de onda esférico. Sin embargo, si la fuente de voz se localiza lo suficientemente lejos con respecto al arreglo de micrófonos, el frente de onda se puede considerar en forma planar como lo explica la teoría del modelo de campo lejano en la Sección 2.1.3. Esta consideración es muy importante para los algoritmos de estimación de DOA porque en general se aprovechan de las diferencias de tiempo de llegada de las señales adquiridas por los micrófonos del arreglo.

Se han desarrollado diferentes métodos para la estimación del ángulo de arribo, los cuales pueden ser divididos en tres campos [10], [36]:

1. Diferencia de tiempo de arribo(TDOA)
2. Técnicas basadas en eigendescomposición
3. Formador de haz (BF)

En el presente capítulo se explica la teoría de la estimación de la dirección de arribo con un arreglo de sensores. En la Sección 3.1 se realiza un análisis de la estimación del ángulo de dirección por medio de la diferencia de tiempo de llegada entre las señales adquiridas por dos micrófonos. En la Sección 3.2 se explica la teoría de las técnicas de eigendescomposición de la matríz de correlaciones de las señales de entrada y el método MUSIC (*Mulple signal clasification*) y finalmente en la Sección 3.3 se detalla la teoría del formador de haz y su aplicación en la estimación del fuentes.

## 3.1. Diferencia de tiempo de arribo (TDOA)

Este método se basa en calcular la dirección de arribo a partir del retardo o diferencia de tiempo de una señal adquirida por un par de micrófonos ($\tau_{1,2}$). Tomando en cuenta el retardo que sufre una señal al incidir en un micrófono, dado que arribó previamente en un micrófono de referencia, la dirección de arribo de la señal se puede calcular por medio de la Ecuación (3.1)

$$\theta = \arcsin\left(\frac{\tau_{1,2}c}{d}\right). \tag{3.1}$$

El retardo de incidencia de una señal en un par de micrófonos se puede calcular por medio de la correlación entre las dos series de tiempo discreto de N datos. La correlación entre dos señales mide el grado de semejanza entre ambas señales. En la Ecuación (3.2) se muestra la correlación entre dos series de tiempo discreto para el l-ésimo parámetro de desplazamiento

$$R_{1,2}(l) = \frac{1}{N}\sum_{N=0}^{N-1} x_1(n)x_2(n+l) \tag{3.2}$$

donde $x_1$ y $x_2$ son las secuencias de tiempo discreto de la señal en el primer y segundo micrófono respectivamente [32]. Como la correlación cruzada se realiza entre dos señales que son prácticamente iguales, el retardo de las señales en términos de muestras se ve reflejado en el valor del índice a partir del máximo global de la correlación $R_{1,2}(k)$, por lo que el tiempo de retardo estimado entre las señales se expresa por medio de la Ecuación (3.3) [10].

$$\tau_{1,2} = \frac{1}{f_s} argmax\left(R_{1,2}(k)\right) \tag{3.3}$$

El problema de calcular la correlación cruzada por medio de la Ecuación (3.2) es que requiere un número de operaciones en el orden de $N^2$ multiplicaciones, sin embargo, cuando se trabaja con arreglos de $M$ micrófonos, realizar la correlación entre todos los micrófonos demanda

un costo computacional alto y dificulta su ejecución en tiempo real. No obstante, se puede realizar una aproximación de la correlación en el dominio de la frecuencia que demanda $N\log_2(N)$ operaciones [13].

El método consiste en obtener la correlación, realizando la esperanza matemática de las señales en el dominio de la frecuencia de cada uno de los micrófonos y, posteriormente aplicar una transformada inversa de Fourier de tiempo discreto ($\mathscr{F}^{-1}$)), es decir:

$$\mathbf{R_{1,2}} = \mathscr{F}^{-1}\left[\vartheta(\omega)\,\Psi_{x_1x_2}(\omega)\right]. \tag{3.4}$$

Se conoce a $\mathbf{R_{1,2}}$ como la correlación cruzada generalizada (GCC), donde $\vartheta(\omega)$ es una función de peso en el dominio de la frecuencia y $\Psi_{x_1x_2}(\omega)$ es la esperanza matemática $E[\cdot]$ de las señales $x_1(n)$ y $x_2(n)$ en el dominio de la frecuencia como se muestra en la Ecuación (3.5) [10], [13]

$$\Psi_{x_1,x_2}(\omega) = E\left[X_1(\omega)\,X_2^*((\omega)\right] \tag{3.5}$$

La función de peso $\vartheta(\omega)$ puede tomar diferentes valores, el más común es realizar un blanqueado de las señales antes de calcular la correlación y de ésta manera aproximar los máximos de interés de la correlación a funciones delta. La forma de obtener esa aproximación es aplicar la llamada transformada de fase como se muestra a continuación:

$$\vartheta(\omega) = \frac{1}{|\Psi_{x_1,x_2}(\omega)|} \tag{3.6}$$

Por lo que al sustituir la Ecuación (3.6) en (3.4), se obtiene la Ecuación (3.7), conocida como correlación cruzada generalizada con transformada de fase (GCC-PHAT)

$$GCC_{PHAT} = \mathscr{F}^{-1}\left[\frac{\Psi_{x_1x_2}(\omega)}{|\Psi_{x_1,x_2}(\omega)|}\right] = \mathscr{F}^{-1}\left[\frac{X_1(\omega)\,X_2^*(\omega)}{|X_1(\omega)\,X_2^*(\omega)|}\right] \tag{3.7}$$

La desventaja de la GCC-PHAT es que su respuesta se degrada cuando se tiene un ruido aditivo fuerte, porque al aplicar la transformada de fase, busca linealizar todas las amplitudes de cada una de las bandas en la transformada de Fourier ($\mathscr{F}$).

## 3.2. Técnicas basadas en eigendescomposición

Estas técnicas fueron inicialmente desarrolladas en radar para la estimación de DOA, centrándose en su implementación para banda angosta, sin embargo, posteriormente se adaptaron para poder realizar la estimación con señales de banda ancha. Estas técnicas se basan en extraer los parámetros de las señales adquiridas por un arreglo de micrófonos aprovechando las características de la matriz de correlación. La extracción de las características se basa en la eigendescomposición de la matriz espacial de covarianza, obteniendo de esta manera los valores (eigenvalores) y vectores propios (eigenvectores) de la misma. Se considera que el ruido aditivo en las señales adquiridas no está correlacionado con la señal de interés, además de ser ruido blanco con media cero y varianza constante.

En la siguiente sección se describe el algoritmo Clasificación de múltiples señales (MUSIC) propuesto por Ralph O. Schmidt en 1986 [4].

### 3.2.1. MUSIC

La idea principal del algoritmo MUSIC, se basa en separar la señal del ruido utilizando la propiedad de ortogonalidad de sus espacios a través de la eigendescomposición de la matriz de covarianza; tal matriz se construye por medio de las señales adquiridas a través del arreglo de micrófonos, como se muestra en la Ecuación (2.35). Supone que el ruido adquirido por los micrófonos no está correlacionado con las señales de la fuente de voz [10], [37].

La matriz de covarianza de las señales a la salida de los micrófonos se calcula a partir de la Ecuación (2.35), misma que puede ser representada por medio de la Ecuación (3.8)

$$\mathbf{R_{xx}} = \mathbf{a}(\theta)\mathbf{R_{ss}}\mathbf{a}^H(\theta) + \sigma_v^2\mathbf{I} = \sigma_s^2\mathbf{a}(\theta)\,\mathbf{a}^H(\theta) + \sigma_v^2\mathbf{I} \tag{3.8}$$

donde $\mathbf{a}(\theta)$ representa el vector que contiene las dirección para cada señal, definido en la Ecuación (2.31), $\sigma_s^2$ y $\sigma_v^2$ son las varianzas de la señal y el ruido respectivamente e $\mathbf{I}$ es la matriz identidad de tamaño $M\ X\ M$. La eigendescomposición de la matriz de covarianza $\mathbf{R_{xx}}$ se observa en la Ecuación (3.9).

$$\mathbf{R_{xx}} = \mathbf{B}\Lambda\mathbf{B}^H \tag{3.9}$$

La eigendescomposición de la matriz de covarianza $\mathbf{R_{xx}}$ en la Ecuación (3.9) consta de la matriz diagonal $\Lambda$, que contiene los valores propios o eigenvalores de de la matriz de covarianza $\mathbf{R_{xx}}$, y la matriz de eigenvectores $\mathbf{B}$, descritos en (3.10) y (3.11), respectivamente.

$$\Lambda = diag\left[\ \lambda_{x,1}\quad \lambda_{x,2}\quad \lambda_{x,3}\quad \cdots\quad \lambda_{x,M}\ \right] = diag\left[\ \lambda_{s,1}+\sigma_v^2\quad \sigma_v^2\quad \sigma_v^2\quad \cdots\quad \sigma_v^2\ \right] \tag{3.10}$$

$$\mathbf{B} = \left[\ \mathbf{b_1}\quad \mathbf{b_2}\quad \mathbf{b_3}\quad \cdots\quad \mathbf{b_M}\ \right] \tag{3.11}$$

Cada eigenvector $\mathbf{b_n}$ está asociado a un eigenvalor $\lambda_{x,n}$ de la matriz $\mathbf{R_{xx}}$. Los valores propios de la matriz representan tanto a la señal de interés como a elementos ruidosos dentro del subespacio de las señales que se muestra en la Figura 3.1. Suponiendo que una señal de banda angosta inside al arreglo de micrófonos, $\mathbf{R_{ss}}$ tendrá un único eigenvalor positivo $\lambda_{s,1}$ asociado a la señal [10] como se muestra en el Ecuación (3.10), entonces los eigenvalores $\lambda_{x,2}, \lambda_{x,3}, ..., \lambda_{x,M}$ representan los elementos ruidosos en el subespacio de la señal. Por lo que para $n \geq 2$, tenemos:

$$\mathbf{R_{xx}}\mathbf{b_n} = \lambda_{x,n}\mathbf{b_n} = \sigma_v^2\mathbf{b_n} \tag{3.12}$$

$$\mathbf{R_{xx}}\mathbf{b_n} = \left(\sigma_s^2\mathbf{a}(\theta)\mathbf{a}^H(\theta) + \sigma_v^2\mathbf{I}\right)\mathbf{b_n} \tag{3.13}$$

sustituyendo (3.13) en (3.12) se obtiene

$$\sigma_s^2\mathbf{a}(\theta)\,\mathbf{a}^H(\theta)\,\mathbf{b_n} = \mathbf{0} \tag{3.14}$$

que es equivalente a:

$$\mathbf{a}^H(\theta)\,\mathbf{b_n} = \mathbf{b}_n^H\mathbf{a}(\theta) = 0 \tag{3.15}$$

El producto punto de la Ecuación (3.15) significa que los eigenvectores ruidosos son ortogonales a la dirección del ángulo de arribo de la señal de interés en el subespacio de la señal.

Para ejemplificar lo anterior, vamos a suponer dos fuentes de sonido de banda angosta posicionadas en dos direcciones diferentes y un arreglo de tres micrófonos para la adquisición de las señales. Al realizar la eigendescomposición de la matriz de covarianza $\mathbf{R_{xx}}$, de tamaño $3X3$, se obtienen tres eigenvalores ($\lambda_1$,$\lambda_2$ y$\lambda_3$) con sus respectivos eigenvectores ($\mathbf{b_1}$, $\mathbf{b_2}$ y $\mathbf{b_3}$), sin embargo, dos de los eigenvalores con sus respectivos eigenvectores ($\mathbf{b_1}$ y $\mathbf{b_2}$) corresponden al subespacio de las señales, mientras que el eigenvector ($\mathbf{b_{min}} = \mathbf{b_3}$), relacionado al eigenvalor mínimo, representa al espacio ruidoso, como se muestran en la Figura 3.1.

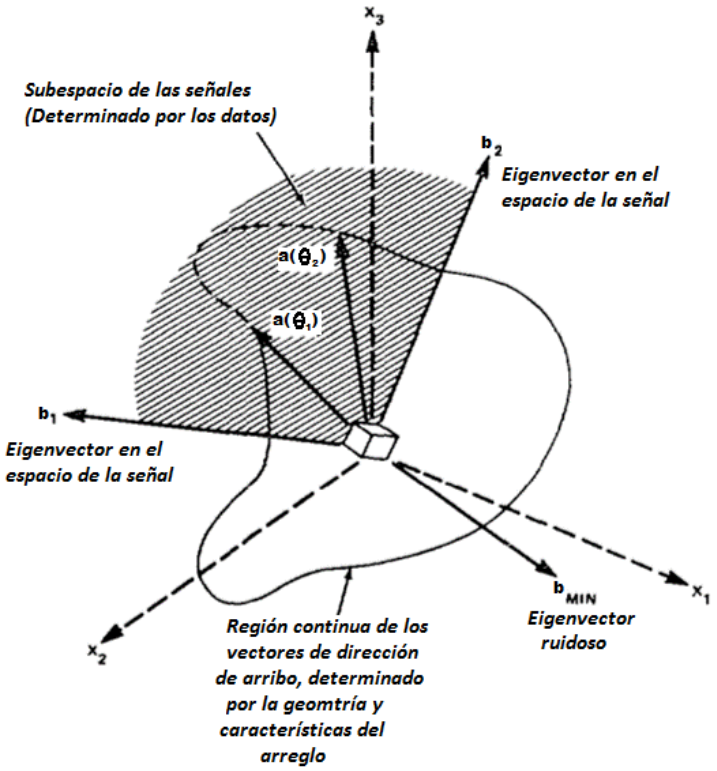



Figura 3.1: Representación de los eigenvectores en el subespacio de las señales [4].

El conjunto de vectores que se muestran dentro de la curva cerrada en la Figura 3.1, representan todas las posibles direcciones de arribo definidas por el vector de direcciones $\mathbf{a}(\theta)$ que es determinado por las característica y la geometría del arreglo, mientras que el área sombreada es el plano definido por los eigenvectores $\mathbf{b_1}$ y $\mathbf{b_2}$, conocido como el subespacio de las señales. La dirección de arribo de las fuentes de banda angosta $\mathbf{a}(\theta_1)$ y $\mathbf{a}(\theta_2)$ se encuentran en el plano formado por el subespacio de la señal, mientras que el eigenvector ruidoso es perpendicular a dichas direcciones de arribo, es decir, el producto punto entre el eigenvector ruidoso y la dirección de la señal es igual a cero como se derminó en las ecuaciones (3.14) y (3.15).

Definiendo a $\mathbf{E_N}$ como la matriz de tamaño $M\ X\ N$ que está formada por los $M$ elementos de los $N$ eigenvectores ruidosos, y la distancia euclidiana $d_{euc}$ desde un vector $\mathbf{Y}$ al subespacio de la señal como:

$$d_{euc}^2 = \mathbf{Y}^H \mathbf{E_N} \mathbf{E_N}^H \mathbf{Y}, \tag{3.16}$$

se puede construir una gráfica de $\frac{1}{d_{euc}^2}$ en función de cada una de las posiciones angulares del vector de direcciones $\mathbf{a}(\theta)$, es decir:

$$MUSIC(\theta) = \frac{1}{\mathbf{a}^H(\theta)\,\mathbf{E_N}\mathbf{E_N}^H\mathbf{a}(\theta)} \tag{3.17}$$

La Ecuación (3.17) modela el espectro MUSIC que se muestra en la Figura 3.2. Los picos máximos en el espectro MUSIC representan la posición o el ángulo de dirección de arribo de las señales.

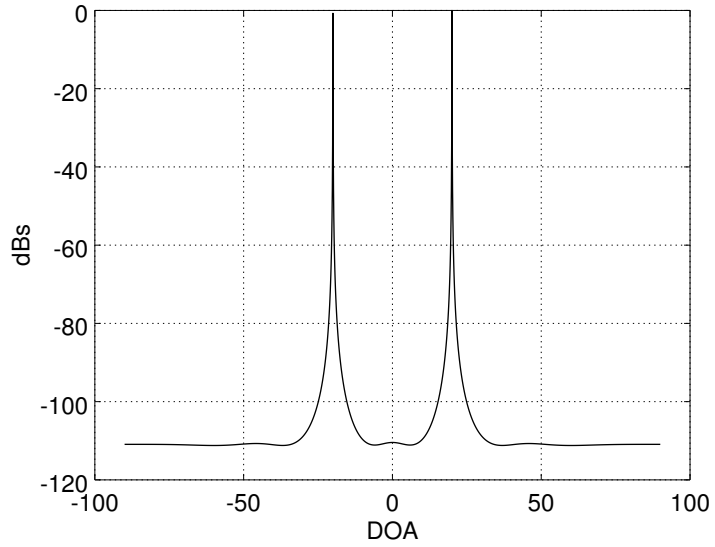



Figura 3.2: Espectro MUSIC. Simulación realizada con un arreglo lineal de 8 micrófonos distanciados a 4 cm entre si, con dos señales de banda angosta de frecuencia f=3500 Hz posicionadas en -20° y 20°.

Si suponemos P señales que inciden en el arreglo de M micrófonos, existirán M-P eigenvalores y eigenvectores asociados al ruido en el dominio de las señales, es decir, el número máximo de fuentes localizadas con este método es de M-1, porque siempre requiere un eigenvector ruidoso en el subespacio de la señal para comprobar la ortogonalidad del mismo con respecto a las direcciones de arribo.

### 3.2.2. MUSIC para señales de Banda ancha

El algoritmo MUSIC mencionado anteriormente es el más utilizado y citado en la bibliografía, sin embargo, cuando se trabaja con un arreglo de micrófonos las señales adquiridas tienen un contenido espectral mucho mayor comparadas con una señal de banda angosta. A pesar de que el ancho de banda de las señales de voz varía dependiendo de las características fisiológicas del hablante, definidas por la longitud y sensibilidad de las cuerdas vocales, resonadores (delimitados por la cavidad nasal, cavidad oral y faringe) y los articuladores (labios, dientes, mandíbula y paladares duro y blando), el ancho de banda promedio de las señales de voz es de

$4kHz$. Es importante tener en cuenta que durante la existencia de voz, no todas las frecuencias de dicha banda tienen contenido de la señal.

El algoritmo MUSIC para señales de banda ancha consiste en construir el espectro MUSIC de la Ecuación (3.17) en función de la frecuencia, ésto significa que el método se lleva a cabo en el dominio de la frecuencia utilizando previamente la transformada discreta de Fourier.

En la Sección 2.1.2 se describe el vector de direcciones que representa los respectivos retardos de incidencia de la señal emitida por una fuente en puntos estratégicos de adquisición, es decir, describe la respuesta de un arreglo de micrófonos. Dicho vector de direcciones está en función de la frecuencia de la señal, por lo que es necesario definir un vector de direcciones para cada valor de frecuencia de la señal, como se muestra en la Ecuación (3.18)

$$\mathbf{X}(\omega) = \mathbf{A}\big(\omega,\theta,\phi\big)\mathbf{S}(\omega) + N(\omega) \tag{3.18}$$

por lo que si extrapolamos la Ecuación (3.17) para cada una de las frecuencias, se tiene la siguiente expresión:

$$MUSIC(\omega,\theta) = \frac{1}{\mathbf{A}^H(\omega,\theta)\,\mathbf{E_N}(\omega)\,\mathbf{E_N}^H(\omega)\,\mathbf{A}(\omega,\theta)} \tag{3.19}$$

En la Figura 3.3 se muestra el espectro MUSIC para una señal de ruido blanco posicionada en $\theta = 50$ grados, utilizando un arreglo lineal uniforme de ocho micrófonos distanciados a 4 $cm$.

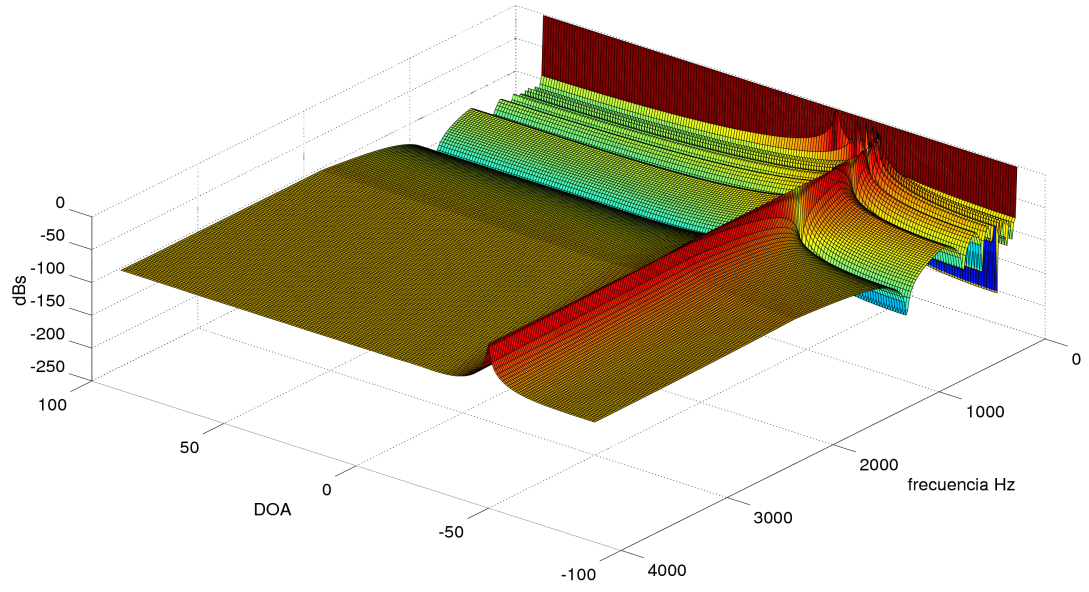



Figura 3.3: Espectro MUSIC para una fuente de ruido blanco posicionada a 50 grados.

### 3.2.3. Root-MUSIC

*Root-MUSIC* es un método que intenta reducir el costo computacional del algoritmo MUSIC encontrando las raíces de un polinomio, en lugar de graficar el espectro MUSIC y localizar los picos máximos del mismo [38].

El denominador de la Ecuación (3.17) se puede simplificar definiendo la matriz $\mathbf{C} = \mathbf{E_N}\mathbf{E_N}^H$, dando lugar a la expresión de *Root-MUSIC* como se muestra en la siguiente ecuación:

$$\mathbf{P_{Root-MUSIC}} = \frac{1}{\left|\mathbf{a}^H(\theta)\,\mathbf{C}\,\mathbf{a}(\theta)\right|} \tag{3.20}$$

Si consideramos un arreglo lineal uniforme de micrófonos, sabemos que el vector de direcciones estará definido como:

$$a(\theta) = e^{jkd(m-1)\sin\theta} \qquad para \qquad m = 1,2,\cdots,M-1$$

por lo que el denominar de la Ecuación (3.20) se puede escribir por medio de la Ecuación (3.21)

$$\mathbf{a}^H(\theta)\,\mathbf{C}\,\mathbf{a}(\theta) = \sum_{m=1}^{M}\sum_{n=1}^{M} e^{-jkd(m-1)\sin\theta} C_{mn} e^{jkd(m-1)\sin\theta} \tag{3.21}$$

$$= \textstyle\sum_{l=-M+1}^{M-1} c_l e^{jkdl\sin\theta}$$

donde $c_l$ es la suma de los elementos de la diagonal de $\mathbf{C}$. La suma de los elementos fuera de la diagonal siempre es menor que la suma de los elementos de la diagonal principal. Además $c_l = c_{-l}^*$. La Ecuación (3.21) se puede simplificar en forma de un polinomio con coeficientes $c_l$ como se muestra a continuación:

$$D(z) = \sum_{l=M-1}^{M-1} c_l z^l \tag{3.22}$$

donde $z = e^{-jkd\sin\theta}$.

El espectro MUSIC de la Ecuación (3.17) llega a ser equivalente al polinomio $D(z)$ de la Ecuación (3.22), de tal manera que los picos en el espectro MUSIC (direcciones de las fuentes) son las raíces del polinomio $D(z)$ que se encuentran más cercanas al circulo unitario [38], [39].

El orden del polinomio de la Ecuación (3.21) es de $2(M-1)$, es decir, que sus raíces complejas son $z_1, z_2, \cdots, z_{M-1}$ y están definidas como $z_i = |z_i|\, e^{j\,arg(z_i)}$ para $i = 1,2,\cdots,2(M-1)$, donde $arg(z_i)$ es la fase de $z_i$.

El ángulo que representa la dirección de arribo de las fuentes se calcula al comparar $e^{j\,arg(z_i)}$ con $e^{-jkd\sin\theta}$ por medio de la Ecuación (3.23).

$$\theta_i = -\sin^{-1}\left(\frac{arg(z_i)}{kd}\right) \tag{3.23}$$

## 3.3. Fundamentos del formador de haz

El formador de haz (*beamforming*) ha sido estudiado por muchas áreas de investigación como en radar, sonar, sismología, robótica y comunicaciones, entre otras. El formador de haz puede ser utilizado para diferentes propósitos tal como la determinación de la dirección de arribo (DOA), mejorar la señal deseada corrompida por ruido y reverberación, filtrado espacial y *coktail party* [10]. En general un formador de haz es un filtro espacial que opera en la salida de un arreglo de micrófonos, con el objetivo de crear un patrón de haz deseado; es decir, adquiere únicamente la señal que incide en una dirección específica, mientras que atenúa a las señales procedentes de las direcciones restantes, esto se logra realizando un direccionamiento del haz de forma electrónica. Un haz direccionado se forma a partir de aplicar retardos en las señales adquiridas antes de sumarlas [5].

El filtrado espacial consiste en dos etapas que son la sincronización y la sumatoria ponderada de las señales. La etapa de sincronización consiste en retrasar o adelantar una cantidad adecuada de tiempo a la salida de cada micrófono, de tal manera que las señales provenientes de una dirección sean sincronizadas. Para lograr esto, es necesario conocer previamente la dirección de arribo, de tal manera que se puedan calcular los retrasos adecuados y realizar la sincronización. La segunda etapa, como su nombre lo dice, consiste en ponderar las señales sincronizadas y posteriormente sumarlas para formar una sola salida. Las dos etapas juegan un papel importante en el proceso de *Beaforming* porque la sincronización controla la dirección del haz, mientras que la suma ponderada el ancho del haz del lóbulo principal y las características de los lóbulos laterales; sin embargo, el estudio del formador de haz se centra primordialmente en determinar los coeficientes para la suma ponderada. En algunos casos los coeficientes pueden ser calculados con base a un patrón de haz previamente especificado, tal es el caso del formador de haz fijo; sin embargo, un mejor enfoque es actualizar los coeficientes con base a las características de la señal y del ruido [5], [8], [10].

Los formadores de haz son desarrollados para señales de banda angosta, lo que significa que pueden ser bien caracterizados para una frecuencia. En el caso de las señales de banda ancha, es necesario caracterizar el formador de haz en todo el contenido espectral de la señal de interés. La desventaja del formador de haz para señales de banda ancha es que el ancho del formador de haz es inversamente proporcional al valor de la frecuencia, es decir, el ancho del formador incrementa para valores de frecuencia pequeños.

En las siguientes secciones se explica la teoría del formador de haz utilizada para determinar la dirección de arribo.

### 3.3.1. Formador de haz fijo

Los coeficientes de un formador de haz fijo $W$ son diseñados de manera permanente, de tal manera que las señales provenientes desde la dirección de la fuente de voz en el ángulo $\theta$ se mantienen en la salida sin distorsión mientras que los sonidos provenientes desde otras direcciones son suprimidos. Si la fuente de voz se desplaza, los coeficientes del formador de haz fijo se deben calcular de nuevo, para ello la fuente de voz requiere ser seguida utilizando un algorit-

mo de localización de fuentes de voz [21].

**Delay and Sum Beamforming**

El *Delay and Sum* (DS) es la estructura más simple para realizar selectividad espacial, y su teoría se origina a partir del procesamiento de arreglos de antenas de banda angosta. La finalidad de este tipo de formador de haz es enfatizar la señal que se localiza en la dirección de interés, mientras que atenúa las señales en otras direcciones. En este tipo de formador de haz se consideran los pesos unitarios, de tal manera que las ondas planas adquiridas por los sensores son únicamente retrasadas apropiadamente y al sumarse quedan exactamente en fase como se observa en la Figura 3.4(a).

En (3.24) se muestra la señal de salida direccionada del DS a partir del conocimiento previo de los retardos $\tau_m$ para cada micrófono $m$ [21], [10].

$$y(n) = \frac{1}{M}\sum_{m=0}^{M-1} x_m(n-\tau_m) \tag{3.24}$$

Sin embargo, la construcción del formador de haz fijo consta de una etapa de ponderación con coeficientes complejos para cada una de las señales alineadas, lo cual hace posible un patrón de radiación más uniforme [21], [5]. En la Ecuación (3.25) se muestra la señal de salida filtrada con un formador de haz fijo incluyendo la etapa de ponderación.

$$y(n) = \frac{1}{M}\sum_{m=0}^{M-1} w_m^* x_m(n-\tau_m) \tag{3.25}$$

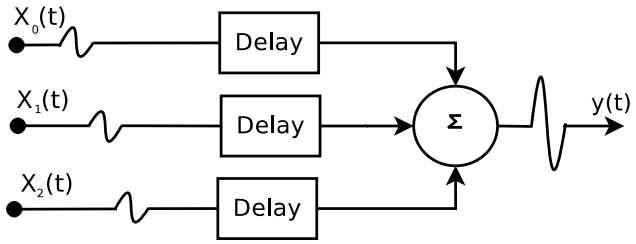
La Figura 3.4(b) muestra un diagrama de bloques de la configuración de un formador de haz fijo con coeficientes complejos $\mathbf{w}^*$ utilizando un arreglo lineal uniforme de $M$ micrófonos. Se observa que el direccionamiento del haz se realiza en el sentido del ángulo $\theta$, aplicando los respectivos retardos $\tau_m$ a cada una de las señales adquiridas y considerando al primer micrófono como el micrófono de referencia.

Las señales adquiridas por un arreglo de $M$ micrófonos se pueden representar con un vector en la forma $\mathbf{x}(n) = \begin{bmatrix} \mathbf{x_1}(n) & \mathbf{x_2}(n) & \mathbf{x_3}(n) & \dots & \mathbf{x_M}(n) \end{bmatrix}^T$, donde $\mathbf{x_m}(n)$ es la señal adquirida por medio del sensor $m$ en el tiempo discreto $n$, y el correspondiente vector de pesos $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_M \end{bmatrix}^T$, donde el superíndice $(\cdot)^T$ denota que es un vector transpuesto. El vector de salida complejo del formador de haz está dado por

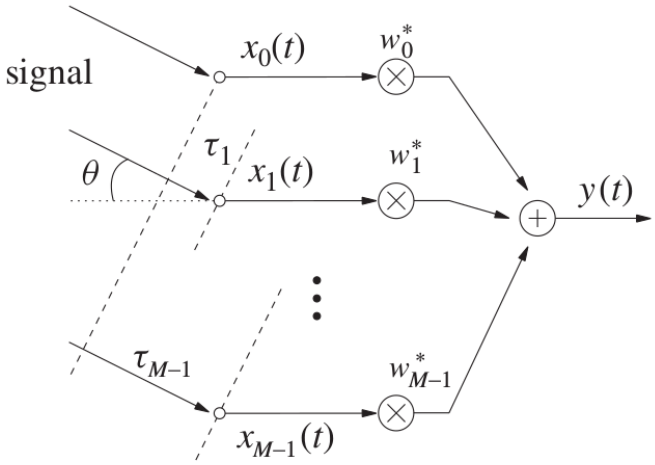
$$\mathbf{y}(n) = \mathbf{w}^H \mathbf{x}(n)\ , \tag{3.26}$$

donde el superíndice $(\cdot)^H$ hace referencia al vector conjugado transpuesto, conocido como Hermitiano.

El formador de haz y el filtro de respuesta finita al impulso (FIR) tienen un funcionamiento parecido, y se puede realizar una analogía del funcionamiento entre ellos cuando el formador de haz opera para una señal de banda angosta. Las entradas de ambos filtros son retrasos en




(a) Delay and sum Beamforming.




(b) Formador de haz fijo [6].

Figura 3.4: Diagrama de bloques de un formador de haz fijo.

tiempo de la señal de referencia multiplicados por pesos conjugados, mientras que la salida es la suma pondera de las señales, como se muestra en la Figura 3.5.

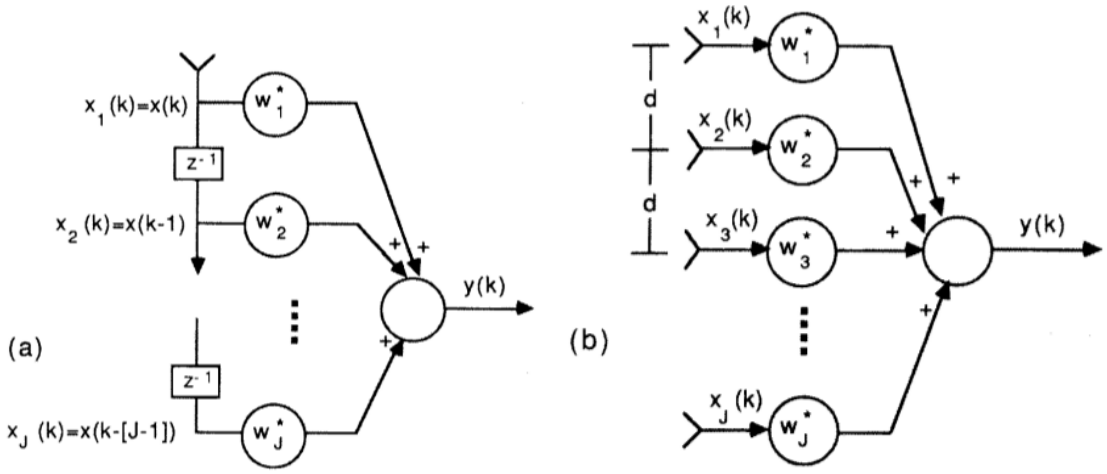
La respuesta en frecuencia de un filtro FIR con pesos $\mathbf{w}^*$, L etapas de retardo y un retardo de T segundos está dada por la Ecuación (3.27)

$$\mathbf{r}(\omega) = \sum_{l=1}^{L} w_l^* e^{-iwT(l-1)} . \tag{3.27}$$

De manera sintetizada se puede expresar como:

$$\mathbf{r}(\omega) = \mathbf{w}^H \mathbf{a}(\omega) \tag{3.28}$$

donde $\mathbf{r}(\omega)$ representa la respuesta de un filtro FIR para una señal compleja de banda angosta de frecuencia $\omega$, y $\mathbf{a}(\omega) = \begin{bmatrix} 1 & e^{j\omega T} & e^{j2\omega T} & \cdots & e^{j(L-1)\omega T} \end{bmatrix}^H$ es el vector que describe la fase de la señal sinusoidal en cada etapa del filtro [5]. De manera similar, el formador de haz se define como la amplitud y fase que representa una onda plana compleja en función de la posición y la frecuencia. Es importante considerar que la posición es una cantidad de tres dimensiones, sin embargo, se puede representar en función de los ángulos azimuth $\theta$ y elevación $\phi$.




(a) Estructura de un filtro FIR

(b) Estructura de un Formador de haz fijo de banda angosta

Figura 3.5: Comparación entre un formador de haz y un filtro FIR [5].

Suponiendo que se adquiere una onda plana compleja de banda angosta, con DOA $\theta$ y frecuencia $\omega$ por medio de un arreglo lineal de micrófonos, los retardos en cada uno de los micrófonos estarán dados por $\Delta_m$, que representa el tiempo de retardo en el m-ésimo micrófono respecto al primer micrófono, de tal manera que la señal en el micrófono $m$ se expresa como $\mathbf{x_m}(t) = e^{j\omega(t-\Delta_m)}$. Sustituyendo en la Ecuación (3.25) se obtiene:

$$y(n) = e^{j\omega t}\frac{1}{N}\sum_{m=0}^{M-1} w_m^* e^{-j\omega\Delta_m} = e^{j\omega t}\mathbf{r}(\theta,\omega) \tag{3.29}$$

donde $\mathbf{r}(\theta,\omega)$ es la respuesta del formador de haz y puede ser representado en forma de vector como:

$$\mathbf{r}(\theta,\omega) = \mathbf{W}^H\mathbf{a}(\theta,\omega) \tag{3.30}$$

Se observa que el vector $\mathbf{a}(\theta,\omega)$ corresponde a las exponenciales complejas $e^{-j\omega\Delta_m}$, y puede ser representado por la Ecuación (3.31)

$$\mathbf{a}(\theta,\omega) = \begin{bmatrix} 1 & e^{-j\omega\tau_1(\theta)} & e^{-j\omega\tau_2(\theta)} & \cdots & e^{-j\omega\tau_{M-1}(\theta)} \end{bmatrix} \tag{3.31}$$

### 3.3.2. Patrón de radiación

El patrón de radiación se puede considerar como la distribución relativa de la energía en función de coordenadas espaciales. En la mayoría de los casos, el patrón de radiación se evalúa considerando el modelo de campo lejano y se puede representar en coordenadas direccionales.

El patrón de radiación de un arreglo de micrófonos lo podemos clasificar en dos formas: patrón de radiación o factor de arreglo de la respuesta del filtro y patrón de respuesta de la dirección. A pesar de que su metodología en ambos casos es muy parecida, se utilizan para aplicaciones diferentes. El patrón de radiación modela la distribución de la energía con base a la

respuesta del filtro $\mathbf{r}(\theta,\omega)$ de la Ecuación (3.30), es decir, $\mathbf{r}(\theta,\omega)^2$ y se calcula para cada uno de los ángulos del intervalo de recepción. Es necesario tomar en cuenta que cada coeficiente $w$ afecta la respuesta temporal y espacial del formador de haz.

El patrón de radiación se direcciona por medio del vector de direcciones que está definido por la Ecuación (2.31), y particularmente cuando el ángulo de elevación $\phi = 90$, el vector de dirección se se puede expresar como se muestra en la Ecuación (2.46).

Si consideramos que los coeficientes de un formador de haz fijo son unitarios, es decir, un formador de haz *Delay and Sum* (DS), el factor de arreglo está definido por la Ecuación (3.32), donde $\lambda$ es la longitud de onda, $d$ la distancia entre micrófonos y $M$ el número total de micrófonos. El procedimiento para calcular dicha ecuación se puede observar en el Apéndice A que parte de la respuesta del filtro (Ecuación (3.27))

$$|H(\theta)| = \left| \frac{\sin\left(\frac{M\pi d \sin(\theta)}{\lambda}\right)}{\sin\left(\frac{\pi d \sin(\theta)}{\lambda}\right)} \right| \tag{3.32}$$

En la Figura 3.6 se muestra un patrón de radiación de un formador de haz DS de banda angosta direccionado en $\theta = 0$ grados y con frecuencia $2500Hz$. La simulación se realizó considerando un arreglo lineal con $M = 8$ micrófonos espaciados a $4cm$.

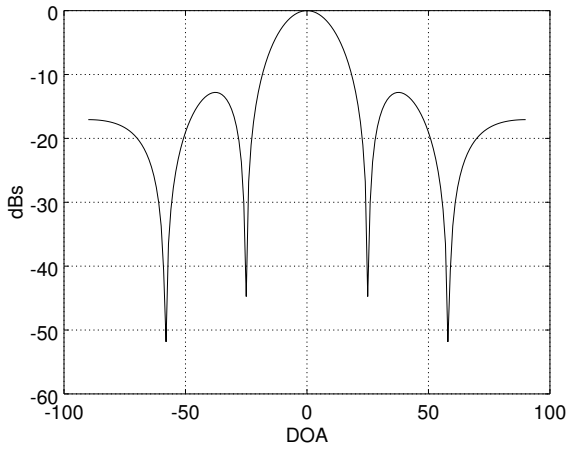
La figura 3.6(a) muestra el patrón de radiación, donde el eje de las abscisas representa la dirección del ángulo azimutal $\theta$, es decir, el direccionamiento del haz, y el eje de las ordenadas que representa la energía en decibeles. Con frecuencia, el patrón de radiación se grafica en coordenadas polares, como se muestra en la Figura 3.6(b), esta representación se utiliza para tener una mejor visualización espacial del comportamiento del haz.

Se puede observar que el patrón de radiación en coordenadas polares de la Figura 3.6(b) presenta un intervalo de $0 \leq \theta \leq 360$ mientras que el intervalo angular de la Figura 3.6(a) es de 90 a 270 grados pasando por cero. El patrón de radiación tiene una respuesta en el intervalo de 0 a 360 grados, sin embargo, para el ejemplo de la Figura 3.6 se utilizó un arreglo lineal de micrófonos, el cuál tiene un factor de arreglo que es simétrico con respecto al lóbulo principal como se muestra en la Figura 3.6(a). Además, es periódico en cualquier múltiplo de $\pi$, dicha periodicidad se puede observar en la Ecuación 3.32, por lo tanto, la Figura 3.6(b) muestra dos lóbulos principales, uno en $\theta = 0$ grados y el segundo en $\theta = 180$, es decir, el patrón de radiación generado en el intervalo $90 \leq \theta \leq 270$ es un patrón espejeado producido por el tipo de arreglo de sensores utilizado.

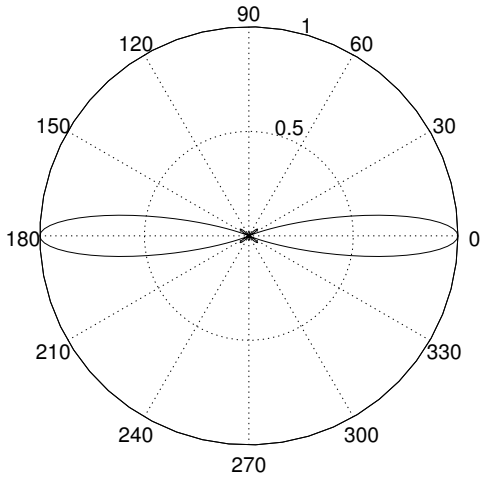
En cambio, el patrón de radiación generado para un arreglo circular de micrófonos tiene un intervalo de detección de $0 \leq \theta < 360$ grados, es decir que dicho patrón es periódico en $2\pi$, de tal manera que en la representación polar no presenta un patrón espejeado como es el caso del arreglo lineal de sensores.

El patrón de la respuesta de dirección muestra la distribución de la energía de las señales adquiridas por un arreglo de micrófonos en el intervalo angular de interés, es decir, realiza un

mapeo de la energía que inciden en los micrófonos del arreglo, mientras que el patrón de radiación modela la respuesta del filtro espacial a partir de los coeficientes **w** y el vector de direccionamiento.




(a) Patrón de radiación.




(b) Patrón de radiación en escala polar.

Figura 3.6: Patrón de radiación de un formador de haz tipo *Delay and Sum* de banda angosta.

De manera similar, el patrón de la respuesta de dirección calcula la energía para cada dirección $\theta$ del vector de direcciones, sin embargo, el patrón de la respuesta de dirección se calcula considerando la energía de la señal. A partir del modelo de la señal, determinado por las ecuaciones (2.30) y (2.33), se puede obtener la energía de las señales adquiridas por los micrófonos como se muestra en la Ecuación (3.33)

$$\mathbf{P}(\theta) = \sum_{m=0}^{M-1} \sum_{n=1}^{N} |w_m^*(\theta) x_m(n)|^2 = \mathbf{w}^H(\theta) \mathbf{R_{xx}} \mathbf{w}(\theta) \tag{3.33}$$

Para el formador de haz *Delay and Sum*, los coeficientes $w$ se consideran unitarios, de tal manera que $w(\theta) = a(\theta)$.

El vector de direcciones $\mathbf{a}(\theta)$ de la Ecuación (3.33) contiene todas las posibles direcciones en donde pudiera localizarse una fuente. Este patrón de respuesta de dirección se utiliza principalmente para localizar las posiciones de fuentes emisoras de señales, esperando que las posiciones reales $(\theta_1, \theta_2, ..., \theta_p)$ de las $p$ fuentes se determinen por medio de los índices de cada uno de los picos que existan en la gráfica del patrón de la respuesta de dirección. De esta manera se estiman las respectivas direcciones $\hat{\theta_1}, \hat{\theta_2}, ..., \hat{\theta_p}$.

Vamos a considerar un arreglo lineal de ocho micrófonos distanciados $4cm$ y dos fuentes que emiten señales sinusoidales, con frecuencia de 2500 $Hz$, posicionadas en $\theta_1 = -20$ y $\theta_2 = 50$ en el campo lejano. En la Figura 3.7 se muestra el patrón de la respuesta de dirección normalizado que fue construido a partir de la Ecuación (3.33) con las señales de banda angosta adquiridas por los micrófonos, donde se puede observar que existen dos picos predominantes que representan la estimación de la posición de cada una de las señales.

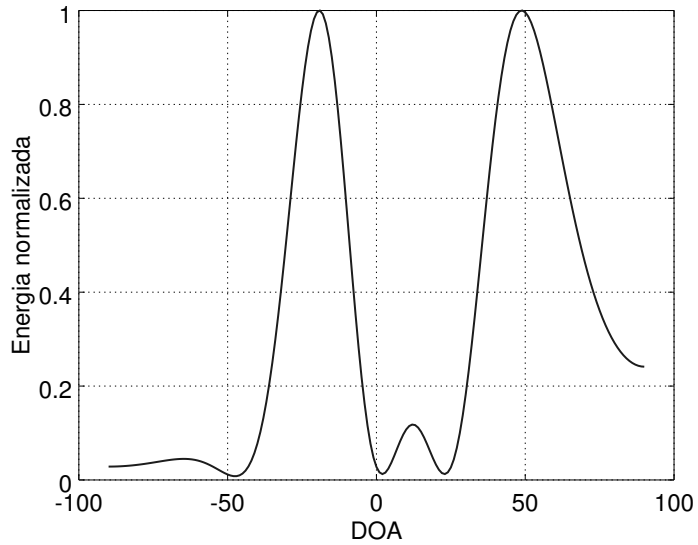



Figura 3.7: Patrón de respuesta de dirección de dos señales de banda angosta posicionadas en $\theta_1 = -20$ y $\theta_2 = 50$.

**Ancho del lóbulo pricipal del formador de haz**

Tanto el patrón de radiación como el patrón de la respuesta de dirección están formados por un lóbulo principal y múltiples lóbulos laterales. El lóbulo principal indica la dirección del haz, es decir, muestra la dirección donde se efectúa el filtrado espacial, o muestra la dirección de la mayor concentración de energía adquirida, mientras que los lóbulo laterales son producidos por las interferencias constructivas y destructivas de las señales alrededor de la fuente de interés.

El ancho del lóbulo principal del formador de haz es un parámetro importante que se debe considerar. Entre menor es el ancho del lóbulo principal, el formador de haz tendrá mejor resolución en la dirección deseada, es decir, el filtrado espacial será más selectivo y capaz de discriminar las múltiples señales provenientes desde otras direcciones. En el caso de la respuesta de dirección, el ancho de los lóbulos indican la direccionalidad de la fuente, por lo que entre menor es el ancho de los lóbulos, la resolución de detección será menor y se podrán estimar los DOA de un mayor número de fuentes.

En el caso de un formador de haz construido por medio de un arreglo lineal de micrófonos, el ancho del lóbulo principal se puede aproximar por medio del *Rayleigh beamwidth* en términos del ángulo de arribo [2], como se muestra en la Ecuación (3.34).

$$\theta_{BW} = \frac{1}{(d/\lambda)M\cos(\theta_0)} \tag{3.34}$$

donde $\lambda$ es la longitud de onda. El ancho del lóbulo es menor cuando $\theta_0 = 0$, mientras que aumenta cuando $\theta_0$ se aproxima a 90 y $-90$ grados. De la misma manera, el ancho del lóbulo principal del formador de haz disminuye cuando aumenta el número de micrófonos, así como de la distancia entre micrófonos y la frecuencia de la señal de banda angosta adquirida [10].

En un diseño de un formador de haz, se busca que el ancho del lóbulo principal sea lo más angosto posible para que el sistema tenga alta resolución en la detección de fuentes sonoras y filtrado espacial, de la misma manera, se busca que la amplitud del lóbulo principal sea lo mas grande posible con respecto a la amplitud de los lóbulos laterales. Para el caso del formador de haz clásico, la amplitud del lóbulo lateral mayor es de *13 dB* por debajo de la amplitud del lóbulo principal [2].

**Lóbulos gratinados**

La distancia entre micrófonos es inversamente proporcional al ancho del lóbulo principal, como se muestra en la Ecuación (3.34), lo que se pensaría que entre mayor sea la distancia $d$ de los micrófonos, se reducirá el ancho del lóbulo principal y de ésta manera tener un haz más nítido. Sin embargo, cuando la distancia entre micrófonos $d$ es muy grande aparece un fenómeno llamado aliasing espacial que genera réplicas del lóbulo principal a lo largo del patrón de radiación, de tal manera que dificulta el filtrado direccional obteniendo un direccionamiento del haz en más de una posición [21], [10], [2].

Los lóbulos gratinados son un problema importante en la localización de fuentes porque también puede existir aliasing espacial en el patrón de respuesta de dirección, apareciendo, al igual que en el patrón de radiación, los lóbulos gratinados. En la Figura 3.8 se muestra una gráfica del patrón de la respuesta de dirección utilizando la técnica de un formador de haz fijo como el de la Figura 3.4(b). En la simulación se utilizó un arreglo lineal uniforme de ocho micrófonos distanciados a $d = 2\lambda$ y una señal de banda angosta con frecuencia de $2kHz$ posicionada en $\theta = 0$ grados.

En la Figura 3.8 se pueden observar tres lóbulos que tienen la misma amplitud localizados

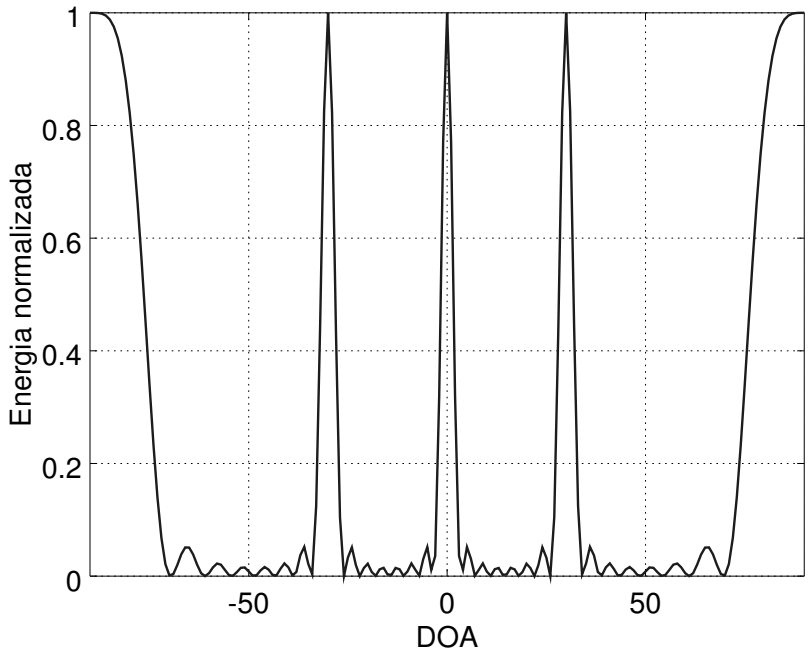



Figura 3.8: Formador de haz con un arreglo lineal de ocho micrófonos si $d = 2\lambda$ y una señal direccionada en $\theta = 0$ con $f = 2[kHz]$.

en $\theta_1 = -35$, $\theta_0 = 0$ y $\theta_2 = 35$, de los cuales el único que corresponde a la dirección de ángulo de arribo real de la señal es el $\theta_0$. Los lóbulos localizados en $\theta_1$ y $\theta_2$ son llamados lóbulos gratinados, los cuales son una réplica del lóbulo principal y se localizan de manera equidistante con respecto al lóbulo principal.

La aparición de los lóbulos gratinados en un patrón de respuesta de dirección dificulta la estimación de la dirección de arribo de una o múltiples señales porque no hay manera de identificar un lóbulo principal que represente la dirección de arribo de una fuente y un lóbulo gratinado, sobre todo cuando existen múltiples fuentes.

A partir de la Ecuación (3.32), se puede observar que el factor del arreglo es una función periódica de $\theta$ en cualquier múltiplo de $\pi$, de tal manera que si existe aliasign espacial los elementos del vector de direcciones son iguales, es decir:

$$e^{-j\omega \frac{2\pi m d \sin\theta_1}{\lambda}} = e^{-j\omega \frac{2\pi m d \sin\theta_2}{\lambda}}$$

Para evitar generar aliasing espacial, se debe cumplir con la siguiente condición:

$$\left| \frac{2\pi d \sin\theta}{\lambda} \right|_{\theta=\theta_1=\theta_2} < \pi \tag{3.35}$$

donde únicamente puede existir un máximo (m=1) dentro del intervalo $-\theta_{max} \leq \theta \leq \theta_{max}$. Para un arreglo lineal de micrófonos el intervalo esta definido como $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$, por lo que $|\sin\theta_{max}| = 1$. Considerando lo anterior, la expresión (3.35) queda:

$$d \leq \frac{\lambda}{2} = \frac{c}{2f} \tag{3.36}$$

donde $d$ es la distancia entre micrófonos, $\lambda$ la longitud de onda mínima de la señal, $f$ la frecuencia máxima de la señal y $c$ la velocidad de propagación del sonido.

## 3.4. Formador de haz de banda ancha

El formador de haz explicado en la Sección 3.3.1 trabaja adecuadamente para señales de banda angosta, sin embargo, su rendimiento se degrada conforme se incrementa el ancho de banda.

Las señales de banda ancha consisten de un número infinito de diferentes componentes frecuenciales, de tal manera que los coeficientes deberían ser diferentes en cada frecuencia, por lo que podemos escribir el vector de coeficientes de la siguiente manera:

$$\mathbf{w}(\omega) = \begin{bmatrix} \mathbf{w_0}(\omega) & \mathbf{w_1}(\omega) & \cdots & \mathbf{w_{M-1}}(\omega) \end{bmatrix}^T \tag{3.37}$$

Los coeficiente dependientes de la frecuencia se pueden utilizar en una línea de retardos para cada micrófono en estructura de filtros FIR o IIR. La estructura de retardos de línea y filtros FIR realizan un proceso de filtrado temporal para formar una respuesta de frecuencia dependiente de la señal de banda ancha en cada sensor y compensar la diferencia de fase en diferentes componentes de frecuencia [6]. En la Figura 3.9 se muestra la estructura utilizada para un formador de haz de banda ancha; dicha arquitectura tiene una respuesta de espacio tiempo.

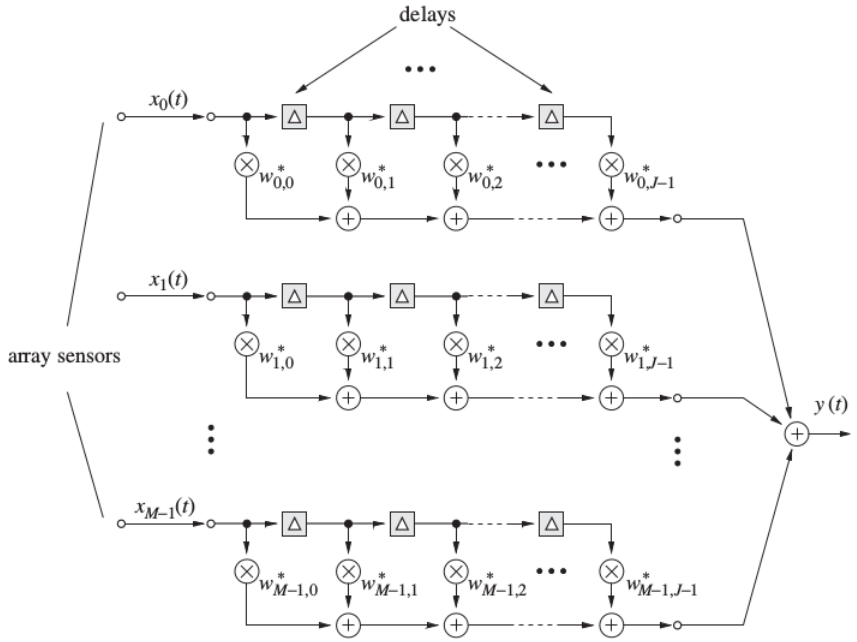



Figura 3.9: Estructura de un formador de haz para señales de banda ancha [6].

La salida del formador de haz de banda ancha mostrado en la Figura 3.9 se describe con la Ecuación (3.38)

$$y(t) = \sum_{m=0}^{M-1} \sum_{i=0}^{J-1} x_m(t - iT_s)\, w^*_{m,i} \tag{3.38}$$

donde $J-1$ es el número de retardos utilizados en cada micrófono $m$, $M$ el número total de micrófonos y $T_s$ es el retardo en cada una de las etapas. Se observa que si utilizamos la notación en forma de vectores, la salida del formador de haz se expresa de la misma manera que en la Ecuación (3.26) con la diferencia de que los vectores $\mathbf{x}(\mathbf{t})$ y $\mathbf{w}$ tienen la siguiente forma:

$$\mathbf{w} = \begin{bmatrix} \mathbf{w_0} \\ \mathbf{w_1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{w_{J-1}} \end{bmatrix} \tag{3.39}$$

donde cada vector $\mathbf{w_i}$ contiene los coeficientes conjugados complejos para cada micrófono como se muestra a continuación

$$\mathbf{w_i} = \begin{bmatrix} w_{0,i} & w_{1,i} & \cdots & w_{M-1,i} \end{bmatrix}^T \tag{3.40}$$

El vector $\mathbf{x}$ está dado por:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x_0}(t) \\ \mathbf{x_1}(t - T_s) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x_{J-1}}(t - (J-1)\, T_s) \end{bmatrix} \tag{3.41}$$

donde $\mathbf{x_i}(t - iT_s)$ corresponde a la i-ésima etapa de retardos en una línea correspondiente a un micrófono, por lo que finalmente:

$$\mathbf{x}(t - iT_s) = \begin{bmatrix} x_0(t - iT_s) & x_1(t - iT_s) & \cdots & x_{M-1}(t - iT_s) \end{bmatrix}^T \tag{3.42}$$

Cabe destacar que el formador de haz para señales de banda angosta es un caso particular del formador de haz de banda ancha cuando $J = 1$.

Otra manera de crear un formador de haz de banda ancha es diseñar múltiples arreglos de micrófonos traslapados con diferentes distancias entre micrófonos, como se muestra en la Figura 3.10, de tal manera que cada subarreglo corresponda a un grupo de frecuencias del ancho de banda total [7].

La desventaja de este método es que utiliza un gran número de micrófonos, por lo tanto implica un costo computacional grande dificultando la implementación en tiempo real [40].

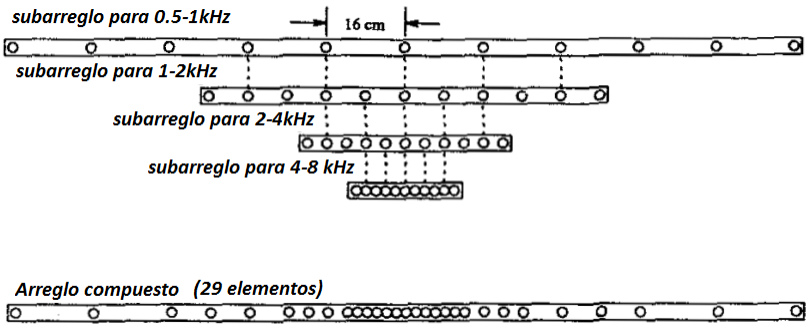



Figura 3.10: Arreglo lineal no uniforme de micrófonos diseñado para cuatro sub bandas [7].

### 3.4.1. Formador de haz en el dominio de la frecuencia

El formador de haz también se puede llevar a cabo en el dominio de la frecuencia. Cuando hablamos sobre el dominio de la frecuencia significa que se requiere realizar una transformación, de cada una de las señales adquiridas por los micrófonos, del dominio del tiempo al dominio de la frecuencia vía la transformada discreta de Fourier (DFT).

El formador de haz de banda ancha se realiza con un formador de haz para cada una de las frecuencias de interés y calculando sus respectivos coeficientes complejos para cada frecuencia. Finalmente, la señal filtrada en la dirección de interés se obtiene calculando la transformada inversa de Fourier, como se observa en la Ecuación (3.43)

$$y(n) = \mathscr{F}^{-1}\left[\sum_{b=0}^{B}\sum_{m=1}^{M} W_{m,b}^{*} X_m(\omega_b)\right] \tag{3.43}$$

donde $\mathscr{F}^{-1}$ denota la operación inversa de la transformada de Fourier (IDFT), $X_m(\omega_b)$ es el valor de la componente frecuencial $\omega_b$ de la señal $x(n)$ en el micrófono $m$ y $W_{m,b}^{*}$ el coeficiente complejo conjugado asignado a la frecuencia $\omega_b$ en el micrófono $m$.

En la Figura 3.11 se muestra el diagrama de bloques de la estructura de un formador de haz de banda ancha en el dominio de la frecuencia.

De la misma manera, la respuesta de un patrón de radiación de un formador de haz de banda ancha está en función de la frecuencia, por lo que cada frecuencia tendrá una respuesta particular dependiendo de los coeficientes calculados. La Ecuación (3.34) explica que el ancho del lóbulo principal del patrón de radiación es inversamente proporcional al valor de la frecuencia de trabajo, de tal manera que el ancho del lóbulo principal de un patrón de radiación en las frecuencias bajas es muy grande y no se realiza adecuadamente el filtrado espacial.

En la Figura 3.12 se muestra un patrón de radiación de un *Delay and Sum* direccionado en cero grados para señales de voz (con un ancho de banda de $4kHz$). En la simulación se utilizó un arreglo lineal de ocho micrófonos distanciados a $4\,cm$. Se puede observar que en el ancho de banda con intervalo de frecuencias $0-1000Hz$, el ancho del lóbulo principal es muy grande de tal manera que cubre todo el intervalo de apertura ($-90$ *a* 90 grados), es decir, el filtro ya no

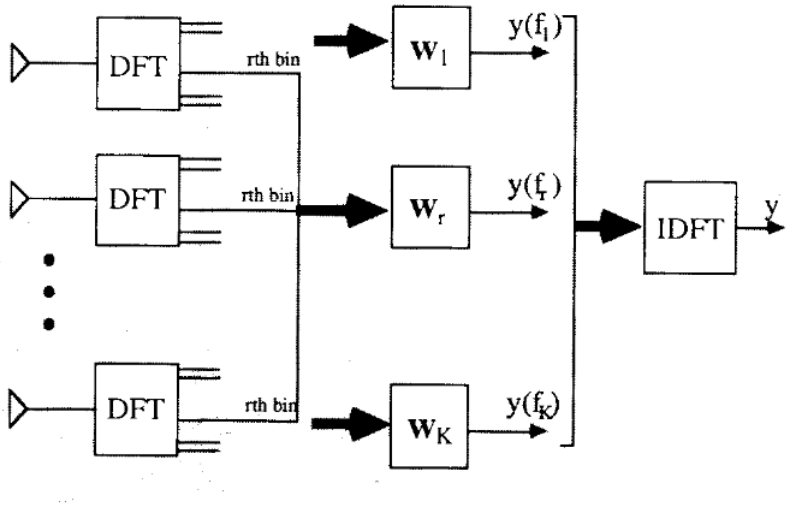



Figura 3.11: Formador de haz en el dominio de la frecuencia [5].

realiza un direccionamiento.

El patrón de la respuesta de dirección en el dominio de la frecuencia se obtiene de manera similar que en la Ecuación (3.33), sin embargo, la matriz de covarianza estará dada en función de la frecuencia. Para determinar la matriz de covarianza en el dominio de la frecuencia vamos a considerar que $\mathbf{X}(\omega)$ es el vector que representa a la señal de entrada $\mathbf{x}(n)$ en el dominio de la frecuencia discreta. De tal manera que la matriz de covarianza en el dominio de la frecuencia se puede calcular como se muestra en la Ecuación (3.44)

$$\mathbf{R}(\omega) = \mathbf{X}_{\omega}\mathbf{X}_{\omega}^{H}, \tag{3.44}$$

donde $\omega$ representa el índice de la frecuencia, $\mathbf{X}_{\omega}$ es un vector columna que está formado por las señales de cada micrófono $m$ en el dominio de la frecuencia, esto es:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_0(\omega) \\ \mathbf{X}_1(\omega) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{X}_{M-1}(\omega) \end{bmatrix} \tag{3.45}$$

Finalmente, el patrón de respuesta de dirección en función de la frecuencia se puede expresar como se muestra en (3.46).

$$\mathbf{P}(\omega,\theta) = \mathbf{A}^{H}(\omega,\theta)\,\mathbf{R}_{\mathbf{xx}}(\omega)\,\mathbf{A}(\omega,\theta) \tag{3.46}$$

donde el vector $\mathbf{A}^{H}$ contiene todas las posibles direcciones $\theta$ en las que puede localizarse una fuente de interés.

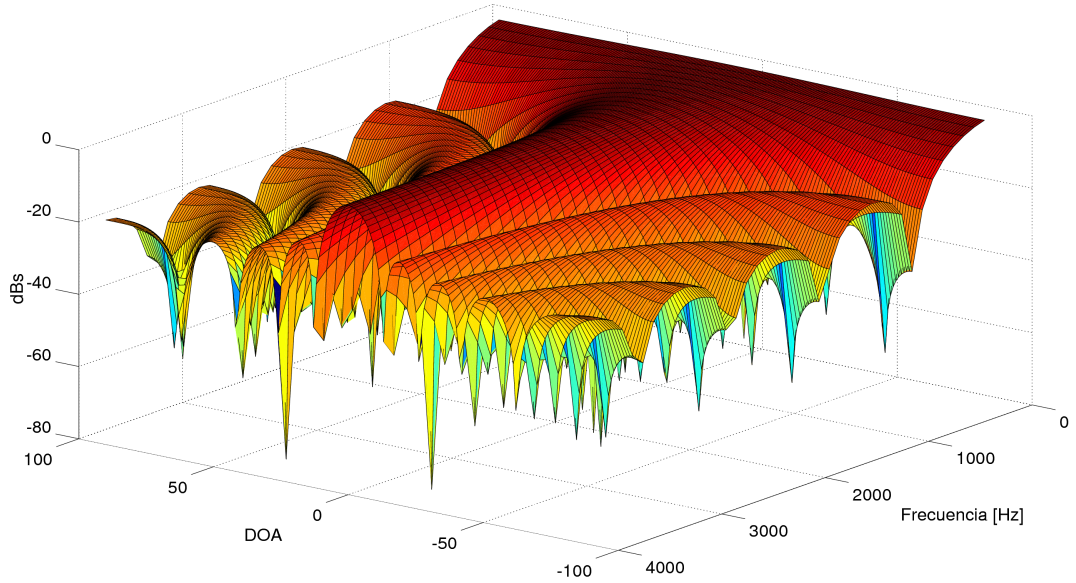



Figura 3.12: Patrón de radiación de un formador de haz de banda ancha.

Vamos a considerar dos señales con el mismo contenido frecuencial dentro de un ancho de banda de 8 $kHz$ posicionadas en el campo lejano en las direcciones $\theta_1 = 50$ y $\theta_2 = -20$ grados respectivamente, un arreglo lineal de ocho micrófonos distanciados a 4 $cm$ y una tasa de muestreo de $fs = 16000$ $Hz$. El patrón de respuesta de dirección en función de la frecuencia se muestra en la Figura 3.13, donde se puede observar que para cada elemento de contenido frecuencial de las señales adquiridas va a presentar un valor máximo de energía en el ángulo de la dirección de la señal, mientras que el valor de la energía en las frecuencias donde no exista señal, la amplitud dependerá del ruido presente en dicha frecuencia.

En la desigualdad de la Ecuación (3.36) se muestra la distancia máxima entre micrófonos para que no exista el aliasing espacial, sin embargo, la distancia entre micrófonos utilizada en la simulación de la Figura 3.13 está calculada de tal manera que el ancho de banda de trabajo sea de 4 $kHz$, es decir, inferior al ancho de banda de las señales, esto significa que para frecuencias superiores a 4 $kHz$ comienza a aparecer el aliasing espacial y dificulta la estimación de la dirección de arribo de las señales por los lóbulos gratinados, no obstante, la estimación de la dirección de arribo se puede realizar por medio de las posiciones de los máximos valores de energía para frecuencias inferiores a $4kHz$. También se puede observar que en el intervalo de frecuencias inferiores a 1000 $Hz$ el ancho de los lóbulos principales es muy amplio mezclándose los lóbulos de ambas señales, dificultando la identificación del número de señales y su posición.

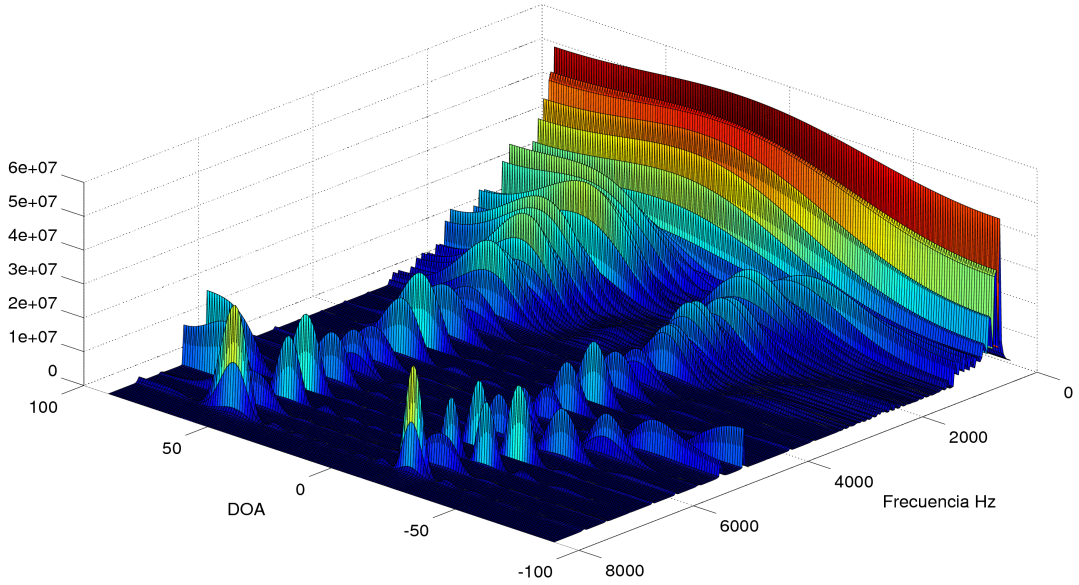



Figura 3.13: Respuesta de dirección de dos señales de banda ancha posicionadas en $\theta_1 = -50$ y $\theta_1 = 20$.

## 3.5. Filtro de respuesta sin distorsión de mínima varianza (MVDR)

El filtro de respuesta sin distorsión de mínima varianza se adapta dependiendo de los valores de entrada y se obtiene por la optimización de un criterio con restricción, minimizando el ruido de la señal ruidosa sin distorsionar a la señal deseada [8],[41]. El problema de optimización se basa en la siguiente expresión:

$$\min_{w} \mathbf{w}^H(\theta)\mathbf{R_{xx}}\mathbf{w}(\theta) \quad sujeto \quad a \quad \mathbf{w}^H\mathbf{a}(\theta) = 1 \tag{3.47}$$

La idea básica de esta técnica es seleccionar los coeficientes del filtro que minimizan la potencia aportada por el ruido y las señales procedentes de cualquier dirección con excepción de la dirección $\theta$, mientras que mantiene con una ganancia fija a la señal en la dirección de interés $\theta$.

La Ecuación (3.47) se puede resolver a través del método de multiplicadores de Lagrange, de tal manera que se obtienen los coeficientes óptimos $w_{opt}(\theta)$ por medio de la Ecuación (3.48) [8], [10].

$$\mathbf{w_{opt}}(\theta) = \frac{\mathbf{R_{xx}^{-1}}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{R_{xx}^{-1}}\mathbf{a}(\theta)} \tag{3.48}$$

De tal manera que el algoritmo MVDR calcula el grupo de coeficientes óptimos basado en las muestras de datos obteniendo un patrón de haz que suprime la respuesta de las señales en las direcciones donde hay fuentes de interferencia.

Sustituyendo la Ecuación (3.48) en (3.33), se obtiene el espectro espacial con los coeficientes óptimos $\mathbf{w_{opt}}$ como se muestra en la Ecuación (3.49) [23], [10] es:

$$\mathbf{P_c}(\theta) = \mathbf{w_{opt}}^H(\theta)\,\mathbf{R_{xx}}\mathbf{w_{opt}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\,\mathbf{R_{xx}^{-1}}\,\mathbf{a}(\theta)} \tag{3.49}$$

Vamos a considerar una fuente de banda angosta con frecuencia $f = 1500\,Hz$ posicionada en $\theta = 30$ en el campo lejano con respecto a un eje de referencia y un arreglo lineal de ocho micrófonos distanciados a $d = 5\,cm$ posicionado en el origen del eje de referencia. El espectro espacial de salida MVDR se muestra en la Figura 3.14.

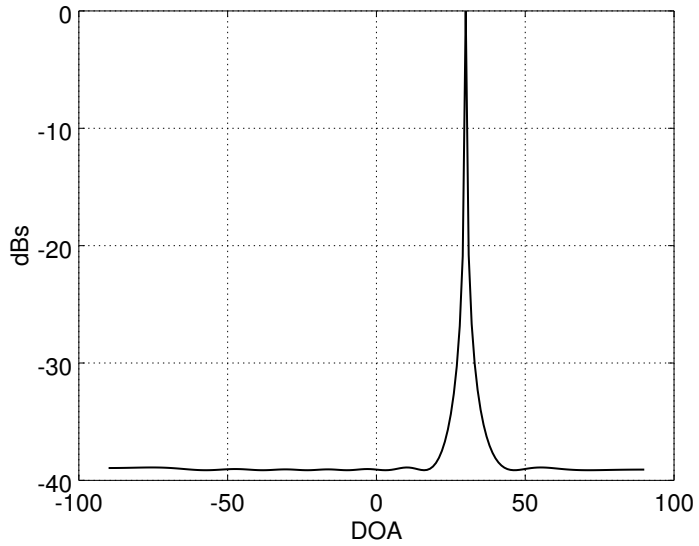



Figura 3.14: Espectro espacial MVDR de banda angosta.

## 3.6. Resumen

En éste capítulo se estudiaron y analizaron los métodos de estimación de dirección de arribo para señales de banda angosta y se extendió la teoría para señales de banda ancha. El método de diferencia del tiempo de arribo (TDOA) es uno de los métodos más utilizado en la literatura para localización de fuentes, principalmente por su fácil implementación y rápida respuesta, sin embargo, éste ha mostrado problemas con la estimación cuando las fuentes de voz están inmersas en ruido acústico aditivo, en especial al utilizar la correlación cruzada generalizada con transformada de fase (GCC-PHAT).

El algoritmo MUSIC, a pesar que requiere un costo computacional elevado, es un método muy utilizado para la estimación de la posición de las fuentes y el número de las mismas. La razón principal es por su alta resolución, es decir, la capacidad de estimar la posición de las fuentes cuando éstas se encuentran relativamente cerca; sin embargo, en la teoría se considera que las señales de las fuentes no tienen correlación entre sí y con el ruido, respectivamente, por

lo que dificulta su implementación en un ambiente acústico no controlado.

El tercer enfoque para determinar la dirección de arribo es el formador de haz. El formador de haz es usado principalmente para realizar filtrado espacial, sin embargo, es posible determinar la estimación del ángulo de arribo por medio de la respuesta de dirección que se que genera en la salida de los micrófonos del arreglo; tal respuesta de dirección se determina de la misma manera que el patrón de radiación, sin embargo, a diferencia del patrón de respuesta de dirección, el patrón de radiación sirve para modelar la respuesta del filtro. Existen muchos tipos de formador de haz, sin embargo, para calcular la dirección de arribo se requiere el patrón de respuesta de dirección. Tanto el patrón de radiación, como el patrón de respuesta de dirección generan un lóbulo principal muy ancho y entre menor es el ancho del lóbulo, es más fácil determinar el ángulo de arribo de las fuentes cuando están muy cercanas cercanas y el filtrado espacial es más selectivo. Para resolver esto, el filtro de respuesta sin distorsión de mínima varianza (MVDR), genera un patrón con un ancho en el lóbulo principal muy reducido, con la desventaja de que tiene un costo computacional elevado, dificultando la implementación de éste en aplicaciones de tiempo real, además para fuentes simultáneas dicha respuesta de dirección se degrada.

Todos las métodos de estimación de la dirección de arribo tienen sus propias ventajas y desventajas y funcionan muy bien bajo ciertas restricciones, sin embargo, no todos tienen una generalización para implementarlos en situaciones reales, es decir, en condiciones no controladas. En el presente trabajo se realizó una combinación de las técnicas de eigendescomposición de la matriz de covarianza con el formador de haz aprovechándose de la correlación entre señales que existe en un ambiente no controlado y construyendo la respuesta de dirección con base a la distribución de la energía en el intervalo de medición como se describió en la teoría del formador de haz.

# Capítulo 4.

# Diseño e implementación del sistema

Los sistemas de dirección de arribo se clasifican principalmente en dos grupos, los sistemas activos y pasivos. Los sistemas activos (como el radar y el sonar) cuentan con una etapa de emisión de señales con características específicas, una etapa de adquisición y una de procesamiento. La señal adquirida por el arreglo de sensores es una reflexión de la señal emitida, tal reflexión se genera en la superficie de un objeto localizado a una distancia $d_{objeto}$ y un ángulo $\theta_{objeto}$, por lo que se puede estimar la posición de dicho objeto realizando el procesamiento de las señales adquiridas comparándolas con la señal emitida. Los sistemas pasivos, a diferencia de los activos, se enfocan en determinar la dirección de arribo de fuentes emisoras de sonido, por lo que no realizan ninguna emisión de señales.

Los sistemas pasivos que estiman la dirección de arribo son ampliamente utilizados en las áreas de robótica, para robots de servicio y robots de rescate, navegación, videoconferencias y sistemas de seguridad; sin embargo, su ejecución en tiempo real no es tarea sencilla en concecuencia a la demanda de recursos computacionales requeridos.

En el presente trabajo se implementó un sistema pasivo que estima la dirección de arribo de señales provenientes de fuentes voz estáticas y con desplazamiento. En la Figura 4.1 se muestra

un diagrama de bloques de la implementación del sistema. El sistema cuenta con una etapa de adquisición de señales por medio de un arreglo de micrófonos, sin embargo, se puede realizar la estimación de dirección de arribo con un banco de señales previamente adquiridas. Se implementó un detector de actividad de voz (VAD) adaptable, de tal manera que si la potencia del ruido acústico en el ambiente incrementa, el VAD sea capaz de detectarlo.

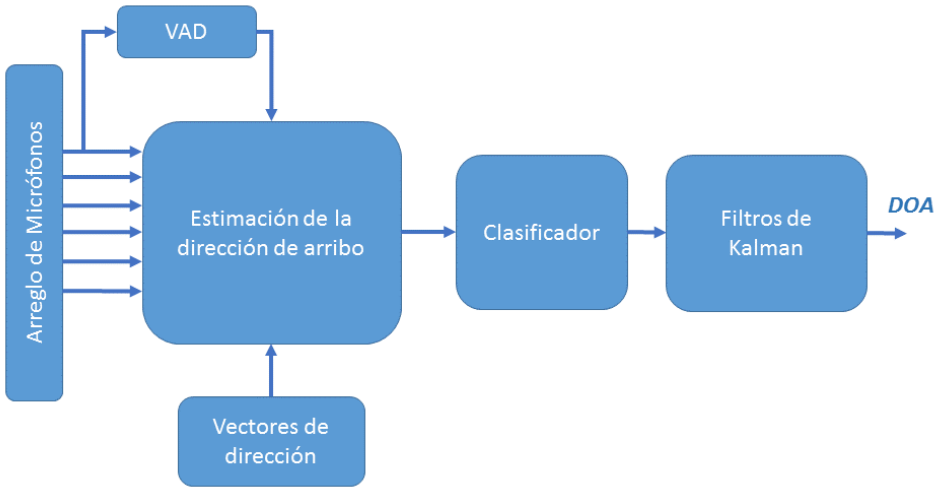



Figura 4.1: Diagrama de bloques del sistema implementado.

En este capítulo se explica el diseño e implementación de un sistema que realiza la estimación de dirección de arribo de señales provenientes de fuentes de voz inmersas en un ambiente ruidoso y con reverberación, todo esto en tiempo real. El sistema se aprovecha de la correlación que existe entre las señales para realizar la estimación, efectuando una eigendescomposición como se explica en la teoría del método MUSIC y generando una respuesta de dirección por medio del formador de haz fijo.

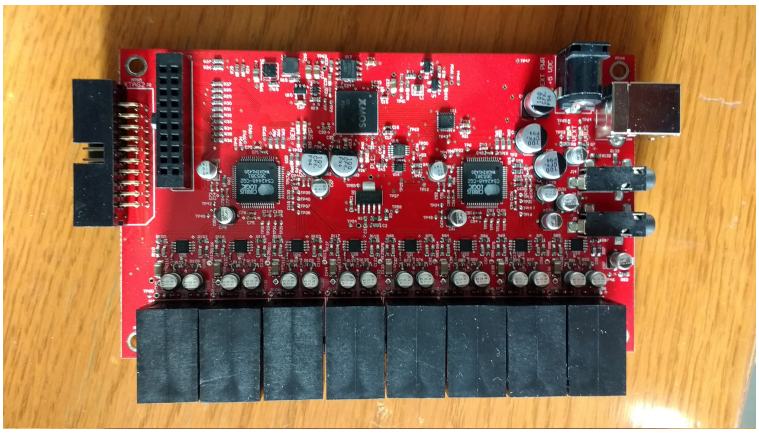
El capítulo comienza explicando la etapa de adquisición de las señales, utilizando la tarjeta embebida *8-Sound USB* y el software *jack audio connection kit*, para trabajar ya sea con las señales adquiridas por un arreglo de micrófonos o con un banco de señales previamente adquiridas. Posteriormente, se explica la implementación del método propuesto basado en la teoría de MUSIC y el formador de haz fijo. En la Sección 4.3 se propone un clasificador estadístico con el objetivo de asignar el posible DOA obtenido de una fuente potencial a la fuente de voz estimada anteriormente y, finalmente, en la sección 4.4 se explica la implementación de un filtro de Kalman para estimar la posición real de la fuente con base en los DOAs ruidosos que se determinan por el algoritmo de estimación dirección de arribo.

## 4.1. Adquisición de señales

Se utilizó la tarjeta de adquisición de señales *8-SoundUSB* con sus respectivos micrófonos para capturar las señales acústicas. La tarjeta de adquisición *8-SoundUSB* fue diseñada en la

Universidad de *Sheerbrook* en Canadá en el proyecto *ManyEars*, con fines de uso en procesamiento de voz para el área de robótica [42]. La tarjeta tiene la capacidad de adquirir hasta ocho señales simultáneas con una frecuencia de muestreo mímima de $fs_{min_{tarjeta}} = 44100\ Hz$ y máxima de $fs_{max_{tarjeta}} = 192\ KHz$; además, tiene una resolución de muestreo de 16 o 24 bits. De la misma manera, cuenta con dos salidas analógicas de tipo estéreo y un procesador tipo *XMOS XS1-L2* que ejecuta como máximo 1000 $MIPS$; sin embargo, en el presente trabajo se utiliza la salida tipo *USB* para la transferencia de datos de las señales con una computadora y efectuar el procesamiento en dicha computadora. Los micrófonos utilizados son de tipo *electret* con directividad onmidireccional. En la Figura 4.2 se muestran los dispositivos utilizados.


(a) Micrófono *electret* con directividad onmidireccional.



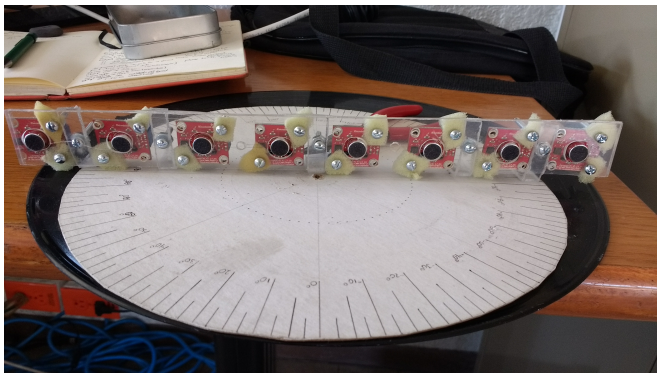
(b) Tarjeta de adquisición de señales utilizada (*8-SoundUSB*).

Figura 4.2: Dispositidos utilizados para la adquisición de las señales.

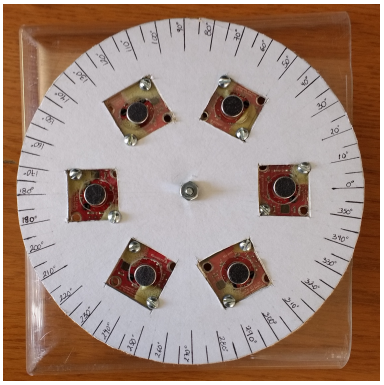
Se implementaron dos geometrías diferentes de arreglos de micrófonos, un arreglo lineal uniforme y un arreglo planar de tipo circular, ambos arreglos con seis micrófonos. La distancia entre micrófonos se determinó a partir del ancho de banda de las señales de voz (4 $KHz$). Para el caso del arreglo lineal uniforme de micrófonos, la distancia entre los mismos se calculó por medio de la Ecuación (3.36), donde la frecuencia máxima es $f_{max} = 4000\ Hz$, de tal manera que la distancia entre los micrófonos del arreglo lineal es de $d = 4\ cm$. El arreglo circular de micrófonos tiene sus elementos de forma equidistante, por lo que cada uno está posicionado cada 60 grados a partir del eje de refencia en $\theta = 0$ grados, es decir, $\theta_m = (m-1)60$ para $m = 1, 2, \cdots, M$.

Para evitar el aliasing espacial, el radio de curvatura de la circunferencia se calculó por medio de la Ecuación (2.52), dando como resultado $r = 0.42\ cm$. En la Figura 4.3 se muestran dos fotografías de los arreglos de micrófonos implementados, donde se puede observar que en ambos casos se utilizó un transportador para localizar la fuente de voz en la posición real.

Se utilizó el programa *JACK Audio Connection Kit* para poder controlar los puertos de entrada y salida de la tarjeta de adquisición de datos con el programa en lenguaje C. *Jack Audio Connection Kit* es un servidor de audio de baja latencia que puede conectar varias aplicaciones tipo *client* a un dispositivo de audio, además puede compartir audio entre varias aplicaciones de forma independiente, de tal manera que se pueden ejecutar diferentes tareas al mismo tiem-



(a) Arreglo lineal uniforme de micrófonos.



(b) Arreglo circular de micrófonos.

Figura 4.3: Arreglos de micrófonos implementados.

po sin ser afectadas [43].

La frecuencia de muestreo en un proceso con *Jack Audio Connection Kit* se puede elegir entre las siguientes: 22050 Hz, 32000 Hz, 44100 Hz, 48000 Hz, 88200 Hz, 96000 Hz y 192000 Hz, sin embargo, entre mayor es la frecuencia de muestreo, se tiene un mayor número de datos para procesar en cada tiempo, lo que significa que se requiere un mayor costo computacional y ésto afecta su ejecución en tiempo real. La frecuencia mínima de muestreo para señales de voz es de $fs_{min} = 8000\ Hz$ según el teorema de Nyquist, por lo que la frecuencia de muestreo utilizada en el presente proyecto es de $fs = 44100\ Hz$, dicho valor se determinó a partir de la tasa mínima que ejecuta la tarjeta *8-SoundUSB*.

La función *callback* del programa en lenguaje C es la encargada de la transferencia de datos y el procesamiento de señales digitales en línea, además hay que tener en cuenta que todos los tipos de datos dentro de Jack son de 32 bits.

Finalmente, para realizar la conexión entre *Jack* y un programa, se requiere seguir los siguientes pasos:

1. Llamar la función *jack-client-open()* para conectar el programa ejecutado con *Jack Audio Connection Kit*.
2. Activar los puertos de entrada y salida para habilitar los datos que van a ser transferidos al programa.
3. Registrar una función *callback* que será llamada cada tiempo por el servidor *Jack*.
4. Notificar a *Jack* que la aplicación está lista para la transferencia de datos.

Además, al utilizar *Jack Audio Connection Kit* es posible enlazar como entradas al sistema las señales adquiridas por el arreglo de micrófonos, o un conjunto de señales que se encuentran almacenadas en alguna base de datos. En ambos casos el sistema funciona en línea, de tal manera que por cada *frame, Jack Audio Connection Kit* ingresa un paquete de 1024 datos en cada canal

de entrada para procesar y reproduce un paquete de datos de la misma longitud. En el presente trabajo se realizó una serie de experimentos grabando señales de fuentes de voz posicionadas en diferentes ángulos de arribo $\theta$, tanto fuentes estáticos como fuentes en movimiento, dichas grabaciones se realizaron con los arreglos de micrófonos mencionados anteriormente. En la Figura 4.4 se muestran las dos formas de ejecutar el sistema con *Jack Audio Connection Kit*.

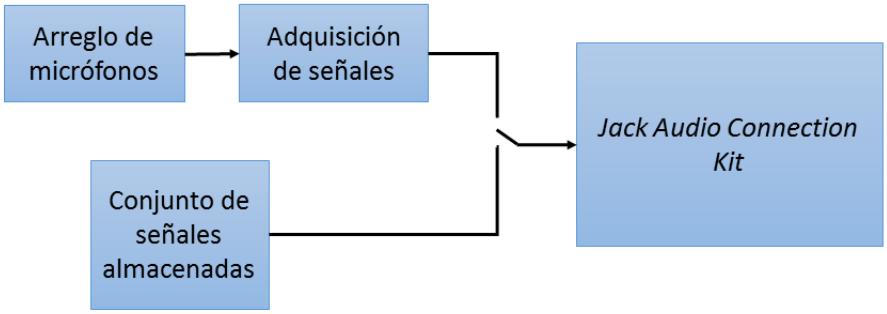



Figura 4.4: Selección de adquisición de señales utilizando *jack Audio- Connection Kit*.

## 4.2. Método implementado

La primer etapa del sistema es un detector de actividad de voz, la cual se implementó con un algoritmo seguidor de ruido y con base en la energía de las señales adquiridas a través de uno de los micrófonos del arreglo. Al iniciar el programa, la etapa de detección de actividad de voz calcula el umbral inicial como se explicó en la Sección 2.3.1, esto con la suposición de que en los primeros instantes de tiempo de la ejecución del programa, únicamente existe el ruido de fondo, y posteriormente el umbral se actualiza en los fragmentos donde no hay presencia de voz.

Se implementó un detector de actividad de voz en el dominio de la frecuencia como el que se explicó en la Sección 2.3.3, el cual, actualiza el valor del umbral en los fragmentos inavtivos de voz realizando un seguimiento de la energía del ruido, como se describe en la Ecuación 2.70. Además, el valor utilizado de la constante $p_{VAD}$ es de 0.25, dicho valor se determinó con base en la mejor respuesta obtenida de las señales utilizadas en los experimentos, dando prioridad al valor del umbral anterior para la actualización del mismo.

El desarrollo de la implementación del sistema se basó en la teoría de MUSIC y del formador de haz fijo, es decir, se aprovecha de las características de la matríz de covarianza para poder estimar la dirección de arribo de las fuentes de voz.

Las señales adquiridas por los micrófonos se trabajaron en el dominio de la frecuencia vía la transformada discreta de Fourier, con el objetivo de analizar las frecuencias donde existe la mayor concentración de la energía de la señal. El análisis se realizó con 20 frecuencias diferentes a lo largo del ancho de banda de las señales de voz.

Posteriormente, se realizó la eigendescomposición de la matriz de covarianza para cada una de las frecuencias de análisis, con el objetivo de seleccionar el eigenvector $\mathbf{b_{max}}$ correspondiente al eigenvalor de mayor magnitud $\lambda_{max}$, mismo que representa a la información de las señales en el eigendominio.

La teoría de MUSIC (Sección 3.2.1) supone que las señales no tienen correlación entre sí y con el ruido, por lo que al realizar la eigendescomposición de la matriz de covarianza, cada eigenvalor y eigenvector corresponde a una señal en el subespacio de las señales (eigendominio), predominando la magnitud de los eigenvalores que son relacionados a las señales. Sin embargo, en los experimentos realizados de este trabajo, predominó únicamente la magnitud de un eigenvalor en cada una de las frecuencias, incluso al utilizar múltiples fuentes de voz. Este fenómeno se le atribuye, en gran medida, a la correlación que existe entre las señales, el contenido espectral en común y a la distribución de la energía de las señales emitidas por la fuente en el recinto acústico. Por lo que, dicho eigenvalor y su respectivo eigenvector se relacionó con todas las posibles fuentes de interés en el subespacio de las señales.

Considerando lo anterior, se implementó un formador de haz fijo como el de la Figura 3.4(b) para cada una de las frecuencias analizadas, pero con la diferencia de que las entradas del mismo son los elementos del eigenvector que representa a las señales en el eigendominio.

La respuesta de dirección se obtuvo calculando la energía a la salida de la estructura del formador de haz fijo construido con los elementos del eigenvector relacionado a las señales de interés, en conjunto con el vector de direcciones, dicho vector de direcciones es calculado fuera de línea. De manera similar que en la Ecuación (3.33), la distribución de la energía del eigenvector en cada ángulo de dirección $\theta$ se calcula de la siguiente manera:

$$\mathbf{P_b}(\omega,\theta) = \sum_{m=0}^{M-1} |a_m^*(\omega,\theta)\, b_m(\omega)|^2 = \mathbf{a}^H(\omega,\theta)\, \mathbf{R_{bb}}(\omega)\mathbf{a}(\omega,\theta), \tag{4.1}$$

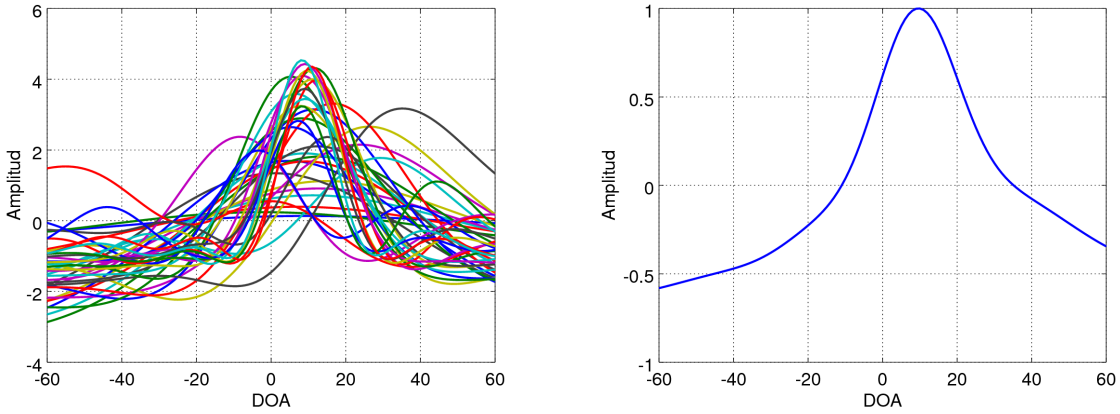
donde $\mathbf{R_{bb}}(\omega)$ es la matriz de autovarianza del eigenvector $\mathbf{b}(\omega)$ relacionado a las señales en la frecuencia $\omega$, dicha matriz tiene una estructura parecida a una matriz Toeplitz y se calcula como se muestra a continuación:

$$\mathbf{R_{bb}}(\omega) = E\left[\mathbf{b}(\omega)\mathbf{b}^H(\omega)\right]. \tag{4.2}$$

Finalmente, se obtiene un patrón de respuesta de dirección para cada una de las frecuencias, donde predominan los lóbulos principales en la dirección de arribo de una fuente, sin embargo, no todos corresponden a la misma posición, es decir, en algunas frecuencias la posición de los lóbulos principales no corresponden a la dirección de arribo real de la fuente, dichos lóbulos son generalmente estimaciones erróneas causadas por ruido. En la Figura 4.5 se muestra un ejemplo de la respuesta de dirección, dicha prueba se realizó con un fragmento de 1024 muestras de la fuente de voz. La posición real de la fuente de voz es en 10 grados y el análisis se realizó con veinte frecuencias dentro del ancho de banda de la voz. Cada una de las curvas de la Figura 4.5(a) representa la respuesta de dirección en una frecuencia diferente, dicha respuesta fue restada con su respectivo promedio.

Se calculó el promedio de las respuestas de dirección en función del ángulo $\theta$, con el objetivo de mantener la amplitud de los lóbulos principales en las direcciones de arribo de la fuente,

mientras que reduce el tamaño de los lóbulos de falsas detecciones, obteniendo así una única respuesta de dirección, en la cual predomina la magnitud de los lóbulos cuyas direcciones de arribo existen en la mayoría de las frecuencias. En la Figura 4.5(b) se muestra el promedio de las respuestas de dirección.




(a) Respuesta de dirección en cada frecuencia.

(b) Promedio de la respuesta de dirección.

Figura 4.5: Patrón de respuesta de dirección de un *frame* con 20 frecuencias diferentes.

El paso final radica en encontrar la posición cuyo valor es el máximo en la respuesta de dirección, sin embargo, ésto se puede realizar cuando únicamente existe una fuente en el recinto. Para el caso de múltiples fuentes, el patrón de respuesta de dirección contendrá tantos lóbulos principales como el número de fuentes activas en el recinto y dichos lóbulos tendrán amplitudes diferentes, por lo que la dirección de arribo va a depender de encontrar los picos máximos de los lóbulos presentes en la respuesta de dirección.

La detección de picos se realizó por medio del cambio de signo en las pendientes de los lóbulos, de tal manera que se detecta un pico cuando el signo de la diferencia entre un número de muestras cambia de positivo a negativo o en su defecto es cero.

Por último, en la Figura 4.6 se exhibe un diagrama de bloques que representa la implementación del método que se ejecutó para la estimación de la dirección de arribo de fuentes de voz en el presente trabajo. El diagrama considera que la actividad de voz es constante.

La matriz $\mathbf{R}(\omega)$, es la matriz de covarianza promedio en la frecuencia $\omega$, formada por las señales de entrada en el dominio de la frecuencia $\mathbf{X}(\omega)$, como se muestra a continuación

$$\mathbf{R}(\omega) = \frac{1}{G}\sum_{g=1}^{G} E\left[\mathbf{X}_g(\omega)\mathbf{X}_g^H(\omega)\right] \tag{4.3}$$

donde $G$ es el número total de ventanas promediadas en el tiempo. En el presente trabajo se realizó el promedio de las matrices de covarianza en función de la frecuencia $\omega$ con 20 ventanas de tamaño 1024. El vector $\mathbf{X}(\omega)$ contiene la información de las señales de los $M$ micrófonos en

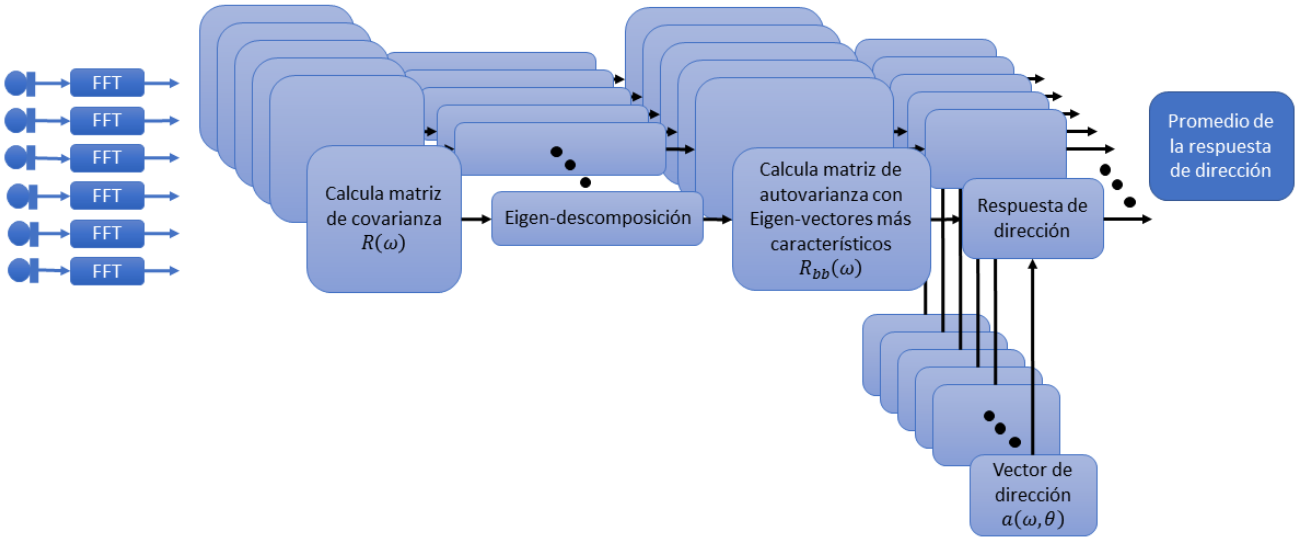



Figura 4.6: Diagrama de bloques del método de dirección de arribo implementado.

la frecuencia $\omega$.

Tomando en cuenta que la dirección de arribo de la fuente de voz puede cambiar en función del tiempo y de la misma manera, aparecer o desaparecer una fuente de voz en el tiempo t, las matrices de covarianza promedio de la Ecuación (4.3) se reinicializan en cero cuando el número de ventanas consecutivas con actividad de voz inactiva excede a ocho. En el Apéndice B se muestra un diagrama de flujo que explica de manera general la implementación del método propuesto.

## 4.3. Clasificador

Las direcciones de arribo de las señales provenientes de las fuentes de voz determinadas por los máximos valores del patrón de respuesta de dirección las vamos a definir como observaciones $O^t$ en el tiempo $t$. Las observaciones $\mathbf{O^t} = \begin{bmatrix} O_0^t & O_1^t & \cdots & O_{Q-1}^t \end{bmatrix}$ corresponden a las $Q$ fuentes potenciales estimadas por el método implementado en la Sección 4.2.

Para cada fuente potencial puede existir cualquiera de las tres posibilidades siguientes [44]:

1. La fuente potencial estimada $q$ en el tiempo $t$ corresponde a una fuente de voz seguida $j$ en el tiempo $t-1$ ($h_0$).

2. La fuente potencial corresponde a una nueva fuente de voz ($h_1$).

3. Corresponde a una falsa detección (puede ser ruido o reverberación) ($h_2$).

En la Figura 4.7 se muestra la asignación de cuatro fuentes potenciales en el tiempo $t$ con base en las fuentes seguidas en el tiempo $t-1$. Se puede observar que una de las fuentes potenciales es una falsa detección. Las fuentes potenciales $q_1$ y $q_3$ corresponden a las fuentes seguidas

$j_1$ y $j_2$ y finalmente una fuente potencial representa a una fuente nueva que requiere un seguimiento. De la misma manera se puede observar que hay una fuente seguida a la cual no se le clasificó ninguna fuente potencial, por lo que después de un intervalo de tiempo se puede determinar como fuente inactiva.

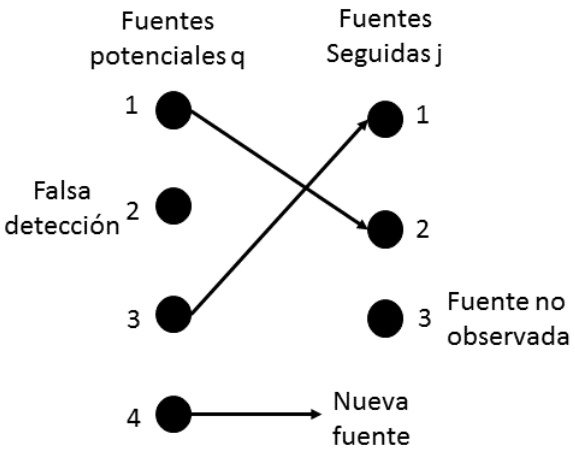



Figura 4.7: Asignación de las fuentes potenciales a las fuentes seguidas.

Para definir el comportamiento de los resultados dados en la salida del formador de haz, vamos a utilizar la distribución Gaussiana o distribución normal. El modelo matemático de la distribución Gaussiana fue definida a principios del siglo XIX por *Carl Friedrich Gauss* y *Adrien-Marie Legendre*. La distribución Gaussiana es un modelo que proporciona una manera fácil de estimar la incertidumbre en muchos fenómenos que acontecen en el mundo. Si se considera una variable $x$, la distribución Gaussiana puede ser descrita como se muestra en la siguiente ecuación:

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp\left\{-\frac{1}{2\pi\sigma^2}\left(x-\mu\right)^2\right\}, \tag{4.4}$$

donde $\mu$ representa la media de la distribución y $\sigma^2$ la varianza, la raíz cuadrada de la varianza $\sigma$ es la desviación estándar. La distribución Gaussiana tiene algunas propiedades matemáticas que son importantes y se tomaron en cuenta para seleccionarla como modelo matemático en el clasificador. La primera es que el producto de varias distribuciones Gaussianas forman otra distribución Gaussiana, de tal manera que no es necesario buscar la forma de la distribución resultante. La otra propiedad, que es muy importante, es la definición del teorema del límite central. El teorema del límite central indica que la distribución de probabilidad de la suma de un conjunto de variables aleatorias independientes e idénticamente distribuidas, se aproxima a una distribución Gaussiana a medida que el número de variables aumenta. Finalmente, la distribución Gaussiana satisface los siguientes requerimientos:

$$\mathcal{N}\left(x|\mu,\sigma^2\right) > 0 \tag{4.5}$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = 1 \tag{4.6}$$

En la Figura 4.8 se muestra una gráfica de la distribución Gaussiana con media cero y varianza unitaria, en donde se puede observar que es simétrica con respecto al valor de la media $\mu$, la media representa la mayor concentración de elementos en la distribución. También se puede observar que el valor de la distribución $\mathcal{N}\left(x|\mu,\sigma^2\right)$ disminuye a medida que $x$ se aleja de la media.

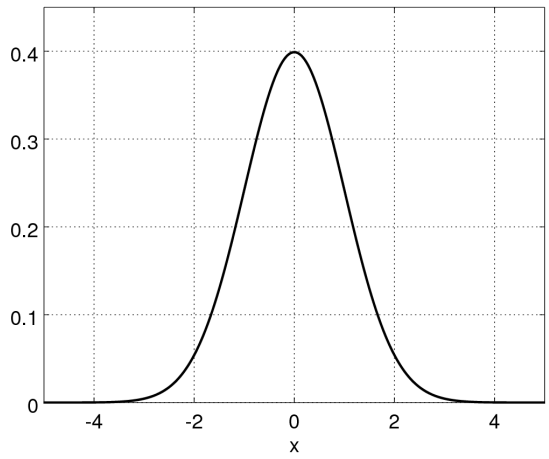



Figura 4.8: Distribución Gaussiana con media cero y varianza unitaria.

En los experimentos realizados, se observó que la distribución de la dirección de arribo para una fuente de voz estática se aproxima a una Gaussiana con media $\mu = \theta_{fuente}$, además el tipo de distribución de la dirección de arribo para una fuente de voz en movimiento depende de la dirección, la velocidad y aceleración de la fuente, por ejemplo, si el movimiento de la fuente es lineal con velocidad constante, la distribución del movimiento será de tipo uniforme, sin embargo, la fuente de voz (en condiciones reales) puede realizar cualquier movimiento dificultando el modelado de la distribución. No obstante, en el presente trabajo se considera que la distribución de la dirección de arribo de una fuente en movimiento en el tiempo $t$ tiene un comportamiento Gaussiano con media $\mu = \theta_{DOA_t}$, de tal manera que en el tiempo $t+1$ la distribución Gaussiana de la dirección de arribo se desplazó con media $\mu = \theta_{DOA_{t+1}}$.

Con base en lo anterior, la clasificación de una observación $q$, con alguna fuente seguida $j$, se realizó efectuando la máxima verosimilitud de las observaciones con cada una de las fuentes seguidas. La verosimilitud es la probabilidad de una observación dados los parámetros de un modelo. Vista como una función de $\mu$ y $\sigma^2$ se define como:

$$p\left(O_q^t|\mu,\sigma^2\right) \tag{4.7}$$

donde $O_q^t$ es la q-ésima observación en el tiempo $t$.

Para el caso de la hipótesis $h_0$, la asignación de la fuente potencial a una fuente seguida se realizó calculando la verosimilitud de la observación con respecto al modelo de cada una de las

fuentes seguidas, de tal manera que la observación en el tiempo $t$ puede ser candidato a pertenecer a la fuente $j$.

La función de densidad de probabilidad que modela el comportamiento de cada una de las fuentes es una distribución Gaussiana como la que se describe en la Ecuación (4.4) con media $\mu = \theta$ (dirección de arribo de la fuente). La desviación estándar $\sigma$ se definió con base en el intervalo angular que establece la mínima separación de dos fuentes de voz para que sean detectadas por el sistema, es decir, $\sigma = 6.2048$ para $\theta_\sigma = \pm 15$ grados.

Finalmente, es necesario verificar que la observación candidata $O_q^t$ realmente pueda ser clasificada con la fuente seguida que obtuvo la máxima verosimilitud, de tal manera que se define el umbral $H$. Si la máxima verosimilitud $p\left(O_q^t | \mu = \theta_j, \sigma^2\right)$ es mayor al umbral $H$, la fuente potencial $q$ pertenece a la fuente seguida $j$. En la Figura 4.9 se muestra un ejemplo en donde se tienen definidos cuatro modelos Gaussianos que corresponden a cuatro fuentes seguidas $j_1$, $j_2$, $j_3$ y $j_4$ con medias $\mu_1 = -45$, $\mu_2 = -17$, $\mu_3 = 0$, y $\mu_4 = 30$ respectivamente. Si en el tiempo $t$ se obtiene la dirección de arribo de una fuente potencia en $O_q^t = 4$ grados, se observa claramente que la máxima verosimilitud corresponderá a la fuente seguida con media $\mu_3 = 0$ grados (punto negro de la Figura 4.9), además supera el valor del umbral $H$ por lo que se puede clasificar con el modelo Gaussiano $j_3$.

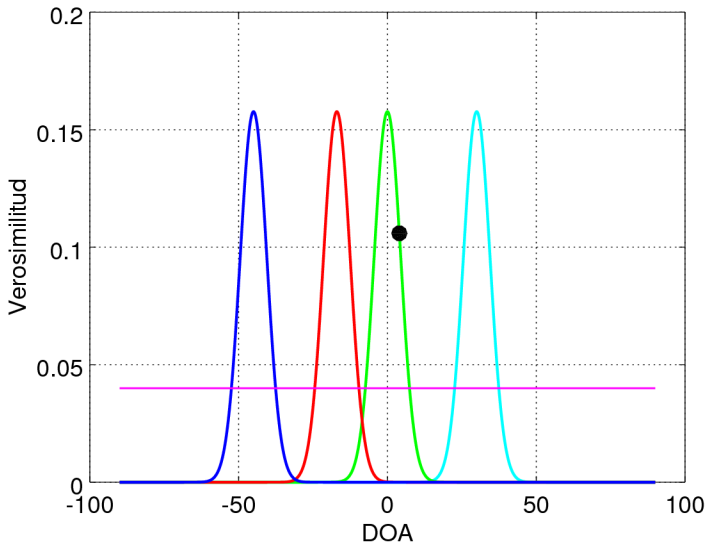



Figura 4.9: Modelo de cuatro fuentes seguidas y un umbral de decisión en el tiempo $t$.

Es necesario tomar en cuenta que una fuente potencial $q$ solamente puede pertenecer a una fuente seguida $j$, es decir, cuando la máxima verosimilitud de dos observaciones $O_1^t$ y $O_2^t$ correspondan a una misma clase o fuente seguida, la observación que tenga la mayor probabilidad de pertenecer al grupo es la que se clasifica en el mismo, mientras que la otra observación puede ser una falsa detección o corresponde a una fuente nueva.

Al inicio de la ejecución del programa ($t = 0$), el clasificador considera que existen tantas clases o fuentes seguidas como el número de observaciones $O_q^t$ (estimaciones de DOA), ésto se debe porque el sistema no tiene información a priori sobre el número de fuentes existentes.

### 4.3.1. Fuentes activas y fuentes inactivas

En un experimento con condiciones reales, el sistema estima la dirección arribo de las fuentes de voz presentes en el tiempo $t$, sin embargo, para tiempos posteriores el número de fuentes de voz puede incrementar o disminuir. De tal manera que las fuentes de voz se han clasificado en dos tipos: fuentes activas y fuentes inactivas. Una fuente se considera como activa en un intervalo de tiempo cuando la fuente de voz tiene una emisión continua de señales durante dicho intervalo, no obstante, la fuente de voz puede dejar de emitir señales, por lo que el sistema no podrá estimar su dirección de arribo y dicha fuente de voz se clasificará como fuente inactiva.

En la Figura 4.10 se muestra una gráfica de la estimación de dirección de arribo de dos fuentes de voz en función del tiempo, donde el eje de las ordenadas representa el ángulo de posición de la fuente y el eje de las abscisas el tiempo en segundos. Se puede observar que la dirección de arribo de la primera fuente detectada es en $t = 5$ segundos después de haber iniciado el sistema, la fuente permanece activa en el intervalo de tiempo de 5 a 50 segundos. En el tiempo $t = 35$ segundos aparece una nueva fuente en $\theta = 80$ grados, y permanece activa de $t = 35$ a $t = 90$ segundos; sin embargo, la fuente en la posición $\theta = 240$ grados tiene dos intervalos de silencio, por lo que en dichos intervalos de silencio la fuente se considera inactiva.

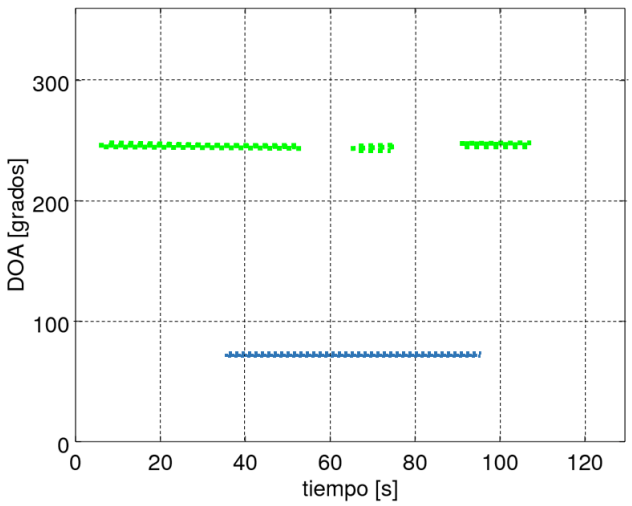



Figura 4.10: Ejemplo de estimación de dirección de dos fuentes en función del tiempo.

En el presente trabajo, una fuente se consideró como inactiva después de que las direcciones de arribo estimadas no pertenecían a dicha fuente después de tres segundos. Además, el calculo de la verosimilitud de una observación con respecto a las fuentes seguidas, se realiza dando prioridad a las fuentes activas, de tal manera que si la observación no pertenece a ninguna de las fuentes activas se procede a verificar con las inactivas, por lo que la creación de una nueva

clase o fuente se realiza cuando la observación no pertenece a ninguna de las clases activas e inactivas.

El clasificador implementado en el presente trabajo no toma en cuenta las características de la fuente de voz si no que clasifica los resultados de la estimación de dirección, es decir, clasifica con base a la posición de las fuentes. En el Apéndice C se muestra un diagrama de flujo que describe el funcionamiento y la implementación del clasificador en el presente trabajo.

## 4.4. Filtro de Kalman

El objetivo del filtro de Kalman es estimar el estado $x \in \Re^n$ de un proceso controlado en tiempo discreto que se rige por la ecuación en diferencias estocástica lineal (4.8)

$$\mathbf{x_t} = \mathbf{A_t x_{t-1}} + \mathbf{B}u_t + \mathbf{w_t}\,, \tag{4.8}$$

con una medición $z \in \Re^n$, definida como:

$$\mathbf{z_t} = \mathbf{H_t x_t} + \mathbf{v_t}\,, \tag{4.9}$$

donde la matriz **A** de tamaño $n\,x\,n$, conocida como matriz de transición de estados, relaciona el estado en el momento previo $\mathbf{x_{t-1}}$ con el estado en el tiempo actual $\mathbf{x_t}$, y la matriz **B** relaciona las entradas de control $\mathbf{u} \in \Re^n$ con el estado estimado en el tiempo actual $\mathbf{x_t}$. La matriz de transformación **H** mapea los parámetros del vector de estado $\mathbf{x_t}$ al dominio de las mediciones $\mathbf{z_t}$.

Las variables aleatorias $\mathbf{w_k}$ y $\mathbf{v_k}$ representan los ruidos del proceso y de la medición respectivamente. Se asume que ambos son ruido blanco, independientes entre si y con densidad de probabilidad normal, ésto es:

$$p\,(w) \sim \mathcal{N}\,(0, Q) \tag{4.10}$$

$$p\,(v) \sim \mathcal{N}\,(0, R) \tag{4.11}$$

donde $Q$ y $R$ son las varianzas del ruido del proceso y de la medición respectivamente.

El filtro de Kalman permite estimar el estado $\tilde{\mathbf{x}}_t$ cuando el estado $\mathbf{x_t}$ no puede ser observado directamente, la estimación se realiza combinando los modelos del sistema con las mediciones ruidosas de ciertos parámetros o funciones lineales de los mismos, por lo tanto, las estimaciones de interés del vector de estado estarán definidas por funciones de densidad de probabilidad.

### 4.4.1. Etapas del filtro de Kalman

El filtro de Kalman consta de dos etapas para su implementación, primero realiza la estimación del estado y posteriormente ajusta la estimación con base a la retroalimentación dada por las mediciones ruidosas. Las ecuaciones de actualización se centran en proyectar hacia delante (en tiempo) el estado actual y el error de covarianza para obtener una estimación a priori. Las ecuaciones de actualización de la medición se enfocan en la retroalimentación, es decir,

se introducen las mediciones a la estimación a priori para obtener una estimación a posteriori mejorada [45].

Las ecuaciones de actualización de tiempo se pueden ver como ecuaciones de predicción, mientras que las ecuaciones de actualización de las mediciones como ecuaciones de corrección. En la Figura 4.11 se muestra un diagrama que explica el flujo de las etapas en el algoritmo del filtro de Kalman implementado.

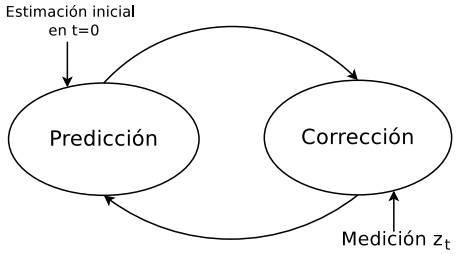



Figura 4.11: Etapas del algoritmo del filtro de Kalman

Las ecuaciones destinadas para la etapa de predicción calculan el estado apriori $\tilde{\mathbf{x}}_t^-$ y la matriz de covarianza **P**. La predicción del estado apriori se calcula por medio de la siguiente expresion:

$$\tilde{\mathbf{x}}_\mathbf{t}^- = \mathbf{A}\,\mathbf{x}_{\mathbf{t-1}} + \mathbf{B}\,\mathbf{u}_\mathbf{t} \tag{4.12}$$

Las varianzas y covarianzas se almacenan en la matriz de covarianza **P**, es decir, los elementos de la diagonal principal de la matriz **P** representan las varianzas asociadas a los parámetros del vector de estado, mientras que los elementos localizados fuera de la diagonal principal representan las covarianzas entre los elementos del vector de estado. La matriz **P** está definida como:

$$\mathbf{P}_\mathbf{t} = E\left[(\mathbf{x}_\mathbf{t} - \tilde{\mathbf{x}}_\mathbf{t})\,(\mathbf{x}_\mathbf{t} - \tilde{\mathbf{x}}_\mathbf{t})^T\right] \tag{4.13}$$

De tal manera que la matriz de covarianza apriori $\tilde{\mathbf{P}}_\mathbf{t}^-$ se calcula como se muestra en la siguiente expresión:

$$\mathbf{P}_\mathbf{t}^- = \mathbf{A}\,\mathbf{P}_{\mathbf{t-1}}\,\mathbf{A}^T + \mathbf{Q}_{\mathbf{t-1}} \tag{4.14}$$

donde la matriz $\mathbf{Q}_{\mathbf{t-1}}$ es la matriz de covarianza del ruido del proceso, asociada con las entradas ruidosas de control.

La ecuación de corrección del estado, como se muestra en (4.15), tiene el objetivo de calcular la estimación del estado a posteriori $\tilde{\mathbf{x}}_\mathbf{t}$ como una combinación lineal de una estimación a priori $\tilde{\mathbf{x}}_\mathbf{t}^-$ y una diferencia ponderada entre la medición actual $z_t$ y la predicción de medición $\mathbf{H}\,\tilde{\mathbf{x}}_\mathbf{t}^-$, conocida como medida residual. Tal medida residual refleja la discrepancia entre la medición actual y la estimada.

$$\tilde{\mathbf{x}}_\mathbf{t} = \tilde{\mathbf{x}}_\mathbf{t}^- + \mathbf{K}_\mathbf{t}\,(z_t - \mathbf{H}\,\tilde{\mathbf{x}}_\mathbf{t}^-) \tag{4.15}$$

La matriz $\mathbf{K_t}$ de tamaño $n\,x\,m$ es conocida como la ganancia de Kalman y su función es minimizar el error de covariaza, dicha ganancia de Kalman se calcula por medio de la Ecuación (4.16).

$$\mathbf{K_t} = \frac{\mathbf{P_t^-}\mathbf{H_t}^T}{\mathbf{H}\,\mathbf{P_t^-}\mathbf{H}^T + \mathbf{R_t}} \tag{4.16}$$

Podemos observar que si la covarianza del error de medición $\mathbf{R_t}$ es muy pequeña, la ganancia de Kalman será grande, de tal manera que la estimación dará mayor peso a la medida residual, esto es:

$$\lim_{R_t \to 0} K_t = H_t^{-1} \tag{4.17}$$

Además, si el valor de $P_t^-$ se aproxima a cero, la estimación del estado $\tilde{\mathbf{x}}_\mathbf{t}$ se va a aproximar a la estimación a priori del estado $\tilde{\mathbf{x}}_\mathbf{t}^-$, es decir:

$$\lim_{\tilde{P}_t^- \to 0} K_t = 0 \tag{4.18}$$

Finalmente, la actualización de la matriz de covarianza del vector de estado a posteriori se calcula por medio de la Ecuación (4.19).

$$\mathbf{P_t} = (\mathbf{I} - \mathbf{K_t}\mathbf{H_t})\,\mathbf{P_t^-} \tag{4.19}$$

donde $\mathbf{I}$ es una matriz identidad de tamaño $n\,x\,n$.

En el presente trabajo, el filtro de Kalman se utiliza para estimar la dirección de arribo real de una fuente de voz con base a los DOAs obtenidos por el algoritmo de estimación de posición (Sección 4.2). Cabe destacar que se aplica un filtro de Kalman para cada una de las fuentes de voz detectadas $j$, de tal manera que las mediciones $z_t$ requeridas en cada filtro, son las observaciones $O_q^t$ clasificadas con cada una de las fuentes seguidas $j$ (como se explica en la Sección 4.3).

El vector de estados está formado por dos elementos, el ángulo de dirección de arribo de la fuente seguida y la velocidad estimada de la fuente seguida en el tiempo $t$, esto es:

$$\tilde{\mathbf{x}}_\mathbf{t} = \begin{bmatrix} \tilde{x}_0 \\ \tilde{x}_1 \end{bmatrix} = \begin{bmatrix} \theta_f \\ v_f \end{bmatrix}, \tag{4.20}$$

además, la variable de control $U_t$ es la velocidad promedio de las mediciones de dirección de arribo calculadas por el algoritmo de dirección de arribo. La predicción del estado se realiza con base al desplazamiento realizado por la fuente, como se muestra en (4.21)

$$\theta_t = \theta_{t-1} + \Delta_\theta\,, \tag{4.21}$$

donde $\Delta_\theta$ es el desplazamiento en grados realizado por la fuente de voz en el intervalo de tiempo comprendido de $t-1$ a $t$. Dicho desplazamiento se puede calcular por la siguiente expresión:

$$\Delta_\theta = \Delta_t\, v_f \tag{4.22}$$

donde $\Delta_t$ es el intervalo de tiempo entre la estimación anterior y la actual. De tal manera que la matriz $\mathbf{A}$ utilizada en cada uno de los filtros se muestra a continuación:

$$\mathbf{A} = \begin{bmatrix} 1 & \Delta_t \\ 0 & 1 \end{bmatrix} \tag{4.23}$$

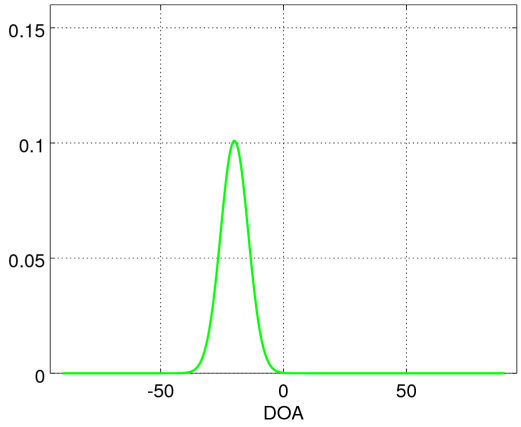
La covarianza del error de medición $R_t$ corresponde al valor de la variación de las muestras, es decir, al valor utilizado para determinar si una muestra pertenece a una fuente seguida en el clasificador (Sección 4.3). Se considera que el ruido del proceso de estimación es blanco, por lo que la matriz de covarianza del ruido utilizada es una matriz identidad como la que se muestra a continuación:

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.24}$$

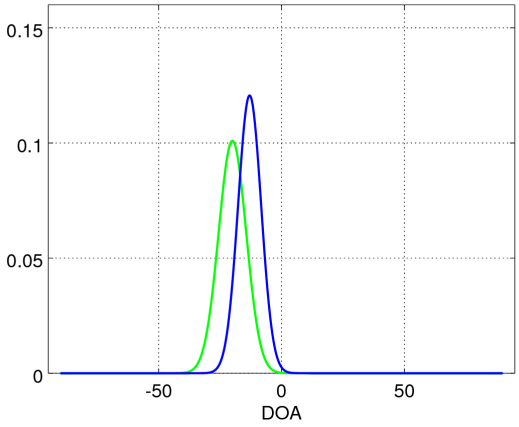
En la Figura 4.12 se muestra un ejemplo de la estimación de la dirección de arribo de una fuente de voz con base en su movimiento. La fuente de voz tiene un movimiento lineal, donde las posiciones reales son $\theta_t = -15$ y $\theta_{t+1} = -2$ grados. La predicción de la posición en el tiempo $t$ se muestra en la Figura 4.12(a), se puede observar que la predicción de la posición de la fuente de voz está centrada en $\theta_t^- = -20$ grados, mientras que el valor de la medición resultó en $O_t = -13$ grados (Figura 4.12(b)), de tal manera que la estimación del estado $\tilde{\theta}_t$ dio como resultado en la posición $\tilde{\theta}_t = -15$ grados. Dicho valor de estimación está en función de la ganancia de Kalman, que da prioridad ya sea a la medida residual o al estado calculado en la etapa de predicción. La Figura 4.12(c) muestra la gráfica de la estimación de la posición de la fuente (curva de color rojo) en el tiempo $t$ con base a la predicción (curva de color verde) y la medición (curva de color azul). Posteriormente, se vuelve a calcular el ciclo de estimación comenzando por la etapa de predicción (Figura 4.12(d)), la medición de las observaciones en el tiempo $t+1$ (Figura 4.12(e)) y finalmente la estimación de la dirección de arribo de la fuente en $\tilde{\theta}_{t+1} = -3$ grados, como se muestra en la Figura 4.12(f).

## 4.5. Resumen

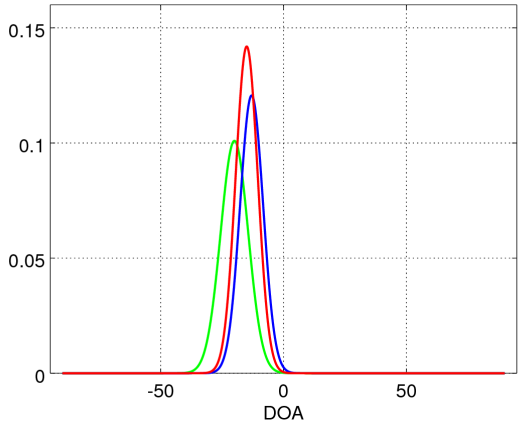
En el presente capítulo se expuso la implementación de un sistema de localización de fuentes de voz que opera en línea, el cual se aprovecha de la correlación entre las señales de las fuentes. Con base en la teoría de MUSIC se realizó la eigendescomposición de la matriz de covarianza creada con las señales capturadas en el dominio de la frecuencia, sin embargo, solamente se utiliza un eigenvector para realizar el análisis de dirección de arribo. El sistema se implementó con dos tipos de arreglos de micrófonos, un arreglo lineal uniforme y un arreglo planar con geometría circular, ambos casos con seis micrófonos y tomando en cuenta el modelo de campo lejano que se explicó en la Sección 2.1.3. Posteriormente, la etapa de seguimiento se implemento con un filtro de Kalman con el objetivo de estimar la posición real de la fuente con base a los resultados ruidosos obtenidos con el algoritmo de dirección. Finalmente, se consideró un clasificador que tiene la función de asignar los resultados del método de dirección de arribo (observaciones) a las fuentes seguidas por el sistema, tal clasificador asigna las observaciones a partir del cálculo de la máxima verosimilitud con modelos Gaussianos. Además tiene la característica de crear una nueva fuente o clase de asignación si la observación dada no corresponde a alguna de las fuentes activas o inactivas existentes.



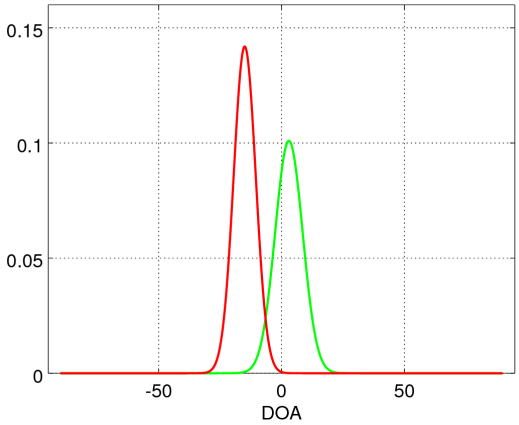

(a) Predicción en $t$



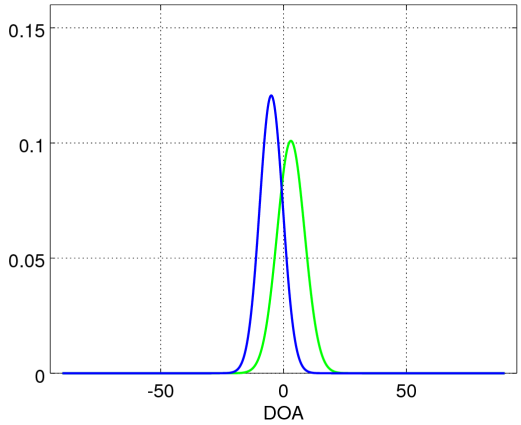

(b) Medición



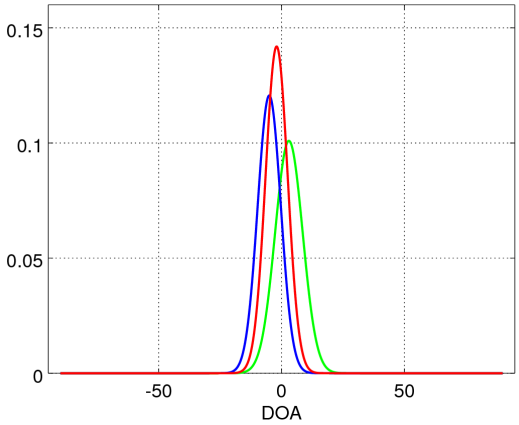

(c) Estimación de la posición de la fuente en $t$




(d) Predicción de nueva posición en $t+1$




(e) Medición de nueva posición




(f) Estimación de nueva posición de la fuente en $t+1$

Figura 4.12: Ejemplo de la estimación de dirección de arribo con el Filtro de Kalman

# Capítulo 5.

# Pruebas y análisis de resultados

Se elaboraron diferentes experimentos para determinar la estimación del ángulo de arribo de señales provenientes de una y múltiples fuentes de voz. Las pruebas se realizaron en el laboratorio de robótica ubicado en el cuarto piso del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS). Dicho laboratorio tiene un ambiente acústico no controlado, es decir, contiene contaminación acústica producida por dispositivos electrónicos posicionados alrededor del laboratorio y en algunos casos ruidos generados por personas que se encontraban trabajando en el laboratorio, además de la reverberación.

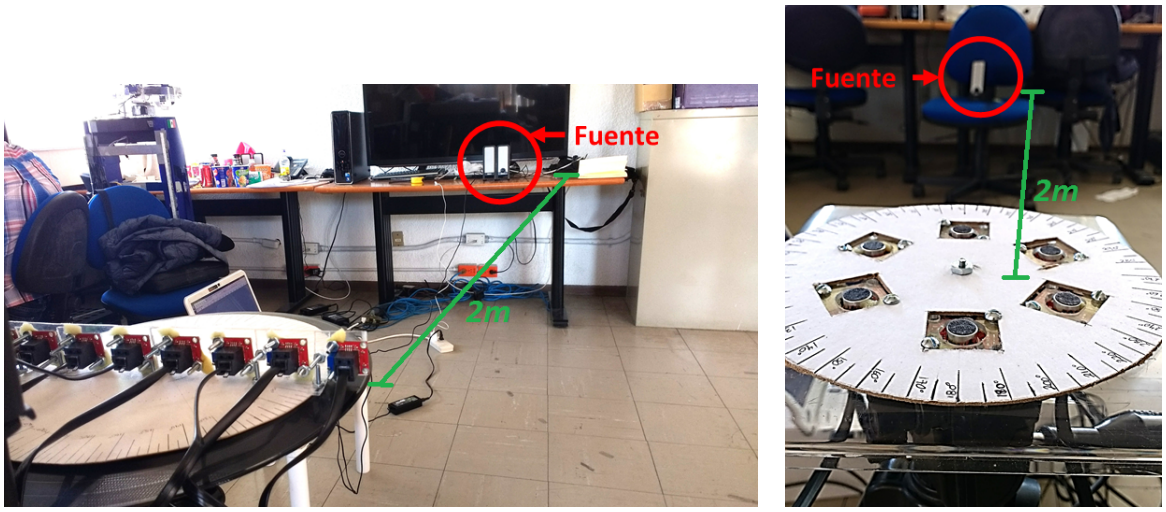
Los experimentos se realizaron reproduciendo audio libros con voces tanto de hombre como de mujer por medio de los altavoces modelo A225 de la marca Dell. Para modelar el comportamiento de una fuente de voz, dichos altavoces se posicionaron en un lugar específico para cada experimento, permaneciendo ya sea de manera estática o en movimiento. Las señales emitidas por las fuentes se adquirieron por medio de los arreglos de micrófonos, ya sea por el arreglo lineal uniforme (Figura 4.3(a)) o el arreglo circular (Figura 4.3(b)). En el cuadro 5.1 se muestran las características de los arreglos de micrófonos utilizados en los experimentos.

Los experimentos se realizaron considerando el modelo de campo lejano descrito en la Sec-

| Característica | Arreglo lineal uniforme | Arreglo circular |
|---|---|---|
| Frecuencia de muestreo $f_s$ | 44100 $Hz$ | 44100 $Hz$ |
| No. de micrófonos utilizados | 6 | 6 |
| Distancia entre micrófonos $d$ | 4 $cm$ | 4.2 $cm$ |
| Radio de la circunferencia $r$ | no aplica | 4.2 $cm$ |
| Ángulo de posición entre micrófonos | 0 grados | 60 grados |

Cuadro 5.1: Características de los arreglos de micrófonos implementados

ción 2.1.3, de tal manera que la distancia definida entre la fuente de voz y el centro del arreglo de micrófonos es de $r = 2\ m$. En la Figura 5.1 se muestran dos fotografías que ejemplifican un experimento realizado con el arreglo lineal de micrófonos, donde la fuente se localiza en la dirección de arribo $\theta = 0$ grados (Figura 5.1(a)), y un experimento realizado con un arreglo circular de micrófonos (Figura 5.1(b)), donde la fuente, de igual manera, se localiza en $\theta = 0$ grados.



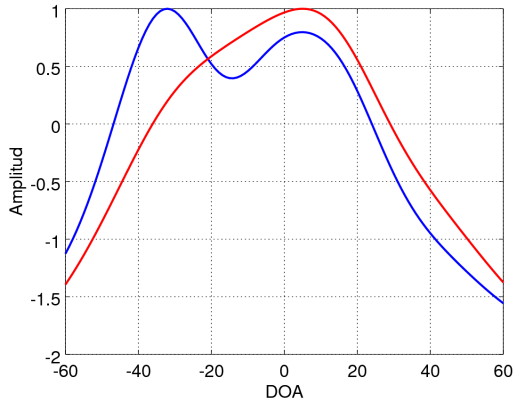

(a) Experimento realizado con una fuente localizada en $\theta = 0$ grados con un arreglo lineal de seis micrófonos.

(b) Experimento realizado con una fuente localizada en $\theta = 0$ grados y un arreglo circular de seis micrófonos.

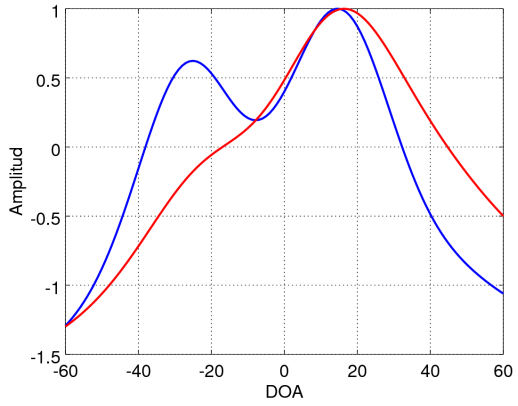
Figura 5.1: Ejemplo de experimentos realizados.

La implementación de una estructura tipo formador de haz con los eigenvectores de la matriz de covarianza ha mostrado mejores resultados en comparación con el formador de haz fijo convencional, eliminando parte del ruido acústico no deseado, así como parte de la reverberación, de tal manera que los lóbulos principales en la respuesta de dirección están mejor definidos y es posible tener una mejor aproximación de la dirección de arribo de la fuente de voz real, sobre todo cuando existe el caso de múltiples fuentes simultáneas. En la Figura 5.2 se muestran dos gráficas, donde cada una representa el patrón de respuesta de dirección de dos fuentes de voz. El experimento de la Figura 5.2(a) se realizó con dos fuentes de voz simultáneas con dirección de arribo $\theta_1 = -35$ y $\theta_2 = 17$ grados, dicha respuesta de dirección se construyó con un *frame* de 1024 muestras de las señales de voz adquiridas por un arreglo lineal de seis micrófonos. La

curva de color azul representa la respuesta de dirección promedio y normalizada, construida con el método implementado, mientras que la curva de color rojo representa la respuesta de dirección promedio y normalizada, la cual, se calculó con un formador de haz fijo. La Figura 5.2(b) muestra un experimento realizado con las mismas características que el anterior, sin embargo, las direcciones de arribo de las señales voz desde las fuentes son $\theta_1 = -25$ y $\theta_2 = 10$ grados.




(a) Respuestas de dirección para dos fuentes de voz simultáneas con DOAs de las fuentes en $\theta_1 = -35$ y $\theta_2 = 17$.




(b) Respuestas de dirección para dos fuentes de voz simultáneas con DOAs de las fuentes en $\theta_1 = -25$ y $\theta_2 = 10$.

Figura 5.2: Comparación de las respuestas de dirección promedio entre el método implementado y el formador de haz fijo.

Se puede observar en ambos casos de la Figura 5.2 que la respuesta de dirección con el método propuesto tiene mejor definición de los lóbulos principales, los cuales corresponden a las direcciones de arribo de las respectivas fuentes de voz, mientras que la respuesta de dirección generada con el formador de haz convencional muestra un lóbulo principal correspondiente a la dirección de arribo de una de las dos fuentes de voz del experimento.

El presente capítulo se divide principalmente en dos secciones, en la Sección 5.1 se presentan los resultados obtenidos con el arreglo lineal uniforme de micrófonos y en la Sección 5.2 se presentan los resultados por medio del arreglo circular de micrófonos. En ambas secciones se muestran experimentos con fuentes de voz estáticas en diferentes posiciones, fuentes de voz estáticas simultáneas y una fuente de voz con desplazamiento.

## 5.1. Resultados con el arreglo lineal uniforme de micrófonos

Los experimentos realizados en el presente trabajo se llevaron a cabo considerando que el arreglo de micrófonos permaneció estático durante el tiempo de la prueba, por lo tanto, no cuenta con un dispositivo para autodesplazarse como en el caso de un robot de servicio. Sin embargo, la fuente de voz puede desplazarse en cualquier dirección y ésta será detectada por el sistema correctamente siempre y cuando la fuente se localice dentro del intervalo de detección.

Para un arreglo lineal de micrófonos, el intervalo de detección de dirección de arribo está comprendido de $\theta = -90$ a $\theta = 90$ grados.

Los experimentos realizados se clasificaron en tres tipos, los cuales son:

1. Estimación de la dirección de arribo de una señal desde una fuente de voz estática.
2. Estimación de la dirección de arribo de múltiples señales desde fuentes de voz no simultáneas.
3. Estimación de la dirección de arribo de múltiples señales desde fuentes de voz simultáneas (máximo tres fuentes).
4. Estimación de la dirección de arribo de una señale desde una fuente de voz en movimiento.

En las siguientes secciones se muestran los resultados obtenidos en los experimentos realizados utilizando el arreglo lineal uniforme de micrófonos. En la Sección 5.1.1 se muestran los resultados de pruebas realizadas con una fuente de voz estática, en la Sección 5.1.2 se presenta un experimento de cuatro fuentes de voz estáticas que son no simultáneas para observar la respuesta del sistema. Los resultados obtenidos con dos y tres fuentes simultáneas se pueden observar en la Sección 5.1.3 y finalmente los experimentos con una fuente de voz en movimiento se presentan en la Sección 5.1.4.

### 5.1.1. Resultados con una fuentes de voz estática

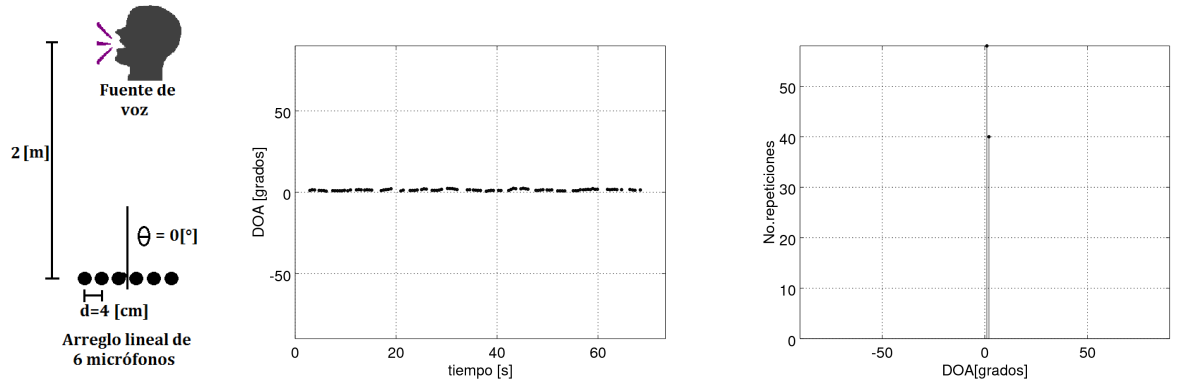
El primer experimento realizado consistió en posicionar una fuente de voz con ángulo de dirección de arribo en $\theta = 0$ grados, como se muestra en la Figura 5.3(a). La adquisición de las señales tuvo una duración de 75 segundos. En la Figura 5.3(b) se muestra la gráfica del seguimiento de la fuente de voz, donde el eje de las ordenadas representa el ángulo de la dirección de arribo en grados y el eje de las abscisas representa el tiempo $t$ en segundos.

En la Figura 5.3(c) se muestra el histograma del ángulo de dirección de arribo de la señal voz, donde se puede apreciar que tiene una variación muy pequeña. La variación resultante de dicho experimento es de un grado.

De la misma manera, para una fuente de voz estática con dirección de arribo en $\theta = 35$ grados, la gráfica de seguimiento de la fuente se muestra en la Figura 5.4(b), mientras que el histograma se puede observar en la Figura 5.4(c). Se puede observar que en ciertos instantes de tiempo el sistema detecta la dirección de arribo de una fuente en $\theta = -20$ grados. Dicha estimación de dirección de arribo se le atribuye a una posible fuente de ruido no deseado, la cual, en esos instantes de tiempo $t$ su energía es suficientemente grande para superar el umbral $H$.

### 5.1.2. Resultado con múltiples fuentes de voz no simultáneas

Se realizó un experimento donde se tienen cuatro fuentes diferentes de voz localizadas en distintas posiciones. En dicho experimento, cada una de las fuentes de voz se activan después

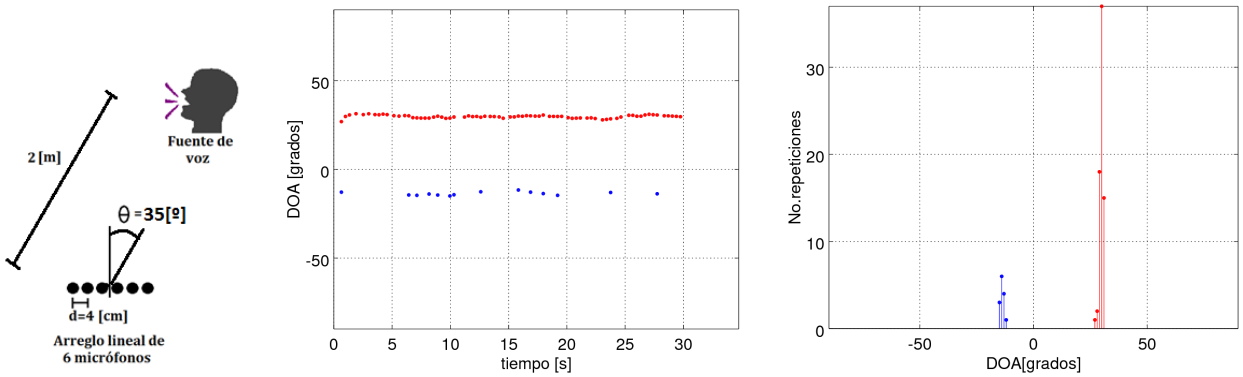



(a) Diagrama espacial.

(b) Seguimiento de la fuente de voz.

(c) Histograma de la dirección de arribo de la fuente de voz.

Figura 5.3: Estimación de la dirección de arribo de una señal proveniente de una fuente de voz estática con ángulo de dirección de arribo en $\theta = 0$ grados, utilizando un arreglo lineal de seis micrófonos.




(a) diagrama espacial.

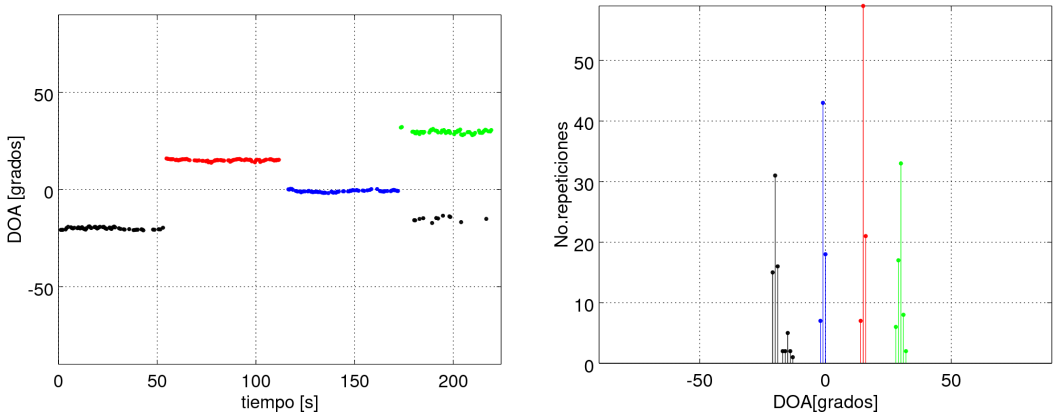
(b) Seguimiento de la fuente de voz.

(c) Histograma de la fuente de voz.

Figura 5.4: Estimación de dirección de arribo de una señal proveniente de una fuente de voz estática con ángulo de dirección de arribo en $\theta = 35$ grados.

de que termina la anterior, ésto con el objetivo de determinar el tiempo de respuesta del sistema. El experimento inicia con una fuente que tiene una dirección de arribo en $\theta_1 = -23$ grados, después de 52 segundos inicia la emisión de la segunda fuente de voz con dirección de arribo en $\theta_2 = 20$ grados, mientras que la primera fuente de voz se vuelve inactiva, como se puede observar en la gráfica de seguimiento de la Figura 5.5(a). Posteriormente se repite el procedimiento activándose otra fuente de voz en la posición $\theta_3 = 0$ grados. Finalmente, el experimento termina con la estimación de una fuente en la dirección de arribo $\theta_4 = 35$. En la Figura 5.5(b) se muestra el histograma del experimento, en donde se puede observar que la fuente que está más lejana al origen tiene una variación máxima de 4 grados, mientras que la varianza de la fuente con dirección de arribo en el origen es de un grado, ésto se le atribuye a que entre más lejos sea el ángulo de arribo de la fuente de voz con respecto al eje de cero grados, el ancho de los lóbulos principales en dicha dirección aumenta, como se observa en la Ecuación (3.34), además, los lóbulos principales obtenidos en las diferentes frecuencias podrían no coincidir exactamente

en el mismo ángulo (como se mostró en la Figura 4.5(a)). También se puede observar que entre más lejano es el DOA de la dirección cero, el sistema es más susceptible a estimar la dirección de arribo de una fuente de interferencia, esto se debe a que la amplitud de los lóbulos principales para estas direcciones disminuye con respecto a las direcciones cercanas al eje cero, de tal manera que al existir una interferencia alrededor del eje cero, ésta se se notará más.




(a) Seguimiento de las fuentes.

(b) Histograma.

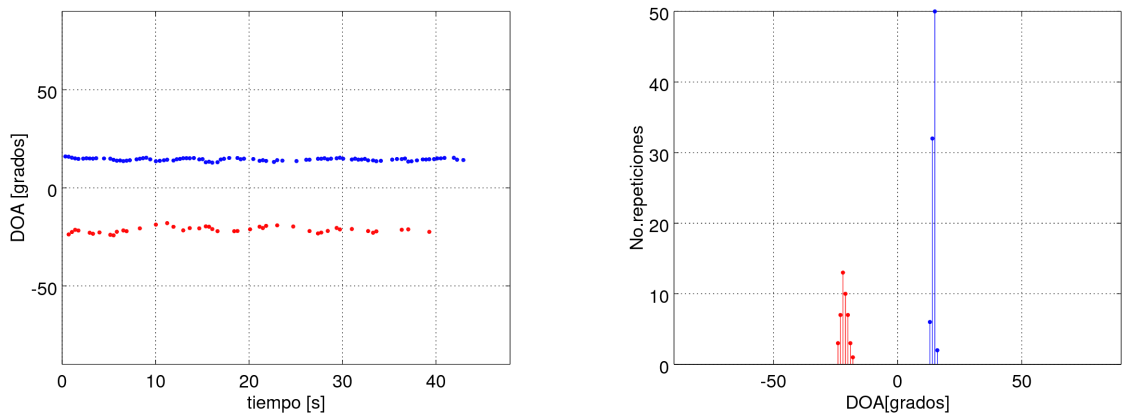
Figura 5.5: Estimación de la dirección de arribo de señales provenientes de diferentes fuentes de voz no simultáneas.

### 5.1.3. Resultados con múltiples fuentes de voz simultáneas

Para el caso de los experimentos realizados con fuentes simultáneas, es importante tomar en cuenta que las señales de voz son variables en el tiempo, de tal manera que en algunos instantes de tiempo $t$, una fuente de voz puede permanecer activa mientras que las demás se encuentran inactivas, o permanecer activas simultáneamente. El sistema es capaz de clasificar correctamente una fuente que se volvió a activar, después de permanecer inactiva durante un tiempo específico, siempre y cuando esta fuente haya permanecido en la misma posición o haya sufrido un desplazamiento mínimo (menor a 5 grados).

En la Figura 5.6 se muestran los resultados de la estimación de dirección de arribo de dos fuentes de voz estáticas. Las direcciones de arribo reales de dichas fuentes de voz son $\theta_1 = -30$ y $\theta_2 = 20$. Se puede observar que la estimación de dirección de arribo de las fuentes tienen una mayor variación comparada con las Figuras 5.3 y 5.4. La fuente con dirección de arribo en $\theta_1 = -30$ grados, tiene una desviación estándar de $\sigma = 1.3717$ que corresponde a una variación de seis grados, mientras que la fuente con dirección de arribo en $\theta_2 = 20$ grados con desviación estándar $\sigma = 0.46491$ que corresponde a una variación de tres grados. Además, la media de las direcciones de arribo estimadas son las siguiente: $\tilde{\theta}_1 = -22$ grados y $\tilde{\theta}_2 = 15$ grados. La discrepancia definida entre las direcciones de arribo reales $\theta$ y las estimadas $\tilde{\theta}$, se le atribuye en gran medida al error de paralelismo generado en la medición física del ángulo de arribo. Esto se debe a que el error obtenido en cada una de las fuentes es muy diferente, es decir, $e_1 = 8$ y $e_2 = 5$

grados respectivamente.




(a) Seguimiento de las fuentes de voz.

(b) Histograma del experimento.

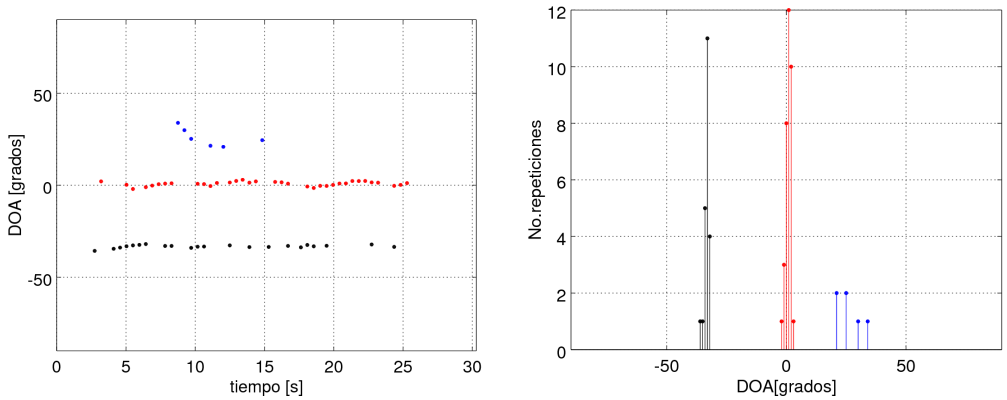
Figura 5.6: Estimación del ángulo de arribo utilizando dos fuentes simultáneas en $\theta_1 = -30$ y $\theta_2 = 20$.

El máximo número de DOAs simultáneas que se estimaron en el presente trabajo fueron tres fuentes, considerando que la separación entre ellas es al menos 25 grados. En la Figura 5.7 se muestran los resultados de la estimación de dirección de arribo de tres fuentes de voz estáticas posicionadas en $\theta_1 = -40$, $\theta_2 = 0$ y $\theta_3 = 40$ grados. A pesar de que la duración de la tercera fuente de voz fue pequeña en comparación con las demás, se puede observar que el sistema presenta dificultades para estimar la dirección de arribo de la misma. Los ángulos de arribo donde se posicionaron las fuentes de voz en este experimento fueron $\theta_1 = -40$, $\theta_2 = 0$ y $\theta_3 = 40$ grados respectivamente, mientras que los DOAs estimados fueron $\tilde{\theta}_1 = -35$, $\tilde{\theta}_2 = 0$ y $\tilde{\theta}_3 = 28$ grados respectivamente, además, la dirección de arribo de la tercer fuente tiene una variación muy grande (13 grados).

De esta última prueba, puede observarse que entre mayor es el número de las fuentes de voz activas simultáneas, como era de esperarse, aumenta la variación de la dirección de arribo de las fuentes y al mismo tiempo el sistema presenta dificultad para estimar la dirección de arribo de todas las fuentes. Sin embargo, el tiempo de respuesta del sistema es de 348 milisegundos cuando la emisión de las señales por las fuentes es constante, de otra manera el tiempo de respuesta está en función de la emisión de las señales.

### 5.1.4. Resultados de una fuente de voz con desplazamiento

Se realizaron tres experimentos diferentes, donde en cada un de ellos las fuentes tienen una trayectoria particular. El primer experimento, con duración de nueve segundos, se llevó a cabo con una fuente de voz que tiene una trayectoria lineal y velocidad aproximada de 4 $\frac{grados}{s}$ en sentido antihorario, comenzando en un ángulo de dirección de arribo positivo hacia uno negativo. La dirección de arribo inicial fue de $\theta_{ini} = 30$ grados, mientras que la dirección inicial estimada es de $\tilde{\theta}_{ini} = 25$ grados, además la dirección de arribo final de la fuente fue de $\theta_{fin} = -25$

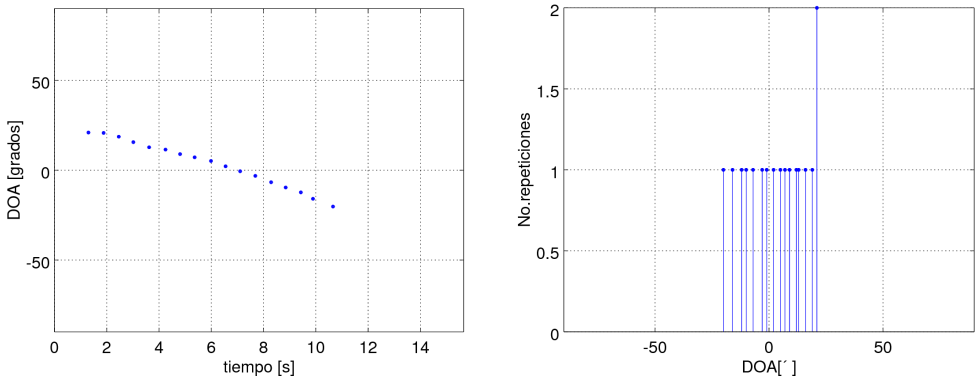



(a) Seguimiento de las fuentes de voz. (b) Histograma del experimento.

Figura 5.7: Estimación de la dirección de arribo utilizando tres fuentes simultáneas posicionadas en $\theta_1 = -40$, $\theta_2 = 0$ y $\theta_3 = 40$ grados.

grados con una estimación de $\tilde{\theta}_{fin} = -20$ grados. En la Figura 5.8(a) se muestra la gráfica del seguimiento de la trayectoria de la fuente, mientras que la gráfica 5.8(b) representa el histograma de dicha trayectoria.

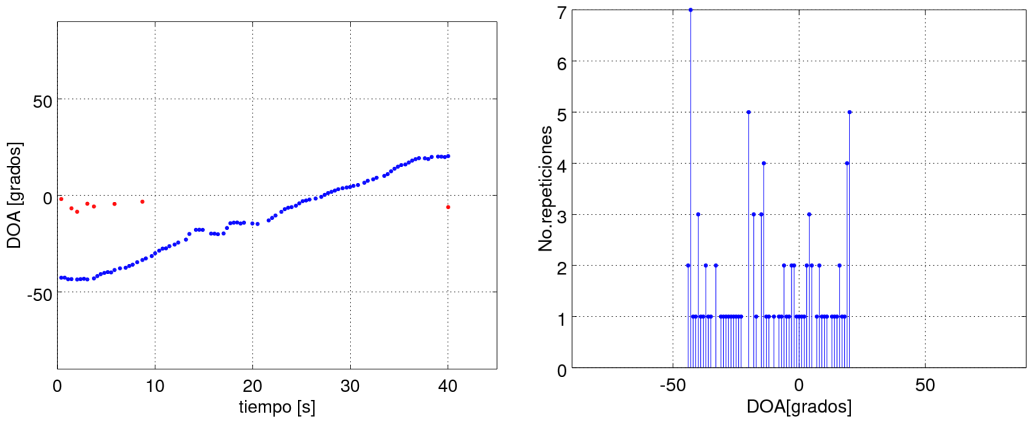
A partir del histograma de la Figura 5.8, se puede observar que el movimiento estimado de la fuente es lineal salvo la primera dirección de arribo que es estática, dando como resultado la forma de una distribución uniforme discreta. Dicha distribución depende del tipo de movimiento, velocidad y aceleración que sufra la fuente de voz, de tal manera que cuando la trayectoria de una fuente de voz sea aleatoria, la distribución que describe su desplazamiento no podrá ser modelada matemáticamente. Aunado a lo anterior en conjunto con el tipo de distribución que presentan los experimentos de fuentes estáticas, el clasificador implementado supone que la dirección de arribo de una fuente de voz en movimiento presenta una distribución Gaussiana en el instante $t$ de estimación, como se explicó en la Sección 4.3.




(a) Seguimiento de la fuente de voz. (b) Histograma del experimento.

Figura 5.8: Estimación del ángulo de arribo utilizando una fuente de voz con desplazamiento lineal.

De manera similar, los siguientes experimentos se realizaron con fuentes en movimiento pero con diferente trayectoria y velocidad. En la Figura 5.9(a) se muestra el seguimiento de una fuente de voz que tuvo una trayectoria que tiende a ser lineal con velocidad promedio de 1.03 $\frac{grados}{segundos}$. La velocidad promedio puede tomar valores tanto positivos como negativos, donde los valores de velocidad positivos representan un desplazamiento de la fuente en sentido horario mientras que los negativos representan al movimiento antihorario de la fuente de voz. La Figura 5.9(b) muestra el histograma de las direcciones de arribo estimadas de la fuente de voz durante el experimento.




(a) Seguimiento de la fuente en movimiento.

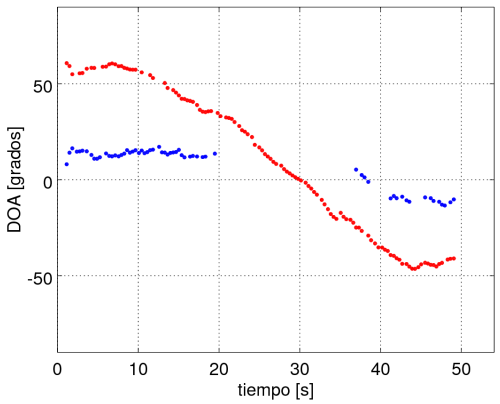
(b) Histograma.

Figura 5.9: Dirección de arribo para una fuente en voz con trayectoria de $\theta_{ini} = -50$ a $\theta_{fin} = 25$.

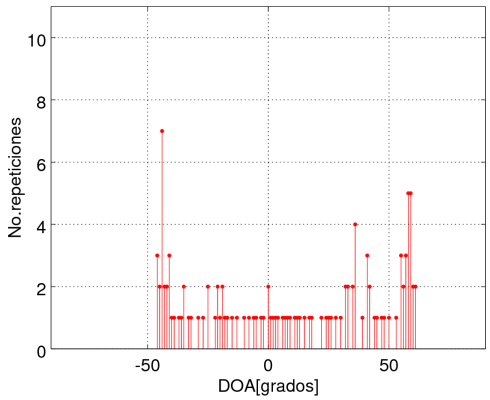
Finalmente, se efectuó un experimento en donde la fuente de voz permaneció estática durante un lapso de tiempo al principio y al final del mismo. El movimiento de la fuente de voz tiende a ser lineal, sin embargo, tiene variación en su velocidad de desplazamiento. En la Figura 5.10 se presentan dos gráficas que describen el comportamiento de la fuente durante el experimento, la Figura 5.10(a) representa el seguimiento de la fuente de voz y la Figura 5.10(b) el histograma resultante de la fuente de voz de interés, además la Figura 5.10(c) muestra el diagrama espacial del experimento.

En la gráfica de seguimiento (Figura 5.10(a)) se puede observar que el sistema estima la dirección de arribo de una fuente de sonido extra a la fuente de interés (Curva de color rojo), dicha dirección de arribo se la atribuimos a una posible fuente de ruido no deseada o de la misma manera, a una posible respuesta de dirección errónea que el sistema construya en algunas de las frecuencias de análisis. Cabe destacar que las direcciones de arribo de las fuentes ruidosas son estimadas cuando el ángulo de arribo de la fuente de interés es lejano al origen ($\theta < -40$ grados y $\theta > 40$ grados), de tal manera que la mejor estimación de dirección de arribo en el presente experimento es cuando la fuente de interés se encuentra en el intervalo $-30 < \theta < 40$ grados.

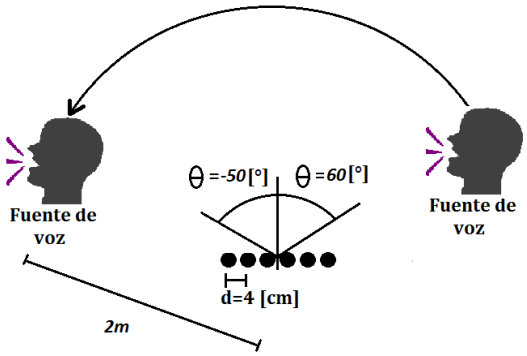
El desplazamiento que describe la fuente de voz en la gráfica de seguimiento (5.10(a)) concuerda con el movimiento real que sufrió la fuente, siendo estática al principio y al final del experimento, y un desplazamiento que tiende a ser lineal en los instantes de movimiento. Dicho




(a) Seguimiento de la fuente.




(b) Histograma de la fuente de voz de interés.




(c) Trayectoria de la fuente de voz de interés.

Figura 5.10: Estimación de la dirección de arribo utilizando una fuente de voz con trayectoria $\theta_{ini} = 60$ a $\theta_{fin} = -50$.

movimiento se puede interpretar en la Figura 5.10(b) donde se puede observar que el histograma tiende a una distribución uniforme discreta en el intervalo de $-45 \leq \theta \leq 60$ grados, salvo los valores en el extremo de la distribución que representan el movimiento estático al principio y al final del experimento.

## 5.2. Resultados con el arreglo circular de seis micrófonos

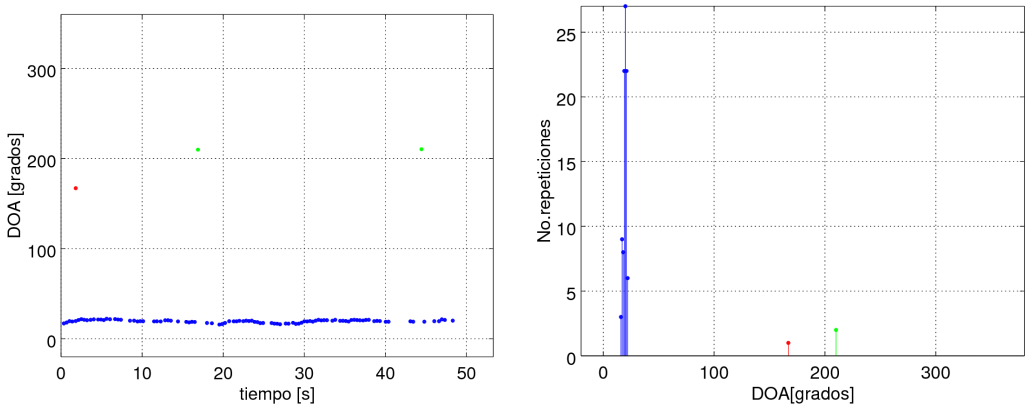
De la misma manera que con el arreglo lineal, se realizaron experimentos, permaneciendo estático el arreglo circular de micrófonos, con una fuente de voz estática y en movimiento, así como con fuentes de voz simultáneas. A diferencia del arreglo lineal de micrófonos, el intervalo de detección para la estimación dirección de arribo es de $0 \leq \theta < 360$ grados.

En las siguientes secciones se presentan los resultados obtenidos a través de los experimentos realizados utilizando el arreglo circular de seis micrófonos, como el que se mostró en la Figura 4.3(b). En la Sección 5.2.1 se muestran los resultados de un experimento con una fuente

de voz estática, en la Sección 5.2.2 los resultados con una fuente de voz en movimiento y finalmente en la Sección 5.2.3 se exponen los resultados con fuentes de voz simultáneas.

### 5.2.1. Resultados de una fuente de voz estática

Los resultados de los experimentos con fuentes de voz estáticas se muestran en las Figuras 5.11 y 5.12. En dichas figuras se exhiben dos pruebas diferentes, donde para cada experimento se muestra una gráfica de seguimiento de la fuente de voz y el histograma. El primer experimento mostrado es de una fuente de voz estática con ángulo de dirección de arribo $\theta_1 = 20$ grados. La Figura 5.11(a) muestra la gráfica de seguimiento de dicha fuente de voz, donde se puede observar que la fuente presenta intervalos de tiempo como fuente inactiva, no obstante, cuando se vuelve fuente activa el sistema la clasifica de manera correcta. También se pueden observar que algunas direcciones de arribo en ángulos superiores que no corresponden a las de la fuente de voz de interés, dichas direcciones de arribo se relacionaron con fuentes de ruido acústico.



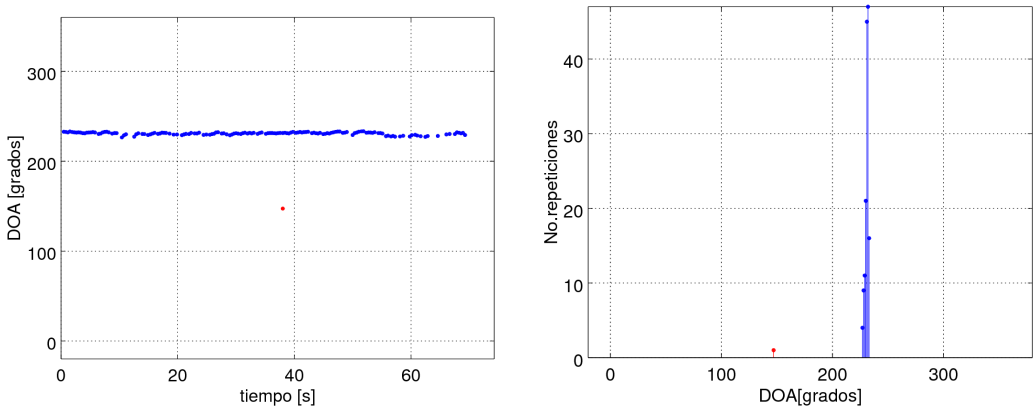

(a) Seguimiento de la fuente de voz. (b) Histograma del experimento.

Figura 5.11: Estimación de la dirección de arribo para una fuente de voz estática con ángulo de arribo en $\theta = 20$ grados, utilizando un arreglo circular de micrófono.

El histograma de la Figura 5.11(b) muestra la media y la variación de la estimación de dirección de arribo de la fuente. La media obtenida a través del experimento es de $\theta_\mu = 19.6$ grados. A pesar de que la ocurrencia de la fuente es muy cercana a la dirección de arribo real, la estimación presenta una variación mayor que los resultados de los experimentos de fuentes de voz estáticas realizados con el arreglo lineal de micrófonos (Sección 5.1.1). El experimento presenta una varianza de $\sigma_{\theta_1} = 1.98$, de tal manera que el máximo valor estimado es de $\theta_{1_{max}} = 22$ grados y el mínimo de $\theta_{1_{min}} = 16$ grados.

Las gráficas de las Figuras 5.12(a) y 5.12(b), muestran la respuesta de seguimiento de una fuente de voz estática con dirección de arribo en $\theta_3 = 240$ grados, considerando que pudiera existir un error de paralelismo en la medición del ángulo de arribo real, y el histograma del experimento respectivamente. La media estimada del experimento realizado es $\theta_{3_\mu} = 231$ grados con varianza de $\sigma_{\theta_3} = 2.337$, teniendo un ángulo de arribo máximo de $\theta_{3_{max}} = 233$ y mínimo de

$\theta_{3_{min}} = 225$ grados.




(a) Seguimiento de la fuente voz. (b) Histograma del experimento.

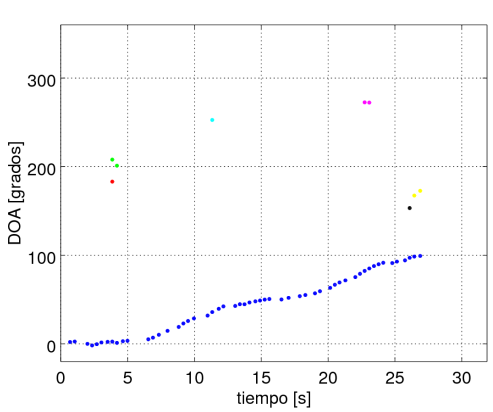
Figura 5.12: Estimación de la dirección de arribo de la señal proveniente de una fuente de voz estática con ángulo de arribo $\theta = 240$ grados.

### 5.2.2. Resultados de una fuente de voz en movimiento

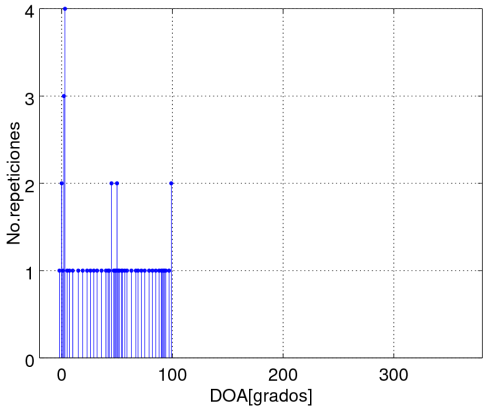
Se realizaron dos experimentos para calcular la estimación de dirección de arribo de fuentes de voz en movimiento, en ambos casos el movimiento de las fuentes tiende a ser lineal, sin embargo, tienen variación en la velocidad de desplazamiento de la misma. La velocidad promedio de la fuente se calculó con base a 15 estimaciones de velocidad instantánea anteriores al tiempo de estimación $t$.

En la Figura 5.13 se muestra el comportamiento de la fuente de voz en movimiento. Dicho experimento comenzó con la fuente de voz posicionada en el ángulo de arribo $\theta_{ini} = 0$ grados, y finalizó con el ángulo de arribo en $\theta_{fin} = 100$ grados. La Figura 5.13(a) muestra el seguimiento de la fuente de voz, la cual realizó su desplazamiento en 28 segundos. El desplazamiento estimado de la fuente de voz está definido por el ángulo inicial y final de dirección de arribo, los cuales fueron calculados como $\tilde{\theta}_{ini} = 2$ grados y $\tilde{\theta}_{fin} = 99$ grados.

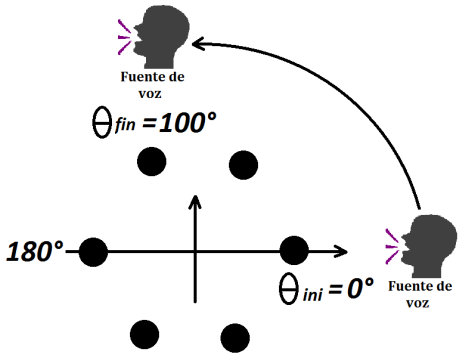
En el histograma de la Figura 5.13(b) se muestra la distribución de los valores estimados de fuente de interés durante el experimento. Se puede observar que la distribución tiende a ser uniforme dentro del intervalo de desplazamiento de la fuente, sin embargo, el movimiento no es completamente lineal porque presenta repeticiones de estimación de dirección de arribo en el ángulo $\theta = 50$ grados , dicha repetición se puede observar en la gráfica de seguimiento en el intervalo $14 < t < 18$ segundos donde la fuente tiene un movimiento que tiende a ser estático. La Figura 5.13(c) muestra un diagrama espacial que describe el desplazamiento de la fuente de voz en este experimento.




(a) Seguimiento de la fuente de voz.




(b) Histograma de la fuente de voz.

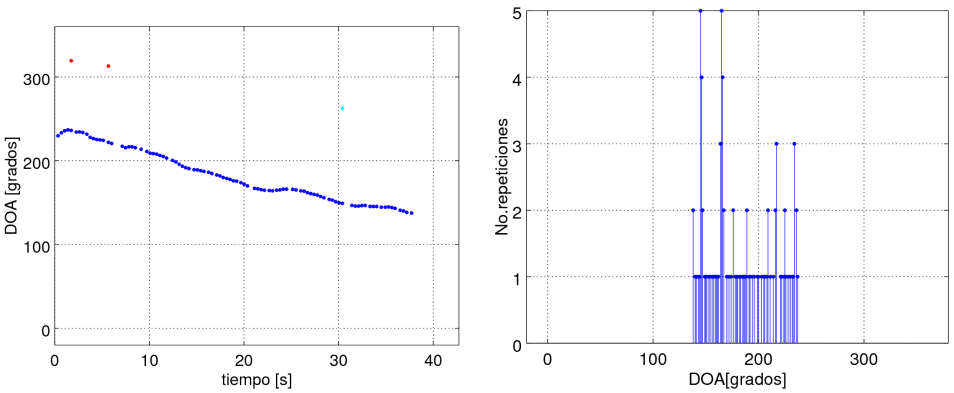



(c) Diagrama espacial.

Figura 5.13: Estimación de la dirección de arribo para una fuente de voz con desplazamiento de $\theta_{ini} = 0$ a $\theta_{fin} = 100$ grados.

Finalmente, se realizó otro experimento, que de igual manera consistió en describir la trayectoria de una fuente de voz, en dicha prueba el comportamiento es lineal como en el caso anterior, sin embargo, en este experimento su movimiento es en sentido horario. El ángulo de dirección de arribo inicial de la fuente de voz es $\theta_{ini} = 245$ grados y el final en $\theta_{fin} = 130$ grados. La gráfica de seguimiento de la fuente de voz en dicho experimento se puede observar en la Figura 5.14(a), donde se puede apreciar que el movimiento de la fuente tiende a ser lineal. También se pueden observar algunas direcciones de arribo estimadas que no pertenecen a dicha fuente de voz, estas direcciones de arribo pertenecen a fuentes sonoras no deseadas que generaron una emisión de señales en instantes específicos cuando se realizaba el experimento.

La Figura 5.14(b) muestra el histograma de las direcciones de arribo estimadas en el experimento, donde se puede apreciar que no es una distribución uniforme completamente, por lo que el movimiento de la fuente de voz estimada no es del todo lineal. El ángulo de arribo inicial estimado fue en $\tilde{\theta}_{ini} = 237$ grados y el final en $\tilde{\theta}_{fin} = 138$ grados. La fuente de voz experimentó una velocidad promedio de $1.61 \frac{grados}{s}$ en sentido horario.




(a) Seguimiento de la fuente de voz. (b) Histograma de la fuente de voz.

Figura 5.14: Estimación de dirección de arribo de una fuente de voz con desplazamiento de $\theta_{ini} = 245$ a $\theta_{fin} = 130$ grados.

### 5.2.3. Resultados con múltiples fuentes de voz simultáneas

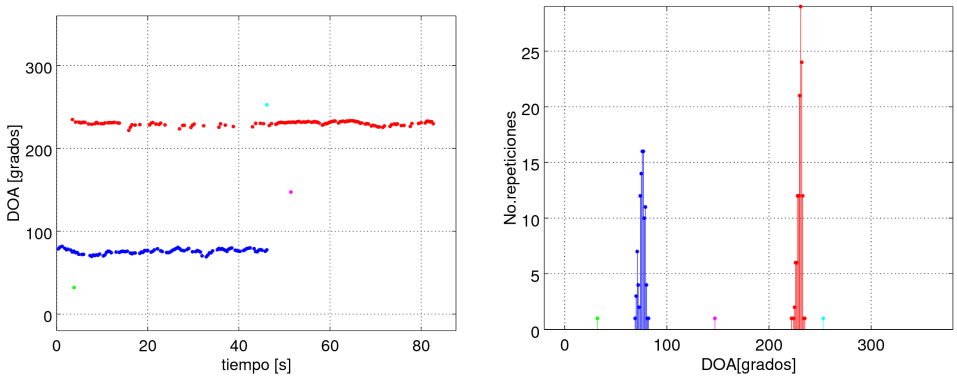
En la presente sección se presentan los resultados de dos experimentos donde se estimó la dirección de arribo de dos fuentes de voz simultaneas. El primer experimento se realizó con las dos fuentes estáticas, dichas fuentes con ángulo de dirección de arribo en $\theta_1 = 80$ y $\theta_2 = 240$ grados, respectivamente. En el experimento, la emisión de las señales por la fuente con dirección de arribo en 80 grados inicia segundos antes que la fuente en 240 grados, sin embargo, tal fuente se vuelve inactiva a partir del segundo 48 mientras que la fuente de voz con DOA en 240 grados permanece activa el resto del experimento.

En la Figura 5.15 se muestra el resultado de la estimación de dirección de arribo durante el experimento por medio de las gráficas de seguimiento (Figura 5.15(a)) e histograma (Figura 5.15(b)). La media de las direcciones de arribo de las fuentes de voz estimadas se centran en $\theta_{1_\mu} = 78$ y $\theta_{2_\mu} = 230$ grados respectivamente. La fuente de voz con DOA en 80 grados tiene una varianza de $\sigma_1 = 5.03$ con una estimación de dirección de arribo máxima en $\theta_{1_{max}} = 80$ grados y mínima en $\theta_{2_{min}} = 69$, mientras que la estimación de las direcciones de arribo de la otra fuente voz tiene una varianza de $\sigma_2 = 4.48$ con ángulo de arribo máximo en $\theta_{2_{max}} = 234$ y mínimo en $\theta_{2_{min}} = 222$ grados.

El siguiente experimento se realizó con una fuente de voz estática con ángulo de arribo en $\theta_1 = 80$ grados y una fuente de voz en movimiento, tal movimiento es de tipo lineal con dirección de arribo inicial en $\theta_{2_{ini}} = 240$ y final $\theta_{2_{fin}} = 130$ grados. La fuente de voz sufrió una velocidad promedio de $1.61 \frac{grados}{s}$ con movimiento en sentido horario.

En la Figura 5.16 se muestra la gráfica de seguimiento de las fuentes, donde la media de la fuente estática dio como resultado en $\theta_{1_\mu} = 78$ grados, mientras que los valores inicial y final de la fuente de voz en movimiento fueron $\theta_{2_{ini}} = 237$ y $\theta_{2_{min}} = 139$ grados respectivamente.

La velocidad promedio de la fuente de voz con movimiento permanece constante en ciertos




(a) Seguimiento de las fuentes. (b) Histograma del experimento.

Figura 5.15: Estimación de la dirección de arribo utilizando dos fuentes de voz simultáneas con ángulo de arribo en $\theta_1 = 80$ y $\theta_2 = 240$ grados.

intervalos de tiempo tomando valores desde $v = 0.98\ \frac{grados}{s}$ hasta $v = 3.10\ \frac{grados}{s}$, los valores de la velocidad promedio que corresponden a la fuente de voz con movimiento en sentido horario.

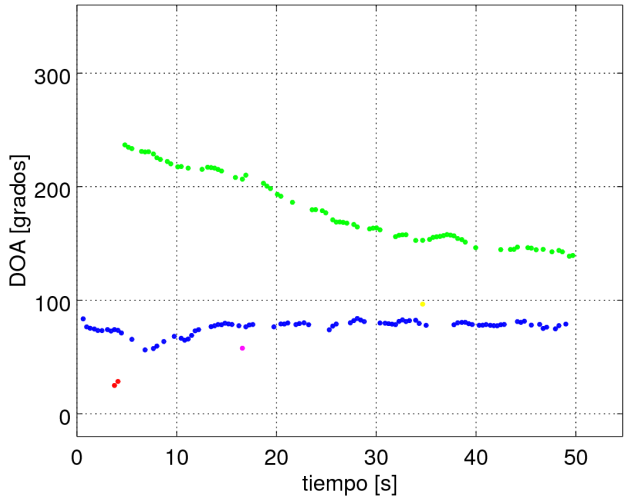



Figura 5.16: Estimación de la dirección de arribo utilizando dos fuentes de voz simultáneas, una fuente estática con ángulo de arribo $\theta_1 = 80$ grados y una fuente con desplazamiento de $\theta_{2_{ini}} = 245$ a $\theta_{2_{fin}} = 130$ grados.

## 5.3. Resumen

En el presente capítulo se mostraron los resultados obtenidos de la estimación de la dirección de arribo de una y múltiples señales provenientes de fuentes de voz, utilizando dos arreglos de micrófonos, el arreglo lineal uniforme y el arreglo circular, ambos con seis micrófonos.

Los resultados con el arreglo lineal de micrófonos dieron resultados más asertivos y con menor variación en el ángulo de arribo, sin embargo, los resultados mostraron que cuando la estimación del ángulo de dirección de arribo de la fuente de voz es mayor a 35 grados, el sistema es más susceptible a estimar la dirección de arribo de fuentes de voz ruidosas o reverberación con direcciones de arribo cercanas al origen. Por otro lado, los experimentos realizados con el arreglo circular de micrófonos mostraron mayor variación en la estimación de dirección de arribo en las pruebas con fuentes de voz estáticas, no obstante, tuvieron un valor más cercano en la media de las estimaciones en comparación con el arreglo lineal. En la Tabla 5.2 se muestra una comparativa de los resultados obtenidos con el arreglo lineal de micrófonos para fuentes de voz estáticas, donde se puede observar que entre más alejado está el ángulo de arribo al ángulo de origen aumenta la variación del ángulo de arribo.

En la Tabla 5.3 se muestran los resultados de los experimentos realizados utilizando el arreglo circular de seis micrófonos, donde se puede observar que la máxima variación de estimación de dirección de arribo llega hasta 18 grados, la mínima es de ocho grados y manteniéndose en promedio con una variación de 10 grados, mientras que las del arreglo lineal la máxima es de 13 grados para $\theta = 45$ grados, con una mínima de 2 grados.

| Posición de la fuente de voz | Media estimada | valor mínimo $\tilde{\theta}_{min}$ | valor máximo $\tilde{\theta}_{max}$ |
|---|---|---|---|
| $\theta = 0$ grados | 1.43 grados | 0 grados | 2 grados |
| $\theta = 15$ grados | 9.015 grados | 7 grados | 10 grados |
| $\theta = 20$ | 15 grados | 14 grados | 16 grados |
| $\theta = -10$ | -6.27 grados | -8 grados | -5 grados |
| $\theta = -25$ | -20.05 grados | -21 grados | -19 grados |
| $\theta = 45$ | 38 grados | 30 grados | 41 grados |
| $\theta = 35$ grados | 29.6 grados | 27 grados | 30 grados |

Cuadro 5.2: Características de los arreglos de micrófonos implementados.

| Posición de la fuente de voz | Media estimada | valor mínimo $\tilde{\theta}_{min}$ | valor máximo $\tilde{\theta}_{max}$ |
|---|---|---|---|
| $\theta = 0$ grados | -0.03 grados | -9 grados | 3 grados |
| $\theta = 20$ grados | 19.61 grados | 15 grados | 25 grados |
| $\theta = 80$ grados | 77.65 grados | 74 grados | 81 grados |
| $\theta = 135$ grados | 126 grados | 123 grados | 131 grados |
| $\theta = 180$ grados | 173.68 grados | 169 grados | 178 grados |
| $\theta = 240$ grados | 231 grados | 225 grados | 233 grados |
| $\theta = 290$ grados | 279.24 grados | 274 grados | 283 grados |
| $\theta = 330$ grados | 323 grados | 317 grados | 330 grados |

Cuadro 5.3: Comparación de los resultados de los experimentos con fuentes de voz estáticas utilizando el arreglo circular de micrófonos.

De la misma manera, los resultados de la estimación del ángulo de arribo de las señales provenientes de fuentes con desplazamiento utilizando el arreglo circular de micrófonos re-

presentan una mejor estimación con respecto a los obtenidos utilizando el arreglo lineal, esto se puede observar comparando las gráficas del histograma de las fuentes de voz, en donde las distribuciones de los ángulos de arribo estimados en los histogramas se aproximan más a una distribución uniforme discreta.

# Capítulo 6.

# Conclusiones

En el presente trabajo se diseñó e implementó un sistema que estima la dirección de arribo de señales de voz provenientes de diferentes fuentes posicionadas dentro de un intervalo angular de $-90 \leq \theta \leq 90$ grados, para el arreglo lineal de micrófonos y $0 \leq \theta < 360$ grados, con el arreglo circular de micrófonos. Dicho sistema tiene un tiempo de respuesta $t_{repuesta} = 348$ milisegundos cuando la emisión de las señales por la fuente de voz es continua.

Se compararon las respuestas de dirección promedio generadas con un formador de haz fijo convencional y el método propuesto, el cual utiliza los elementos del eigenvector más representativo que es relacionado con las señales adquiridas. En dicha comparación se observó un patrón de respuesta de dirección promedio mejor definido, sobre todo cuando se tienen dos o más fuentes de voz activas de manera simultánea.

A diferencia de la teoría del método MUSIC, la respuesta de dirección se obtuvo utilizando únicamente el eigenvector asociado al mayor eigenvalor, el cual, mostró una gran diferencia con respecto al valor del resto de los eigenvectores. Lo anterior se lo atribuimos a la posible correlación entre las señales emitidas por las fuentes de voz; de tal manera que se utilizó un único eigenvector para representar a las señales activas en el subespacio de las señales, toman-

do en cuenta que los experimentos se realizaron dentro de un ambiente acústico no controlado.

Se utilizaron dos tipos de arreglos de micrófonos, un arreglo lineal uniforme y un arreglo circular equidistante, ambos con seis elementos. Los resultados obtenidos con el arreglo lineal de micrófonos mostraron menor variación en la estimación del ángulo de arribo en los experimentos con fuentes estáticas; dicha variación fue de $\pm 1$ grado con respecto a la media, además, ésta aumenta considerablemente cuando la dirección de arribo de una fuente se encuentra fuera del intervalo $-40 \leq \theta \leq 40$ grados, estimando un DOA mínimo de $\theta_{min} = 8$ grados y máximo de $\theta_{max} = 3$ grados alrededor de la media.

La variación mínima obtenida en las pruebas realizadas con el arreglo circular de micrófonos fue de $\pm 4$ grados y máxima de $\pm 7$, sin embargo, la media de los DOA calculada se acerca más a la dirección de arribo de la fuente de voz real, con un error mínimo de un grado. No obstante, dicho error entre la media de las estimaciones de los DOA y la dirección de arribo real de la fuente depende, en gran medida, al error de paralelismo en la medición del ángulo de arribo real.

Por otro lado, el seguimiento de las fuentes de voz con desplazamiento se realizó implementando un filtro de Kalman, estimando el estado real con base a las observaciones ruidosas. A pesar de que la distribución de los DOA de una fuente en movimiento depende del tipo de desplazamiento que ésta experimente, se puede suponer que dicha fuente presenta una distribución tipo Gaussiana en el instante de estimación $t$, de tal manera que la media de la distribución en el tiempo $t+1$ puede cambiar. Aunado a estas pruebas, se pudo implementar un clasificador estadístico que compara la máxima verosimilitud de la observación, dadas las fuentes de voz activas e inactivas.

El clasificador implementado en el presente trabajo agrupa los ángulos de arribo estimados con las fuentes seguidas. Dicha clasificación se realiza con base en los valores de los DOA y no por las características sonoras de la fuente, dando prioridad a la clasificación con las fuentes de voz activas en el tiempo $t$.

## Trabajo a futuro

Se propone interactuar con una etapa de filtrado espacial, de tal manera que se pueda obtener la información procedente de una de las fuentes de voz de interés, aunque se tengan múltiples fuentes de voz simultáneas; además, con dicha información mejorar el clasificador tomando en cuenta las características de las señales de voz emitidas por cada una de las fuentes. De la misma manera, realizar un análisis teórico en el cual se decremente la frecuencia de muestreo para evitar el solapamiento de algunos de los *frames* con información de las señales de entrada, sin afectar la respuesta del sistema, y finalmente, extender dicho sistema para estimar el ángulo de dirección de arribo de elevación $\phi$.

# Bibliografía


[1] Donald E Hall. *Basic acoustics*. Wiley, New-York, 1993.

[2] Sergios Theodoridis and Rama Chellappa. *Array and Statistical Signal Processing*, volume 3. Academic Press Library in Signal Processing, 2013.

[3] R Venkatesha Prasad, Abhijeet Sangwan, HS Jamadagni, MC Chiranth, Rahul Sah, and Vishal Gaurav. Comparison of voice activity detection algorithms for voip. In *Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on*, pages 530–535. IEEE, 2002.

[4] Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280, 1986.

[5] Barry D Van Veen and Kevin M Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine*, 5(2):4–24, 1988.

[6] Ewi Liu and Stephan Weiss. *Wideband Beamforming Concepts and techniques*. Wiley, 2010.

[7] Thomas Chou. Frequency-independent beamformer with low response error. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 5, pages 2995–2998. IEEE, 1995.

[8] Hamid Krim and Mats Viberg. Two decades of array signal processing research: the parametric approach. *IEEE signal processing magazine*, 13(4):67–94, 1996.

[9] James H Justice. *Array signal processing*. Prentice Hall, New Jersey, 1985.

[10] Jacob Benesty, Jingdong Chen, and Yiteng Huang. *Microphone array signal processing*, volume 1. Springer Science & Business Media, 2008.

[11] MA Alrmah, Stephan Weiss, Soydan Redif, Sangarapillai Lambotharan, and John G McWhirter. Angle of arrival estimation for broadband signals: A comparison. In *Intelligent Signal Processing Conference 2013 (ISP 2013), IET*, pages 1–6. IET, 2013.

[12] Sylvain Argentieri and Patrick Danes. Broadband variations of the music high-resolution method for sound source localization in robotics. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 2009–2014. IEEE, 2007.

[13] J-M Valin, François Michaud, Jean Rouat, and Dominic Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1228–1233. IEEE, 2003.

[14] Jiri Tuma, Patrik Janecka, Milan Vala, and Lukas Richter. Sound source localization. In *Carpathian Control Conference (ICCC), 2012 13th International*, pages 740–743. IEEE, 2012.

[15] Ivan Meza, Caleb Rascon, Gibran Fuentes, and Luis A Pineda. On indexicality, direction of arrival of sound sources, and human-robot interaction. *Journal of Robotics*, pages 1–13, 2016.

[16] Enzo Mumolo, Massimiliano Nolich, and Gianni Vercelli. Algorithms for acoustic localization based on microphone array in service robotics. *Robotics and Autonomous systems*, 42(2):69–88, 2003.

[17] Rong Liu and Yongxuan Wang. Azimuthal source localization using interaural coherence in a robotic dog: modeling and application. *Robotica*, 28(07):1013–1020, 2010.

[18] Kazuhiro Nakadai, Hiroshi G Okuno, Hiroaki Kitano, et al. Real-time sound source localization and separation for robot audition. In *INTERSPEECH*, 2002.

[19] Hong Wang and Peter Chu. Voice source localization for automatic camera pointing system in videoconferencing. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 1, pages 187–190. IEEE, 1997.

[20] V Krishnaveni, T Kesavamurthy, and B Aparna. Beamforming for direction-of-arrival (doa) estimation-a survey. *International Journal of Computer Applications*, 61(11), 2013.

[21] Simon Doclo. *Multi-microphone noise reduction and dereverberation techniques for speech applications*. PhD thesis, Katholieke Universiteit Leuven, 2003.

[22] Henry Cox, Robertm Zeskind, and Markm Owen. Robust adaptive beamforming. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(10):1365–1376, 1987.

[23] Aboulnasr Hassanien, Shahram Shahbazpanahi, and Alex B Gershman. A generalized capon estimator for localization of multiple spread sources. *IEEE Transactions on Signal Processing*, 52(1):280–283, 2004.

[24] N Jablon. Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections. *IEEE Transactions on Antennas and Propagation*, 34(8):996–1012, 1986.

[25] P Mowlaee Begzade Mahale. Robust adaptive crls-gsc algorithm for doa mismatch in microphone array. In *Signal Processing Conference, 2008 16th European*, pages 1–5. IEEE, 2008.

[26] Prabhakar S Naidu. *Sensor array signal processing*. CRC press, USA, 2009.

[27] Eugen Skudrzyk. *The foundations of acoustics: basic mathematics and basic acoustics*. Springer Science & Business Media, New York, 2012.

[28] Robert J Mailloux. *Phased array antenna handbook*, volume 2. Artech House Boston, 2005.

[29] F Dunn, WM Hartmann, DM Campbell, and NH Fletcher. *Handbook of acoustics*. Springer, New York, 2015.

[30] Shing-Chow Chan and Carson KS Pun. On the design of digital broadband beamformer for uniform circular array with frequency invariant characteristics. In *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, volume 1, pages I–I. IEEE, 2002.

[31] S Gökhun Tanyer and Hamza Ozer. Voice activity detection in nonstationary noise. *IEEE Transactions on speech and audio processing*, 8(4):478–482, 2000.

[32] John G Proakis, Dimitris G Manolakis, Verónica Santalla del Rio, and José Luis Alba Castro. *Tratamiento digital de señales*, volume 3. Prentice Hall, Madrid, 1998.

[33] Alan V Oppenheim and Ronald Schafer. *Tratamiento de señales en tiempo discreto*. 2000.

[34] Kirill Sakhnov, Ekaterina Verteletskaya, and Boris Simak. Approach for energy-based voice detector with adaptive scaling factor. *IAENG International Journal of Computer Science*, 36(4):394, 2009.

[35] Philippe Renevey and Andrzej Drygajlo. Entropy based voice activity detection in very noisy conditions. *threshold*, 5(5.5):6, 2001.

[36] Hassan Elkamchouchi and Mohamed Abd Elsalam Mofeed. Direction-of-arrival methods (doa) and time difference of arrival (tdoa) position location technique. In *Radio Science Conference, 2005. NRSC 2005. Proceedings of the Twenty-Second National*, pages 173–182. IEEE, 2005.

[37] Revati Joshi and Ashwinikumar Dhande. Direction of arrival estimation using music algorithm. *signal*, 1(1):3, 2014.

[38] Andy Vesa. Direction of arrival estimation using music and root-music algorithm. In *18th Telecommunications Forum, Pg*, pages 582–585, 2010.

[39] Tanuja S Dhope. Application of music, esprit and root music in doa estimation. *Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia*, 2010.

[40] Vázquez S. Samuel S. *Reducción de ruido en señales de voz mediante un formador de haz*. Tesis de licenciatura, Universidad Nacional Autónoma de México, México, 2012.

[41] Jack Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.

[42] François Grondin, Dominic Létourneau, François Ferland, Vincent Rousseau, and François Michaud. The manyears open framework. *Autonomous Robots*, 34(3):217–232, 2013.

[43] *Jack Audio-Connection-Kit*. 22-5-2017(Online).

[44] Jean-Marc Valin, François Michaud, and Jean Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55(3):216–228, 2007.

[45] Greg Welsh and Gary Bishop. An introduction to the kalman filter. *University of North Carolina at Chapel Hill Chapel Hill NC*, 95:95–041, 1995.

# Apéndice A.

# Factor del arreglo

El factor de arreglo para el caso de un arreglo de micrófonos lineal en banda angosta se puede considerar a partir de la Ecuación (A.1).

$$H(\theta) = \sum_{m=1}^{M} a_m e^{iKd(m-1)\sin\theta} \tag{A.1}$$

Donde $a_m$ es el factor de escala de cada uno de los micrófonos. Si consideramos a $a_m = 1$ y $G = Kd\sin\theta$, además se desarrolla la Ecuación (A.1) se obtiene:

$$H(\theta) = 1 + e^{jG} + e^{j2G} + \cdots + e^{(M-2)jG} + e^{(M-1)jG} \tag{A.2}$$

Multiplicado la Ecuación (A.2) por $e^{jG}$ tenemos:

$$H(\theta)\, e^{jG} = e^{jG} + e^{j2G} + \cdots + e^{(M-1)jG} + e^{MjG} \tag{A.3}$$

Si realizamos la sustracción entre las ecuaciones (A.3) y (A.2) se obtiene la Ecuación (A.4).

$$H(\theta)\, e^{jG} - H(\theta) = e^{jG} + e^{j2G} + \cdots + e^{MjG} - \left(1 + e^{jG} + e^{j2G} + \cdots + e^{(M-)jG}\right) \tag{A.4}$$

Desarrollando la Ecuación (A.4) tenemos:

$$H(\theta) = \frac{e^{jMG} - 1}{e^{jG} - 1} \tag{A.5}$$

Factorizando $e^{j\frac{MG}{2}}$ en el numerador de la Ecuación (A.5) y $e^{\frac{jG}{2}}$ de la misma, obtenemos la Ecuación (A.6).

$$H(\theta) = \frac{e^{j\frac{MG}{2}}\left(e^{j\frac{MG}{2}} - e^{-j\frac{MG}{2}}\right)}{e^{j\frac{G}{2}}\left(e^{j\frac{G}{2}} - e^{-j\frac{G}{2}}\right)} \tag{A.6}$$

Sabemos que la función seno se puede representar en la forma de Euler como se muestra en la Ecuación (A.7).

$$\sin x = \frac{e^{jx} - e^{-jx}}{2j} \tag{A.7}$$

Por lo que al sustituir la Ecuación (A.7) en (A.6) y $G = Kd\sin\theta$, se obtiene el factor del arreglo con amplitudes unitarias en los micrófonos como se muestra en la Ecuación (A.8).

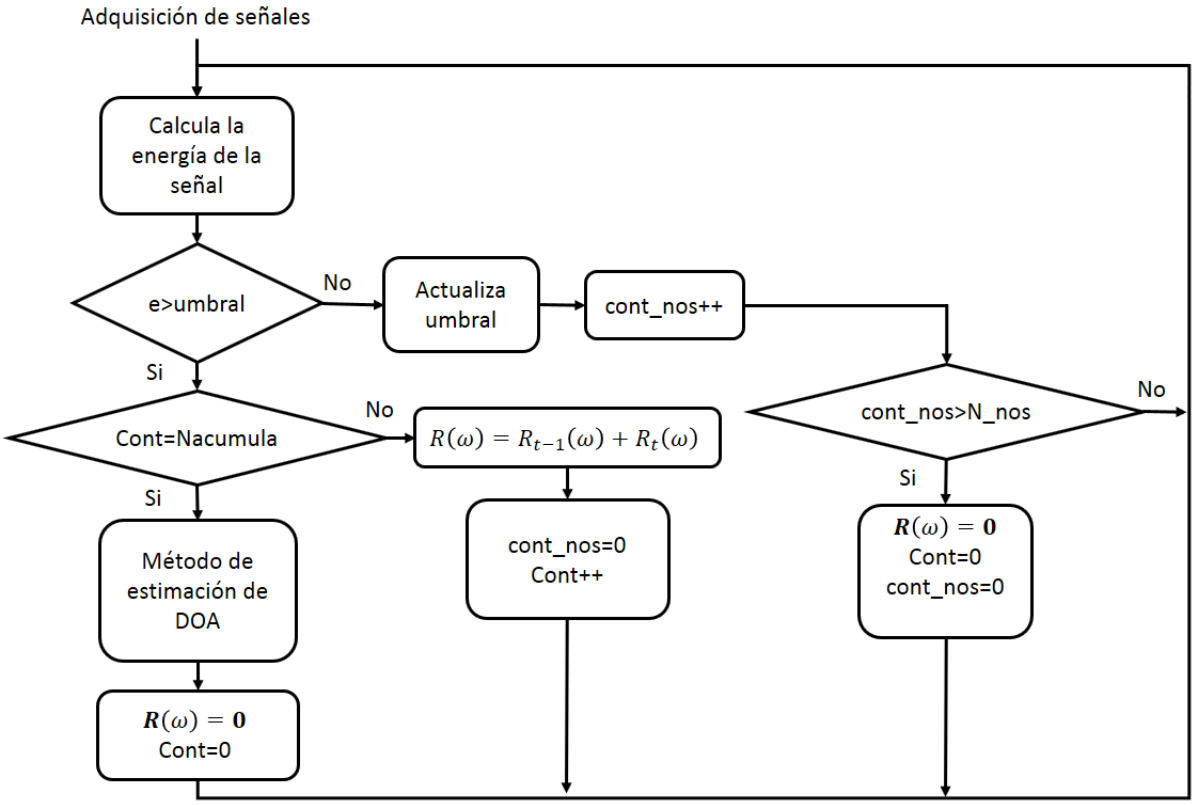
$$H(\theta) = \frac{\sin\left(\frac{\pi M d \sin\theta}{\lambda}\right)}{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)} e^{j\frac{\pi d \sin\theta}{\lambda}} \tag{A.8}$$

De tal manera que al obtener la magnitud de la Ecuación (A.8) se obtiene el patrón de radiación para un arreglo de micrófonos lineal uniforme de banda angosta como se observa en la Ecuación (A.9).

$$|H(\theta)| = \left| \frac{\sin\left(\frac{\pi M d \sin\theta}{\lambda}\right)}{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)} \right| \tag{A.9}$$

# Apéndice B.

# Diagrama de flujo de la implementación del sistema



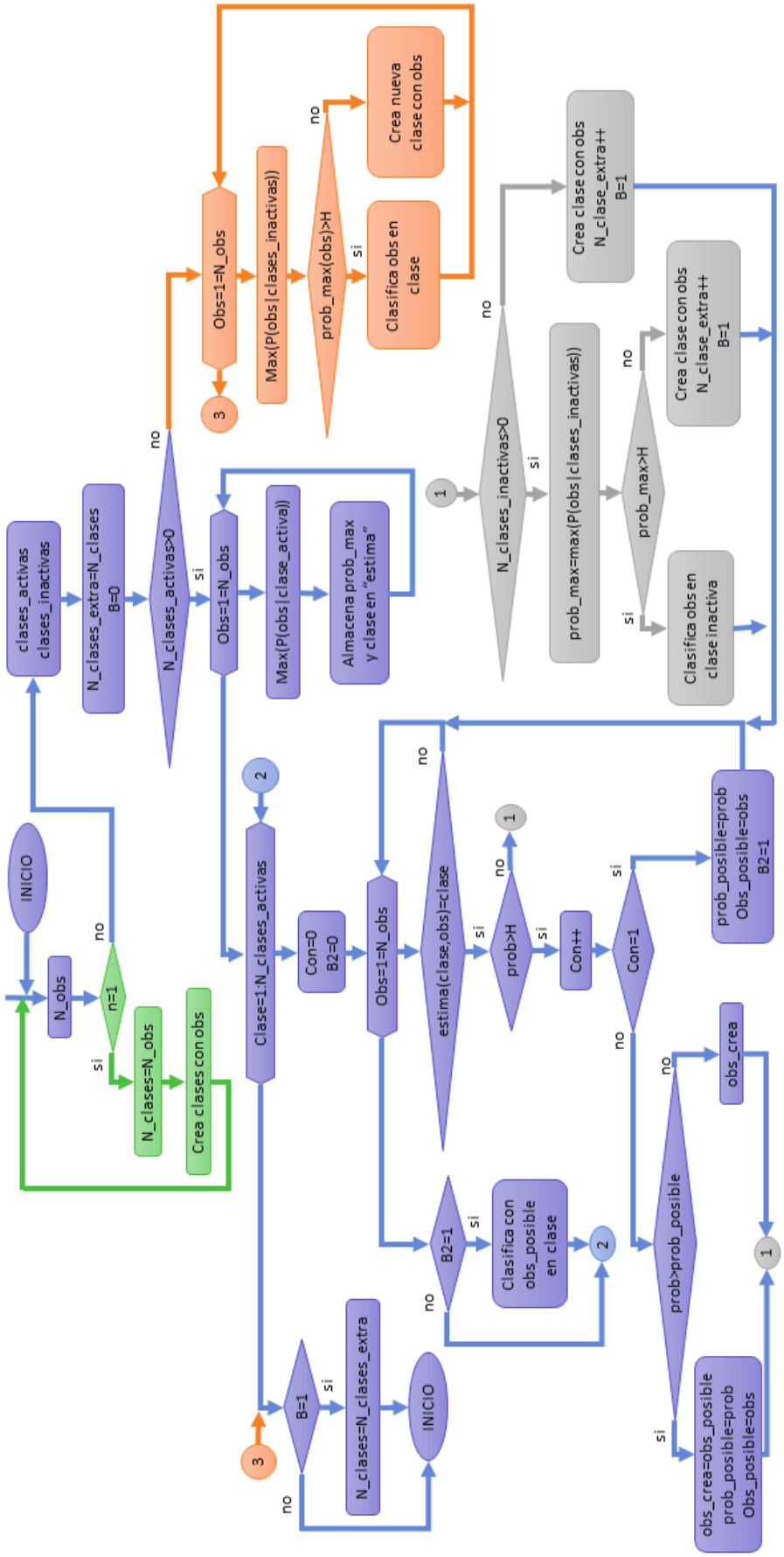
# Apéndice C.

# Diagrama de flujo del clasificador implementado

La clasificación de las observaciones con las fuentes de voz seguidas, se calcula por medio de la máxima verosimilitud tanto con el modelo Gaussiano de la fuente de voz seguida en el tiempo $t$, como con su respectivo modelo Gaussiano de las observaciones agrupadas con dicha fuente de voz, si la suma de ambas es mayor al umbral $H$ entonces se dice que la observación pertenece a la fuente de voz seguida.

El clasificador da prioridad para asignar la observación a las fuentes activas, si dicha observación no pertenece a ninguna de las clases activas, entonces verifica con las clases inactivas previamente definidas, de tal manera que si la observación pertenece a alguna fuente inactiva, ésta es clasificada con dicha fuente y pasa a ser una fuente activa. Si el observación no corresponde a ninguna de las clases tanto activas como inactivas, se crea una nueva clasificación (fuente de voz seguida).

A continuación se muestra un diagrama de flujo que describe la implementación del clasificador utilizado en el presente trabajo.


INICIO
N_obs
n=1
si
no
N_clases=N_obs
Crea clases con obs
clases_activas
clases_inactivas
N_clases_extra=N_clases
B=0
N_clases_activas>0
si
no
Obs=1=N_obs
Max(P(obs|clase_activa))
Almacena prob_max
y clase en "estima"
Obs=1=N_obs
Max(P(obs|clases_inactivas))
prob_max(obs)>H
si
no
Crea nueva
clase con obs
Clasifica obs en
clase
3
1
N_clases_inactivas>0
si
no
prob_max=max(P(obs|clases_inactivas))
prob_max>H
si
no
Clasifica obs en
clase inactiva
Crea clase con obs
N_clase_extra++
B=1
Crea clase con obs
N_clase_extra++
B=1
2
Clase=1:N_clases_activas
Con=0
B2=0
Obs=1=N_obs
estima(clase,obs)=clase
si
no
prob>H
si
no
1
Con++
Con=1
si
no
prob_posible=prob
Obs_posible=obs
B2=1
B2=1
si
no
Clasifica con
obs_posible
en clase
2
prob>prob_posible
si
no
obs_crea
obs_crea=obs_posible
prob_posible=prob
Obs_posible=obs
1
B=1
si
no
N_clases=N_clases_extra
INICIO
3

# Glosario

**aliasing espacial** En el área de procesamiento digital de señales el aliasing espacial se produce cuando la tasa de muestro es lo suficientemente pequeña de tal manera que no se puede representar la forma de onda de la señal maestreada, por lo que que la frecuencia de muestreo mínima debe de ser al menos el doble del ancho de banda de la señal adquirida. Sin embargo, en el presente trabajo también se le llama a aliasing espacial cuando la distancia entre los micrófonos del arreglo es lo suficientemente grande para producir lóbulos gratinados tanto en el patrón de radiación como en la respuesta de dirección.

**cámara anecoica** Recinto acústico diseñado para absorber las reflexiones producidas por ondas acústicas o electromagnéticas en cualquiera de las superficies que la conforman, además, dicha cámara se encuentra aislada del exterior de cualquier fuente de ruido o influencia sonora externa.

**direccionamiento del haz** Orientación del lóbulo principal del patrón de radiación con el objetivo de filtrar únicamente en dicha orientación.

**distribución Gaussiana** Proporciona la incertidumbre de una variable aleatoria x en forma de una campana Gaussiana, dicha distribución describe el comportamiento de muchos fenómenos que acontecen en el mundo.

**eigendescomposición** Proceso que se efectúa a partir de una matriz cuadrada con el objetivo de obtener los valores propios (eigenvalues) y vectores propios (eigenvectores) de dicha matriz.

**espectro MUSIC** Respuesta de dirección construida por el método MUSIC a través de la ortogonalidad entre los eigenvectores ruidosos y las direcciones de arribo.

**filtrado espacial** Adquisición de la señal que proviene en una sola dirección mientras que discrimina las señales provenientes del resto de las direcciones de arribo.

**lóbulo principal** También conocido como haz, contiene la mayor concentración de la energía orientada a una dirección deseada en el patrón de radiación y la dirección de arribo de las señales de la fuente en la respuesta de dirección.

**lóbulos gratinados** Replicas del lóbulo principal que aparecen tanto en el patrón de radiación como en la respuesta de dirección (para el formador de haz y el método MUSIC) cuando la distancia entre micrófonos es muy grande.

**lóbulos laterales** En un patrón de radiación, son los intervalos angulares en el que se encuentra un valor máximo relativo de potencia, dichos lóbulos se encuentran a un costado del lóbulo principal y son no deseados para el filtrado espacial .

**modelo de campo lejano** Región en el espacio donde el frente de onda de las señales se consideran como frentes de onda planos.

**máxima verosimilitud** Máxima probabilidad de una observación dado un modelo establecido.

**patrón de radiación** Forma en la que se distribuye la energía ya sea recibida o emitida por un dispositivo, en un intervalo de dirección establecido por el arreglo de sensores utilizado.

**rarefacción** Proceso por el que un cuerpo o sustancia se hace menos denso y se contrapone al fenómeno de compresión.

**respuesta de dirección** Representación gráfica que relaciona la dirección de arribo DOA con la energía, perpendicularidad o correlación, depende del método implementado.

**reverberación** Copia de la señal original con menor energía y desfasada que se produce por la reflexión de dicha señal en las pareces del recinto.

**teorema del límite central** La distribución de probabilidad de la suma de un conjunto de variables aleatorias independientes e idénticamente distribuidas, se aproxima a una distribución Gaussiana a medida que el número de variables aumenta .

**transición de estados** Cambio de estado producido por un evento.

**verosimilitud** Probabilidad de una observación dados los parámetros de un modelo, tales como $\mu$ y $\sigma$.

**ángulo de arribo** Ángulo definido entre el eje de referencia(cero grados) y el vector normal al frente de onda de las señales de voz (Dirección relativa en la que arriban las señales producidas por una fuente posicionada en un lugar en el espacio).

# Acrónimos

**ADC** conversión analógica a digital.

**BF** formador de haz.

**DFT** transformada discreta de Fourier.

**DOA** dirección de arribo.

**DS** *Delay and Sum*.

**FIR** filtro de respuesta finita al impulso.

**GCC** correlación cruzada generalizada.

**GCC-PHAT** correlación cruzada generalizada con transformada de fase.

**GSC** cancelador de lóbulos laterales.

**IDFT** transformada discreta de Fourier inversa.

**IIR** filtro de respuesta infinita al impulso.

**LMS** algoritmo *least-mean-square*.

**MUSIC** Clasificación de múltiples señales.

**MVDR** respuesta sin distorsión de mínima varianza.

**PDS** Procesamiento digital de señales.

**TDOA** diferencia del tiempo de arribo.

**VAD** detector de actividad de voz.

# Lista de símbolos