UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

Viabilidad de Cancelación Generalizada de Lóbulos Laterales en los auxiliares auditivos

TESINA

Que para obtener el título de:

Ingeniero en Telecomunicaciones

P R E S E N T A

Dulce María Apolonio Apolonio

DIRECTOR DE TESINA

Dr. Caleb Antonio Rascón Estebané

Ciudad Universitaria, Cd. Mx., 2017

Terminar una obra vale más
que comenzarla: lo que cuenta
es la
perseverancia, y no la pretensión.

Eclesiastés 7:8

A mi madre.
Ella se viste de fortaleza y dignidad,
y mira esperanzada al porvenir.
Cuando habla lo hace con sabiduría,
cuando instruye lo hace con amor.
Se levantan sus hijos para felicitarla,
su marido para elogiarla:
“Muchas mujeres han demostrado su valor,
pero tú las superas a todas”
Proverbios 31: 25-26, 28-29

Agradecimientos

Gracias Dios por permitirme terminar este proyecto.

Gracias por las personas tan maravillosas que tengo a mi lado: mi mamá, esa mujer buena y valiente que me ama con todo su corazón. Papá, que a pesar de todo, siempre ha estado a mi lado. Mis hermanos, a los que amo infinitamente y a quienes no cambiaría por nada. Mi tía Lena, mi segunda mamá, a quien adoro con todo mi ser. Mis amigos, en ellos puedo confiar porque a pesar de no frecuentarnos tanto, siempre podemos contar el uno con el otro.

Gracias por darme la oportunidad de conocer al Dr, Caleb, él ha sido muy paciente conmigo y me ha enseñado que una de las claves del éxito es compartir el conocimiento con los demás.

Gracias por darme la oportunidad de estudiar en la mejor universidad del país, la UNAM.

Gracias Dios por estar siempre a mi lado y demostrarme, en cada instante de mi vida, el gran amor que me tienes:

El amor con que te amo es eterno.
Aunque las montañas cambien de lugar,
y se desmoronen los cerros, no cambiará mi amor por ti
-dice el Señor, que te ama-
Isaías 54:8,10

Índice

Índice de Tablas

Índice de Figuras

Resumen

En México, casi 7.2 millones de personas padecen algún tipo de discapacidad, de este grupo el 33.5% tiene discapacidad auditiva. Para mitigar los efectos de dicho padecimiento, los pacientes utilizan auxiliares auditivos.

El objetivo de esta tesina es valorar qué tan viable es utilizar el filtro espacial llamado Cancelación Generalizada de Lóbulos Laterales (Generalized Sidelobe Canceller, GSC por sus siglas en inglés) en los auxiliares auditivos. La implementación de GSC se realizó en el lenguaje de programación C y se utilizó un procesador de señales de audio en tiempo real llamado *Jack Audio Connection Kit.*

El desempeño del filtro espacial se midió en base a la Relación Señal Interferencia (SIR) que se obtuvo de las grabaciones obtenidas del programa realizado. Dichas grabaciones se adquirieron con ayuda del corpus AIRA (Interacciones Acústicas para Audición de Robots), el cual fue recolectado por el grupo GOLEM (grupo de robótica del departamento de ciencias de la computación del IIMAS). AIRA es una recopilación de grabaciones hechas en dos escenarios: la cámara anecóica del Laboratorio de Acústica y Vibraciones del CCADET-UNAM y en el laboratorio de estudiantes del Departamento de Ciencias de la Computación del IIMAS-UNAM. Las evaluaciones realizadas permitieron detectar fallas y puntos de mejora, los cuales se contemplan en un trabajo futuro.

Capítulo 1

Introducción

Oír es un sentido admirable que posee el ser humano. El sistema auditivo humano puede capturar los sonidos, distinguir unos de otros e identificar su dirección de procedencia de una manera asombrosa. Oír permite aprender a hablar, a comunicarse. Contar con un sistema auditivo sano es una fortuna, sin embargo no todas las personas gozan de buena salud auditiva. Padecer discapacidad auditiva puede hacer que los individuos no puedan integrarse a la sociedad sintiéndose solos y con baja autoestima.

La Organización Mundial de la Salud (OMS), en su portal de Internet, informa que más del 5% de la población mundial (360 millones de personas, de ellos 328 millones de adultos y 32 millones de niños) padece pérdida de audición. La mayoría de esas personas vive en países de ingresos bajos y medianos [1].

Por otra parte, en México, en el año 2014, según los resultados de la Encuesta Nacional de la Dinámica Demográfica, había cerca de 120 millones de personas. De ellos, casi 7.2 millones reportaron tener algún tipo de discapacidad, de este grupo el 33.5% tiene discapacidad auditiva [2].

La siguiente gráfica muestra en porcentajes los tipos de discapacidad que existen en México para el año 2014. Se observa que la deficiencia auditiva ocupa el cuarto lugar a nivel nacional, por lo tanto, es un problema que merece atención.

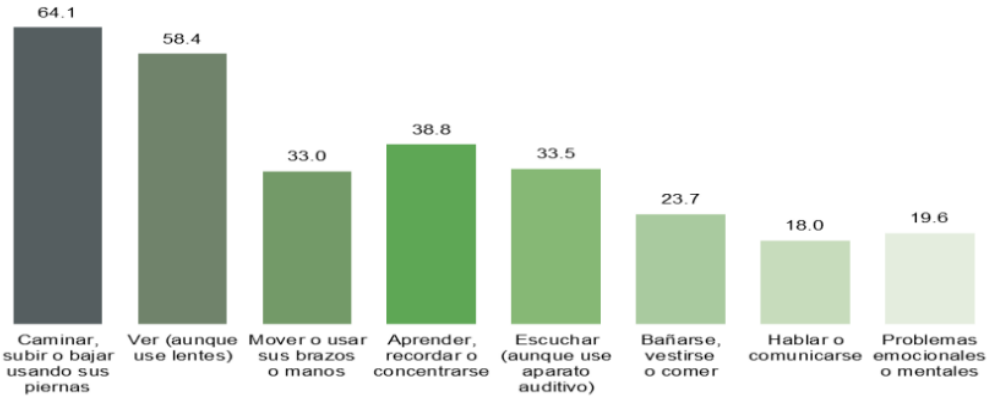

Figura 1.1 Porcentajes de la población con algún tipo de discapacidad [3].

Los datos de la gráfica suman más del 100 % debido a que una persona puede tener más de una discapacidad.

Existen distintos niveles de pérdida auditiva, ésta puede ser leve, moderada, grave o profunda. Afecta a uno o a ambos oídos. Las personas con pérdida auditiva leve, moderada o grave

pueden hacer uso de auxiliares auditivos.

En un principio, sólo existían en el mercado auxiliares analógicos, los cuales amplifican todas las señales de audio que capturan, incluyendo las interferencias. Esto, lejos de minimizar los efectos de la discapacidad auditiva, genera molestias en los usuarios forzándolos a dejar sus auxiliares en el cajón [4].

Para finales de los años 90 surge una nueva generación de auxiliares auditivos: los digitalmente programables. Esta generación ofrece hardware estético, cómodo y discreto. Otra de las ventajas que brinda esta tecnología es que los auxiliares se ajustan a las necesidades del usuario con ayuda de una computadora.

En el mercado también se comercializan auxiliares auditivos con tecnología digital la cual ofrece un software capaz de amplificar selectivamente el volumen de los sonidos que se desean escuchar. El software utiliza un filtro que discrimina las señales que causan interferencia de las señales deseadas, logrando que el usuario sea capaz de entender con mayor facilidad lo que escucha, aún en ambientes con demasiadas fuentes de interferencia. Sin embargo, los consumidores indican que aunque sus auxiliares cuentan con procesamiento digital de señales, esto no es suficiente cuando se encuentran en situaciones en dónde hay mucha reverberación.

Actualmente se han desarrollado técnicas de filtrado que permiten minimizar el ruido existente en el ambiente y amplificar la señal deseada. Una de esas técnicas es el Conformador de Haz (en inglés *Beamforming*). El Conformador de Haz tiene como objetivo amplificar la señal deseada y minimizar señales interferentes. Este sistema tiene sus inicios en el área de las telecomunicaciones, y consta de un arreglo de dos o más micrófonos acomodados a una cierta distancia conocida entre ellos. La dirección de arribo de la señal deseada debe ser conocida a-priori para llevar acabo el filtrado. Mientras mayor sea la cantidad de micrófonos el filtrado se llevará a cabo con mayor éxito.

Una técnica popular de *Beamforming* actualmente es la de Cancelación Generalizada de Lóbulos Laterales (GSC). Teóricamente, su mayor ventaja es que es adaptativo, lo cual le permite adecuarse en ambientes dinámicamente acústicos.

El objetivo de esta tesina es evaluar el desempeño de GSC en los auxiliares auditivos asumiendo que el usuario se encuentra frente a la fuente de sonido de su interés. El programa fue implementado en C en conjunto con un servidor que procesa señales de audio en tiempo real llamado *Jack Audio Connection Kit.*

Las pruebas del programa se realizaron con un corpus llamado AIRA el cual fue recolectado en dos ambientes diferentes: La cámara anecóica del Laboratorio de Acústica y Vibraciones del CCADET-UNAM y en el laboratorio de estudiantes del Departamento de Ciencias de la Computación del IIMAS-UNAM.

Para determinar la viabilidad de GSC en los auxiliares auditivos se tomó como referencia la relación señal interferencia (SIR) obtenida en cada prueba. El SIR se obtuvo gracias a un código implementado en Octave llamado ***bss_eval_sources***, el cual hace referencia a la Separación Ciega de Fuentes de Audio.

Capítulo 2

Marco Teórico

2.1 Auxiliares auditivos

La discapacidad auditiva ha existido desde tiempos remotos. Para mitigar los efectos de este padecimiento, en la antigüedad se utilizaban cuernos secos y huecos. Las trompetillas también fueron populares, los materiales que se empleaban para su construcción eran hojas de hierro, plata, madera, caparazón de caracol o cuernos de animales.

Durante el siglo XX la evolución de los audífonos tuvo aciertos significativos: la miniaturización de todos sus elementos, el aumento en la ampliación del sonido y el bajo consumo de energía.

En 1902 comenzó la comercialización del "Acousticón", un aparato que podía introducirse debajo de la ropa o en el bolsillo. Desgraciadamente, su tamaño no era práctico, ya que constaba de tres partes: el transmisor, el amplificador y el lugar que alojaba la pila. Diez años después, en 1912, apareció el primer control de volumen para las prótesis.

Los circuitos impresos representaron un enorme avance en cuanto al diseño de los auxiliares auditivos. Gracias a su potencial miniaturización, se pudieron eliminar las soldaduras y cableados que, hasta entonces, requerían de un mayor espacio físico donde alojarse.

Alrededor de 1948 nació el transistor. Este dispositivo fue la base para los futuros auxiliares analógicos y para los digitalmente programables. Gracias a esto, apareció en 1953 el primer auxiliar de bolsillo que utilizaba solamente transistores para amplificar el sonido.

Entre los años 1952 y 1987 se agregaron controles y funciones que mejoraron la respuesta y el rendimiento del auxiliar en algunas situaciones.

En 1982 se conoce la primera patente de un aparato auditivo eléctrico en Estados Unidos creada por T. Edison, E. Berliner y H. Hunnings.

Entre 1985 y 1990 se desarrollaron los primeros auxiliares que incorporaban tecnología digital. Sin embargo, éstos no eran auxiliares realmente digitales: sólo utilizaban esa tecnología para aumentar sus posibilidades de calibración. Dichos dispositivos son los que se conocen como "auxiliares digitalmente programables" y podían almacenar varias calibraciones. El usuario podía seleccionarlas según la situación sonora en la que se encontrara, ya sea mediante un control remoto o por medio de una llave selectora.

El más prometedor adelanto tecnológico es el auxiliar digital. En 1995 tuvo lugar el lanzamiento de los primeros productos comerciales con características de procesamiento digital [5].

2.1.1 Descripción física y funcionamiento de un auxiliar auditivo

Los auxiliares auditivos tienen como objetivo incrementar selectivamente el volumen de los sonidos que se desean escuchar. Tienen la capacidad de hacer los sonidos suaves audibles, mientras que al mismo tiempo hacen cómodos los sonidos moderados o fuertes, proporcionando alivio en ambientes ruidosos así como tranquilos [6].

La Figura 2.1 muestra cada una de las partes de un auxiliar auditivo y su ubicación.

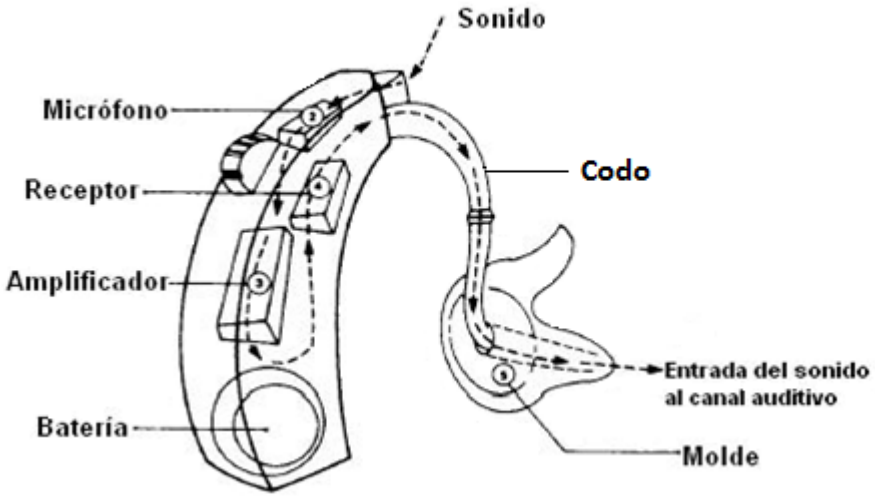

Figura 2.1 Partes de un auxiliar auditivo [9].

A continuación, se proporciona una descripción de cada una de las partes de un auxiliar auditivo:

- **Micrófono**. Captura los sonidos del ambiente y los convierte en señales eléctricas.
- **Amplificador.** Aumenta la intensidad de las señales que provienen del micrófono. El amplificador requiere el uso de técnicas especiales de filtrado para el procesamiento de

las señales, dichas técnicas reducen el ruido de fondo y amplifican la señal de interés. En este trabajo de tesina, esta es la parte del auxiliar que se pretende mejorar.

- **Receptor.** Convierte las señales eléctricas a señales acústicas para ser escuchadas.

- **Codo.** Se encarga de enviar el sonido amplificado a través de una manguera de plástico hasta el molde. Dicha manguera debe reunir ciertas características de acuerdo a la potencia de amplificación del auxiliar auditivo. Si la pérdida auditiva es profunda, las paredes de la manguera deben de ser gruesas para que el auxiliar no“pite”. Por otro lado, si el auxiliar es de mucha potencia y se requiere mayor amplificación, es preferible tener una pared gruesa para evitar que el sonido salga por sus poros (entre más delgada, mas porosa). Otra función que tiene el codo es sostener el auxiliar alrededor de la oreja.
- **Molde.** Es la parte del auxiliar que se inserta dentro del canal auditivo de la oreja. Es una parte fundamental dentro de todo el proceso de amplificación ya que permite que el auxiliar entregue adecuadamente el sonido que amplifica.
- **Batería.** Las baterías son especiales para cada tipo de auxiliar. Su duración depende del tiempo que el usuario utilice el auxiliar, el tipo de pila que se emplee y el tipo de audífono con el que se cuente.
- **Dispositivo de encendido.** Tiene la tarea de prender y apagar el auxiliar. En algunos modelos de auxiliares el compartimiento de la pila es adicionalmente el interruptor de encendido y apagado.

- **Control de volumen.** El control de volumen permite al usuario ajustar manualmente la amplificación del sonido.

2.1.2 Tipos de auxiliares auditivos

Los auxiliares auditivos se clasifican de acuerdo a la parte del oído en la que son colocadas. También se pueden catalogar conforme a la tecnología con que operan. En este trabajo se describirá ésta segunda clasificación.

Analógicos

Los auxiliares analógicos fueron los primeros que se ofrecieron comercialmente. Funcionan amplificando las señales que reciben de forma general, es decir, no discriminan una señal objetivo de entre señales que causan interferencia. Desafortunadamente, esto causaba molestias a los consumidores. Los ajustes electro-acústicos se llevan a cabo manualmente dependiendo de las necesidades del usuario.

Funcionamiento de un auxiliar analógico.

1. Se convierte la señal acústica a una eléctrica por medio de transductores de entrada, como un micrófono y una bobina de inducción.
2. Se amplifica la señal.
3. Se convierte la señal eléctrica amplificada a una señal acústica.

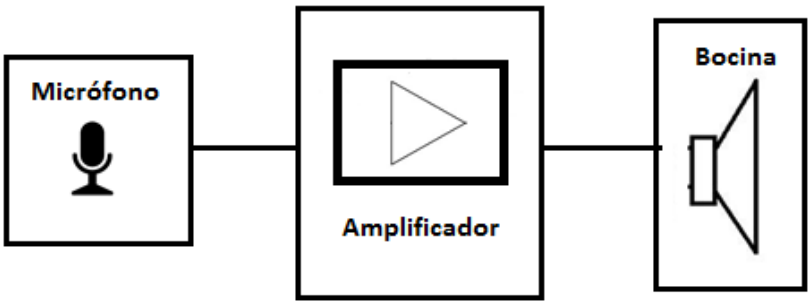

Figura 2.2 Diagrama de bloques de un auxiliar auditivo analógico.

Digitales

Trabajan con procesamiento digital de señales, lo cual les permite ser más selectivos, amplificando la señal deseada y minimizando el ruido de fondo así como de interferencias. Ofrecen grandes beneficios con respecto a los auxiliares analógicos ya que proveen una alta calidad de sonido, menor distorsión y mayor comprensión del habla en ambientes con ruido. Otra ventaja es que pueden ser ajustados por medio de un programa que se corre en una computadora externa, otorgando una mayor flexibilidad que los auxiliares analógicos. Una de sus desventajas es que no permite amplificaciones muy grandes, por lo que la adaptación se reduce a pérdidas entre débiles y moderadas.

Funcionamiento de un auxiliar digital.

1. Se convierte la señal acústica a una eléctrica por medio de transductores de entrada, como un micrófono y una bobina de inducción.

2. Se convierte la señal eléctrica (analógica) a una digital.
3. Se procesa la señal digital con los ajustes programados externamente.
4. Se convierte la señal digital a señal eléctrica (analógica).
5. Se convierte la señal eléctrica a una señal acústica.

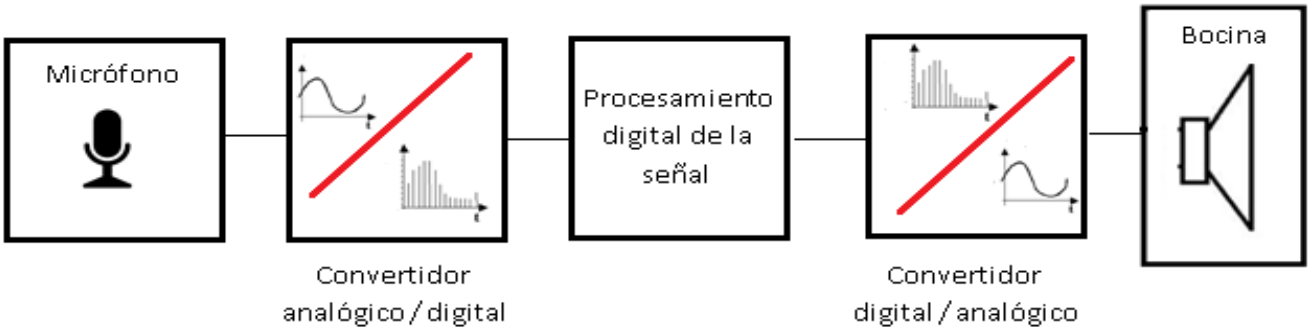

Figura 2.3 Diagrama de bloques de un auxiliar auditivo digital.

Digitalmente Programables
Es importante mencionar que este tipo de auxiliares son analógicos, es decir, no existe en su funcionamiento un proceso de conversión analógico-digital o viceversa. La parte "digital" de su nombre hace referencia a que se pueden ajustar y calibrar, por medio de una computadora externa, ciertos parámetros de operación (descritos más adelante) en su circuito de control interno. Y, ya que son analógicos, permiten amplificaciones grandes.

Una de las alternativas de software que se utilizan para ajustar los auxiliares a las necesidades del usuario es "*EasyFit*". Este programa permite adaptar las características del auxiliar auditivo de acuerdo a los requerimientos del paciente y seleccionar un buen modelo rápidamente. Los ajustes prescritos calculados en *EasyFit* se basan en métodos estándar para audífonos lineales, como [10]:

- **Control A-GRAM:** disminuye las bajas frecuencias o las altas frecuencias dependiendo de la curva global del audiograma.
- **Ganancia:** ajusta la ganancia pre-ajustada del auxiliar. En teoría, esta ganancia debería ser ajustada a un nivel confortable cuando el control de volumen esté alrededor de 3 [dB]. De esta manera se dejan alrededor de 10 [dB] de ganancia de reserva.
- **UCL:** ajusta la salida máxima (MPO) según el umbral de dolor (UCL) del usuario.

- **Método MPO**, proporciona tres opciones de limitación de la salida:
    - PC: *Peak clipping*, para los casos en que se necesita la mayor salida posible, por ejemplo pérdidas auditivas severas.
    - AGC Rápido: AGC de salida de acción rápida, funciona como la combinación tradicional de AGC + PC, proporcionando una limitación con baja distorsión.
    - AGC Lento: Combinación de AGC de salida + PC de actuación más lenta, proporcionando una limitación confortable de baja distorsión con un mínimo efecto de bombeo. Preferiblemente para pérdidas auditivas leves y moderadas.
- **Control de volumen:** permite ajustes del control de volumen del audífono desde la pantalla durante la adaptación. El ajuste manual del control de volumen anula el ajuste realizado desde la pantalla.
- **Control de Feedback:** debe activarse para eliminar el *feedback* si fuera necesario.
- **Opciones de la Bobina Telefónica:** permite cualquier combinación M-MT-T.
- **Nivel de la señal de aviso:** Una señal de aviso se activa cuando el usuario pulsa el botón para activar la bobina telefónica. Esta señal se puede graduar en 7 intensidades distintas (indicadas por niveles de entrada equivalentes) o se puede suprimir.
- **Incremento de ganancia al activar la bobina telefónica:** se puede incrementar en 6 [dB] la ganancia a través de la bobina telefónica con respecto a la ganancia a través del micrófono.

Discusión sobre la razón costo-beneficio de los tipos de auxiliares

Cada tecnología implementada en los auxiliares auditivos ofrece distintas ventajas.

Los auxiliares análogos son baratos pero no son selectivos: amplifican de igual manera la señal deseada que la del ruido.

Por otro lado, los que se utilizan por la mayoría de los consumidores son los digitales y los digitalmente programables. Ofrecen al usuario un mejor sentido del oído ya que son más fáciles de ajustar, son más discretos y estéticos. Por el lado de los auxiliares digitales, los consumidores que han emigrado a esta tecnología han notado cambios favorables, por ejemplo escuchan mejor en conversaciones uno a uno[11]. Pero, estos dos tipos de auxiliares son caros, no solo por lo que va de la inversión inicial, sino también por el costo de su consecuente mantenimiento y reparación.

2.1.3 Problemas conocidos

Para el año 2015 la Organización Mundial de la Salud indicó que 360 millones de personas en todo el mundo padecen discapacidad auditiva. Los problemas de audición que padecen estas personas pueden minimizarse usando auxiliares auditivos o implantes cocleares, sin embargo, la producción actual de audífonos satisface menos del 10% de las necesidades mundiales [1].

Como se mencionó en el apartado anterior, existen distintas tecnologías implementadas en los auxiliares auditivos, algunos estudios realizados han dado a conocer el nivel de satisfacción de los usuarios de auxiliares auditivos respecto a su funcionamiento. En el año 2005 se publicó un artículo en *The Hearing Journal* con el título "*Satisfacción del cliente con los auxiliares auditivos en la era digital",* escrito por el Dr. Sergei Kochkin [11]. La publicación tuvo el objetivo de valorar la satisfacción del usuario de auxiliares auditivos.

El estudio realizado valoró la eficiencia de los auxiliares que trabajaban con procesamiento digital de señales. En dicho escrito se menciona que casi el 90% de los consumidores de auxiliares auditivos utilizaban tecnología digital y que notaban diferentes ventajas con respecto a la tecnología analógica. Entre las ventajas que se encontraron se hizo mención de las siguientes:

- El procesamiento de la señal de audio es superior, en ambientes con ruido es más fácil comprender la señal de voz.
- Mayor comodidad del usuario en situaciones ruidosas.
- El funcionamiento del auxiliar se adapta mejor a las necesidades de cada consumidor.
- Se reduce la retroalimentación acústica y mecánica.
- Se reduce el número de micrófonos en los dispositivos unidireccionales.

Para conocer el sentir de los usuarios se consultó a más de 1500 usuarios de auxiliares auditivos, entre ellos personas que nos los utilizaban pero que estaban conscientes de que tenían algún problema con su audición. Seis de cada diez fueron hombres. En las encuestas realizadas se preguntó qué tan satisfechos estaban con respecto a las características de los auxiliares, el rendimiento de los mismos, el servicio que les brindaban los proveedores y qué tan bien escuchaban en 15 situaciones distintas. Los resultados arrojaron que el 84% de las personas encuestadas tenían pérdida en ambos oídos, de ese porcentaje el 52% señaló que su pérdida era moderada. En cuanto al mejor escenario en el que escuchan mejor el 48% dijo que

escuchar un grito en una habitación era la mejor situación. Por otra parte, el 60% dijo que se tenía dificultad para oír en situaciones con ruido de fondo. En cuanto a las características del producto, el 74% de los usuarios dijo que utilizaban audífonos en ambos oídos. El 57% de los usuarios contaba con tecnología programable mientras que el 47% con tecnología digital. Y el 69% de los usuarios indicaron que sus auxiliares auditivos contaban con control de volumen.

Para calificar el desempeño de los auxiliares auditivos se tomaron en cuenta los porcentajes de diez factores que los usuarios consideraron claves para decir si el auxiliar era bueno o malo. Dichos factores se mencionan en la Tabla 2.1:

Factores evaluados	Porcentaje de personas satisfechas
Satisfacción total del usuario	74 %
Claridad del sonido	72 %
Costos (rendimiento del audífono en relación con los precios)	69 %
Fiabilidad del audífono	69 %
Sonido audible naturalmente	68 %
Capacidad para escuchar en grupos pequeños	66 %
Fidelidad del sonido	64 %
Conversaciones uno a uno	63 %
Actividades de ocio	63 %
Escuchar TV	62 %

Tabla 2. 1 Satisfacción de los usuarios medida en porcentajes [11].

En la Tabla 2.1 se puede observar que en cuanto a la claridad del sonido, la fidelidad y la naturalidad, tres de cada cuatro consumidores estaban satisfechos con la primera y segunda característica y siete de cada diez creían que el auxiliar les proporcionaba un sonido natural.

Otro dato que es muy importante es que dos de cada tres estaban contentos con la capacidad de detectar la dirección de los sonidos y la facilidad para escuchar sonidos suaves. Sin embargo, sólo seis de cada diez se sentían cómodos con los sonidos fuertes, mientras que uno de cada cuatro no estaba conforme; poco más de la mitad estaban satisfechos con la capacidad de los auxiliares para eliminar la retroalimentación, en cuanto a situaciones ruidosas sólo la mitad estaba conforme.

En lo que se refiere a las situaciones auditivas más difíciles se encuentran los grupos grandes: sólo el 63% estaba satisfecho. En lugares como la escuela o el salón de clases nada más el 59% estaba cómodo. De una manera similar, durante el uso de un teléfono celular, sólo el 59% mostró satisfacción.

Algunas personas encuestadas dijeron que no sabían si sus auxiliares auditivos contaban con procesamiento digital de señales y otras tantas dijeron que sus auxiliares no tenían tecnología digital, dando un total del 53%, el restante 47% aseguró que sus auxiliares auditivos procesaban las señales digitalmente. El hecho de que trabajaran con procesamiento digital parece otorgar una plusvalía con respecto a la tecnología analógica. Los usuarios con pérdida de moderada a severa que habían emigrado de tecnología analógica a digital dijeron que: escuchaban mejor los sonidos fuertes, la fidelidad del sonido era mejor, la retroalimentación era suprimida de una mejor manera, escuchaban mejor en situaciones con ruido, y en general estaban más satisfechos que con sus anteriores auxiliares analógicos. Por otro lado, las personas con pérdida auditiva moderada no percibieron tantas diferencias entre la tecnología digital y la analógica.

Cinco años después, en enero del 2010, Kochkin escribió otro artículo *"Consumer satisfaction with hearing aids is slowly increasing"* [12]. En esta publicación se explican los estudios que se hicieron sobre la satisfacción de los usuarios respecto a los auxiliares auditivos. Los aspectos que se evaluaron fueron los siguientes: las características físicas del auxiliar, el rendimiento del auxiliar y la satisfacción en 19 escenarios de la vida diaria. Los datos que se reportan en el artículo son evaluaciones de usuarios con pérdida auditiva leve, moderada, severa y profunda. Se encuestó a 3174 personas, seis de cada diez eran hombres.

El artículo menciona que cerca del 100% de los auxiliares auditivos evaluados eran programables y digitales. El 55% de usuarios de auxiliares auditivos digitales informó que está satisfecho con el funcionamiento de dicho auxiliar.

Se evaluaron las siguientes particularidades:

- Beneficio general que ha obtenido el usuario
- Claridad del sonido
- Valor (rendimiento del audífono con relación al precio)
- Sonido natural
- Confiabilidad del auxiliar
- Riqueza o fidelidad del sonido
- Uso en situaciones ruidosas
- Capacidad para escuchar en grupos pequeños
- Confort con sonidos fuertes
- Sonido de la voz

Como se puede ver en el listado anterior, la evaluación recae en la capacidad del auxiliar para llevar a cabo el procesamiento de las señales de audio y la calidad del sonido que entrega. Respecto a este punto, según el artículo, tres de cada cuatro consumidores están satisfechos con la claridad del tono y el sonido de sus auxiliares, así como el sonido de su voz. Siete de cada diez están satisfechos con la direccionalidad, la naturalidad del sonido, el silbido y la retroalimentación, la capacidad de escuchar sonidos suaves y la fidelidad sonora. Dos de cada tres están satisfechos con cómo escuchan los sonidos fuertes y el sonido de masticar y tragar. Por otra parte, seis de cada diez consumidores están satisfechos con el uso de su auxiliar en situaciones ruidosas y al escuchar el ruido del viento.

En cuanto a situaciones reales, los escenarios evaluados fueron los siguientes: conversaciones uno a uno, grupos pequeños, viendo televisión, actividades al aire libre, uso del teléfono y del celular, actividades al aire libre, ir de compras y mientras viajan en automóvil.

El 91% de los consumidores encuestados respondió que estaban satisfechos con la capacidad de sus audífonos para mejorar la comunicación en conversaciones uno a uno. Tres de cada cuatro están satisfechos cuando conviven en grupos pequeños. El 80% dijo estar satisfecho mientras miran televisión. En cuanto a actividades al aire libre el 78% de los usuarios dijeron estar cómodos. El 77% de los usuarios encuestados dijo estar satisfecho con el funcionamiento de sus auxiliares auditivos al realizar compras y mientras viajan en automóvil. Cuando hablan por teléfono el 73% se sentía cómodo con el comportamiento de los auxiliares. Siete de cada diez reportaron satisfacción en conciertos y películas. En cuanto al uso del teléfono celular el 69% mostró satisfacción. Los porcentajes más bajos fueron reportados para eventos deportivos con el 66%, en el lugar de trabajo con un 65% y en la escuela con un porcentaje de 59%.

Los resultados de este estudio muestran que los usuarios de auxiliares auditivos no están del todo satisfechos, pero los auxiliares que usan les permiten llevar una vida normal. El artículo resalta que el grado de pérdida auditiva está ligado al grado de comodidad del usuario: cuanto más grave es la pérdida auditiva, mayor es la dificultad de satisfacer las necesidades del consumidor en los diferentes escenarios en los que se desenvuelve en su día a día debido a las múltiples fuentes de audio que están presentes.

El hecho de que haya distintas fuentes de sonido combinadas con la fuente de interés del

usuario complica el funcionamiento de los auxiliares auditivos. La problemática anterior fue expuesta por el Dr. Dennis Van Vliet, en el año 2012, en el artículo *"When Technology Bites Back"* [13]. Dos de sus pacientes (uno de 11 años y otro de 78 años) que utilizaban auxiliares auditivos se quejaron de la eficiencia de las audio-prótesis porque se sintieron incómodos cuando estaban en un escenario en el cual oían los distintos sonidos que genera un automóvil. Cuando intentaron ajustar el volumen de sus auxiliares terminaron por eliminar el molesto ruido y por escuchar nada de lo que les interesaba. Además, la direccionalidad del micrófono no funcionó del todo bien ya que discriminó las señales que causaban interferencia de igual manera que lo hizo con la señal objetivo. Es posible mejorar esta característica porque el usuario, por lo general, siempre pone su atención a la fuente de sonido de su interés y trata de estar frente a ella. Por ejemplo, cuando una persona conversa con otra es común que se estén mirando cara a cara.

Los auxiliares auditivos han mejorado con el paso del tiempo logrando que el usuario se sienta cómodo en sus actividades cotidianas. Actualmente los auxiliares auditivos se venden a precios accesibles, son discretos y estéticos. Sin embargo, un problema que persiste es que los usuarios no se sienten del todo satisfechos en situaciones donde hay mucha reverberación y fuentes de interferencia, es en esas circunstancias donde la técnica de filtrado espacial puede ser utilizada.

2.2 Procesamiento de señales de audio.

Procesar una señal de audio ayuda a interpretar los datos que contiene. El procesamiento digital de señales permite modificar una señal o mejorarla de una manera flexible. Se caracteriza por representar los datos de la señal en un dominio discreto, por tal motivo, procesar la señal digitalmente es muy útil para representar señales analógicas en tiempo real.

Una de las herramientas que se utilizan en el procesamiento de la señal de audio es el filtrado. Esta técnica se ejecuta para limpiar la señal deseada de interferencias y de esta manera poder hacer un reconocimiento de voz, conocer la dirección de arribo de la señal o simplemente poder interpretar mejor los datos que contiene. Dicho filtrado normalmente se lleva a cabo para disminuir el nivel de ruido y así mejorar los datos capturados o combinarlos con el fin de crear nuevos datos útiles para la aplicación en cuestión

2.2.1 Señal de audio

El sonido es un fenómeno producido por ondas sonoras longitudinales generadas por el

movimiento vibratorio de un cuerpo, que se propagan por un medio elástico y que son captadas por un receptor. El sonido puede incluir información ondulando en diversas frecuencias, de las cuales el rango audible va de los 20 [Hz] a los 20 [kHz].

Las ondas sonoras están definidas por los siguientes parámetros (presentados en la Figura 2.4): longitud de onda (λ), frecuencia (f), velocidad del sonido (c), periodo (T), amplitud (*A*).

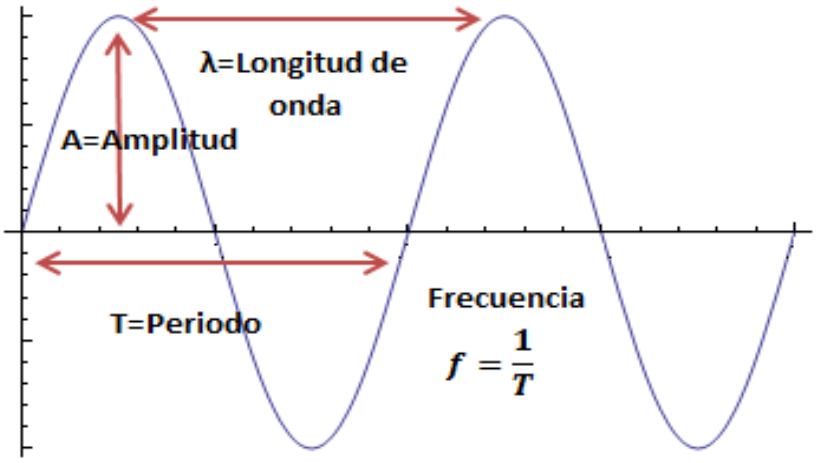

Figura 2.4 Parámetros de la señal sonora [16].

El período (T) es el tiempo necesario para que se produzca un ciclo completo de la onda sonora, su unidad es el segundo [s].

La frecuencia (f) corresponde al número de ciclos que se realizan por segundo (es la inversa del período), se miden en Hertz [Hz]:

$$f = \frac{1}{T} \qquad (2.1)$$

La velocidad (v) a la que se propaga la onda acústica en un medio elástico dependerá de las características de éste, su unidad son los metros sobre segundos [m/s].

La longitud de onda (λ) es la distancia que abarca una onda en un período completo, se mide en metros [m] y está relacionada con la velocidad del sonido (c), frecuencia y período:

$$\lambda = \frac{c}{f} \qquad (2.2)$$

En general los sonidos no son sinusoidales. Su amplitud y su frecuencia pueden variar con el tiempo debido a que están compuestos por más de una frecuencia. Se puede considerar que es una combinación de un gran número de sonidos sinusoidales superpuestos.

La magnitud de una onda se puede determinar por medio de la presión sonora la cual se refiere a las variaciones de presión producidas por una onda sonora en su propagación a través del

aire.

Las ondas sonoras pueden sufrir degradaciones debido a algunos fenómenos como la absorción, la cual hace referencia a la energía que pierden las ondas sonoras cuando, al propagarse, inciden en un material (la cantidad de energía absorbida depende del material), el resto de la energía es reflejada provocando lo que se conoce como un eco. El cúmulo de los ecos producidos por una señal y su reflejo en los materiales en el ambiente se conoce como reverberación, y es otro factor que contribuye a la degradación de la onda. Esto crea el fenómeno de que la energía de la onda persiste un determinado tiempo después que la fuente original ha dejado de emitirla. La reverberación, al ser capturada, se puede considerar como una adición acumulada de la misma señal a si misma varias veces, cada una sufriendo un desfasamiento y degradación de energía. Esto cambia significativamente la señal capturada originalmente.

La difracción es una propiedad de la onda sonora que le permite desalinear su trayectoria para propagarse a través de un obstáculo o rendija. Lo que determina qué tanto se puede difractar una onda es la dimensión del obstáculo o la abertura y la longitud de onda de la señal sonora.

Por otro lado, la interferencia también es un fenómeno que degrada la onda sonora, ésta se presenta cuando dos o más ondas sonoras se superponen. Esto puede originar una amplificación o atenuación de las mismas. Cuando dos señales de la misma frecuencia están totalmente desfasadas se anulan entre sí y esto genera una interferencia destructiva; por el contrario, si las señales están en fase, éstas se suman y se tiene una interferencia constructiva.

La manera más práctica de expresar los parámetros acústicos es con una relación logarítmica la cual permite obtener resultados manejables en la unidad llamada Bel. Esta unidad se define como el logaritmo en base 10 de la relación entre dos potencias sonoras o dos intensidades. Para evitar utilizar magnitudes demasiado grandes, normalmente se utiliza la décima parte del Bel, el decibelio (dB).

Debido a que la intensidad acústica es proporcional, en el campo lejano, al cuadrado de la presión sonora, se definió para las medidas sonoras una escala conveniente que está dada por el nivel de presión sonora.

$$L_p = 10\log\left(\frac{p}{p_0}\right)^2 = 20\log\left(\frac{p}{p_0}\right) \quad (2.3)$$

Donde p es la presión sonora medida, p_0 es la presión sonora de referencia (normalmente 20 [μPa]) y la palabra nivel se añade a presión sonora como una indicación de que el valor medido tiene un cierto nivel por encima de un valor de referencia definido previamente.

El uso de la escala logarítmica disminuye el margen dinámico de la presión sonora el cual va de 0 a 120. El 0 hace referencia al umbral mínimo y 120 representa el umbral de dolor [18].

2.1.2 Modelo de la señal de audio

Las señales de audio, por lo regular, forman mezclas convolutivas debido a que registran retardos y reflexiones al propagarse en ambientes reales en los cuales se irradian otras señales.

Se les llama mezclas convolutivas cuando se generan en un ambiente reverberante ya que las componentes de las observaciones (fuentes originales que se mezclan) no sólo almacenan la información de las fuentes sino también la de las reflexiones de dichas fuentes. Aunque se tenga una propagación lineal, también se generan múltiples trayectos con distintos tiempos de propagación.
Esta mezcla convolutiva se puede expresar de la siguiente manera:

$$x(n) = A * s(n) \qquad (2.4)$$

donde cada elemento de *A* es una función de transferencia desconocida

Las señales de audio, en la mayoría de los casos, son analizadas en el dominio de la frecuencia. Para cambiar las señales de audio al dominio de la frecuencia se utilizan diferentes cálculos matemáticos, por ejemplo la Transformada de Fourier. Sin embargo, dicho método no es del todo útil debido a que no contempla los cambios en frecuencia a lo largo del tiempo.

Una variante de la Transformada de Fourier es la llamada Transformada de Fourier de Tiempo Reducido (*"Short-time Fourier Transform", STFT*). Este método divide la señal en segmentos cortos, y a cada segmento le aplica la Transformada de Fourier, resultando en lo que se conoce como un espectrograma: una representación en donde un eje es el tiempo, otro es la frecuencia, y un tercero representa la energía de una frecuencia en un momento en el tiempo.

La STFT fracciona la señal en segmentos aplicando una función ventana, la cual es una envolvente que se utiliza para el análisis espectral. El rango de duración de la ventana va de 1 [ms] a 1 [s], cabe mencionar que los fragmentos se pueden traslapar. La expresión de la STFT para una señal discreta es la siguiente:

$$STFT\{s[n]\} = S(f, w) = \sum_{n=-N}^{N} s[n]\, w[n-m]\, e^{-j2\pi fn} \qquad (2.5)$$

Donde, *s[n] es la señal que se transformará* y *w[n]* es la ventana que se utiliza, ambas en el domino discreto. N representa la mitad del tamaño de la ventana w , y $S(f, w)$ es el valor (en frecuencia f) de $s[n]$ cuando $w[n]$ .le aplica una función ventana.

Una de las desventajas que presenta la STFT es que produce distorsiones en la medida del espectro, debido a que el analizador de espectro mide la señal de entrada y el producto de la misma por la ventana. Por lo tanto el espectro resultante es la convolución del espectro de la señal de entrada y el espectro de la ventana [22].

Elegir una ventana adecuada permite obtener de manera menos destructiva las propiedades espectrales de la señal de audio, evitando discontinuidades en el dominio del tiempo que se reflejan con valores en frecuencias no presentes en la señal original. Al remover estas discontinuidades, se hace posible que la Transformada de Fourier en Tiempo Reducido pueda

ser invertible. Las ventanas más utilizadas son la ventana Hamming y la ventana Blackman, por su habilidad de ser sobrelapadas con un desfase conocido y obtener una función cercana a ser unitaria

2.1.3 Procesamiento de la señal de audio

El audio digital es el resultado de dos procesos: el proceso de muestreo y el de cuantificación.

El muestreo puede interpretarse como la multiplicación de una señal continua en el tiempo por un tren de impulsos unitarios. En el muestreo existe un periodo de tiempo T llamado periodo de muestreo, este tiempo T se relaciona con la frecuencia de muestreo de la siguiente manera:

$$f_s = \frac{1}{T} \qquad (2.6)$$

Donde f_s representa la frecuencia de muestreo y T es el periodo de muestreo-

La frecuencia de muestreo se refiere a la velocidad con que se obtienen las muestras y nos señala la cantidad de muestras tomadas o registradas en un tiempo determinado. Mientras más alta sea la velocidad de muestreo, la señal digital se parecerá más a la señal original.

Para una señal analógica el valor de dicha señal en cada instante de tiempo es conocido, sin embargo, para la señal muestreada sólo se conocen los valores que toma en determinados puntos en el tiempo. Si se desea reconstruir la señal original a partir de la versión digital, se deben realizar interpolaciones calificadas de los valores entre los puntos de muestreo. Si los valores registrados en las interpolaciones son distintos a los de la señal original entonces la señal reconstruida será una señal distorsionada.

La situación anterior, se puede resolver utilizando el teorema de muestreo de Nyquist-Shannon, el cual indica que para poder reconstruir con exactitud la forma de una señal analógica es necesario que la frecuencia de muestreo (f_s) sea superior o igual al doble de la máxima frecuencia a muestrear (f_m).

$$f_s \geq 2 f_m \qquad (2.7)$$

El siguiente proceso que se lleva a cabo en el procesamiento de la señal de audio es el de la cuantificación. En la cuantificación lo que se hace es convertir una señal continua en amplitud en una señal discreta en amplitud.

Durante dicho proceso se mide el nivel de tensión de cada una de las muestras adquiridas en el proceso de muestreo, y se les asigna un valor discreto de amplitud, seleccionado por aproximación dentro de un margen de niveles previamente fijado.

Estos niveles, también llamados bandas, están espaciados equitativamente en un intervalo dado, dicho espaciamiento puede ser no lineal y esto generaría una distorsión armónica. Una vez que se determina a qué banda pertenece el nivel de la señal, el código correspondiente puede ser usado para representar ese nivel.

Las cantidades preestablecidas para ajustar la cuantificación se escogen de acuerdo a la resolución que utilice el código empleado durante la codificación. Si el nivel obtenido no concuerda exactamente con ninguno, se toma como valor el inferior más próximo.

La mayoría de los sistemas utilizan hoy el código binario, el número de intervalos de cuantificación, N , es:

$$N=2^n \qquad (2.8)$$

Donde *n* es la longitud de la palabra en código binario. Mientras se tengan más bandas, más larga es la longitud de la palabra, y esto brinda una mejor resolución. Esto a su vez, da paso a que la representación de la señal sea más exacta.

Hay muchas maneras de codificar la información digital en forma de señales eléctricas. Este proceso se llama modulación, el método más común es la modulación por impulsos codificados (Pulse Code Modulation, PCM). Existen dos maneras de llevar a cabo la modulación PCM, las cuales son el modo paralelo y en serie.

El modo paralelo codifica la información como niveles de voltaje en un número de cables llamados buses paralelos. Se utilizan señales binarias, lo que significa que sólo se utilizan dos niveles de voltaje, +5 [V] corresponden a un número binario 1 y 0 [V] que corresponde al número binario 0. Por lo tanto, cada cable que lleva 0 ó 5 aporta un dígito binario (bit). Un bus paralelo consistente de ocho hilos por lo tanto tiene 8 bits. Un bus paralelo es capaz de transferir altas tasas de transmisión.

El modo serial transfiere la información en un solo hilo. Los tiempos de transmisión son más largos, pero sólo se necesita un cable.

Hay otras formas de modulación como la Modulación por Amplitud de Pulsos (PAM), Modulación por Posición de Pulsos (PPM), entre otras.

2.3 Beamforming

El termino *Beamforming* surge gracias a los diseñadores de filtros espaciales, los cuales crearon estos filtros para formar rayos para transmitir una señal por una dirección específica. Por lo tanto *Beamforming* significa, en español, formador de haz.

Se puede definir *Beamforming* como un procesador de señales que, utilizado junto con un arreglo de sensores, provee una forma de filtrado espacial [26]. Por lo tanto, *Beamforming* es un sistema conformado por sensores que están acomodados con cierta geometría con el fin de obtener la información de una señal objetivo de manera fiel y sin interferencia.

Beamforming (formador de haz) tiene su origen en el área de las Telecomunicaciones, se empleó en los arreglos de antenas con el fin de aprovechar toda la energía que radian éstas. En dichos arreglos la potencia es transmitida hacia la antena objetivo pero también en otras direcciones. El hecho de "repartir" la energía por todas partes representa un problema: la antena objetivo no recibe todo y por lo tanto, no lo recibe bien. Para recibir toda la energía transmitida es necesario tener un "camino exclusivo" por el cual dicha energía pueda llegar sólo a la antena deseada.

2.3.1 Beamforming clásico: Delay and Sum

El formador de haz realiza un filtrado espacial que tiene por objetivo capturar la señal deseada lo más fiel posible y minimizar o eliminar las señales que causan interferencia, esto lo logra a partir del conocimiento de la dirección de procedencia de la señal de interés.

La forma en la que *Beamforming* funciona es por medio de un algoritmo que realiza la suma ponderada de la señal recibida por cada elemento dispuesto en el arreglo con una configuración espacial particular. La suma ponderada de la señales permite obtener una señal de interés en una dirección particular y atenúa el resto que se encuentran en direcciones distintas a las de interés [27].

En otras palabras, *Beamforming* requiere de un arreglo de sensores, los cuales son colocados con cierta geometría (circular, lineal, triangular). Cada elemento del arreglo recibe la señal objetivo y también las demás señales que causan interferencia. Si se conoce la dirección de la señal deseada, se calcula el desfase apropiado por cada micrófono del arreglo. Para este cálculo es importante mencionar que normalmente se utiliza el concepto de la onda plana, el cual hace referencia a que la onda de sonido normalmente se propaga de una manera esférica. Cuando la fuente que origina el sonido es suficientemente lejana del dispositivo que captura la señal, dicha onda se puede considerar como plana. Esta suposición reduce los modelos utilizados para describir la relación entre el desfase entre dos señales capturadas y la dirección de arribo de la fuente de interés a una función arcoseno.

Posteriormente se desfasa cada señal y se suma toda la información recolectada por cada componente del arreglo, y la suma se divide entre el número total de sensores dispuestos en el arreglo.

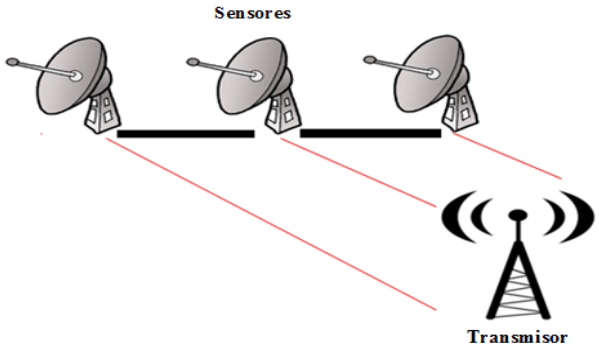

Figura 2.5 Arreglo de sensores que recolectan la señal emitida por un transmisor.

Como lo muestra la figura 2.5, cuando el transmisor emite la señal, ésta llega combinada con las otras señales existentes en el ambiente a los distintos sensores colocados en el arreglo. Debido a la distancia que existe entre cada sensor respecto al receptor, el tiempo de llegada a cada elemento del arreglo es distinto; para compensar dicho tiempo las señales son desplazadas mediante un retraso.

Una vez aplicado el retraso a las señales, el *Beamforming* suma las señales punto-a-punto, y divide el resultado entre el número de sensores existentes en el sistema con el fin de atenuar las señales que provienen de otras direcciones. Por tal motivo, mientras mayor sea el número de sensores mayor será dicha atenuación. Este es el modelo más simple de *Beamforming* y recibe el nombre de Retardo y Suma (*Delay and Sum*).

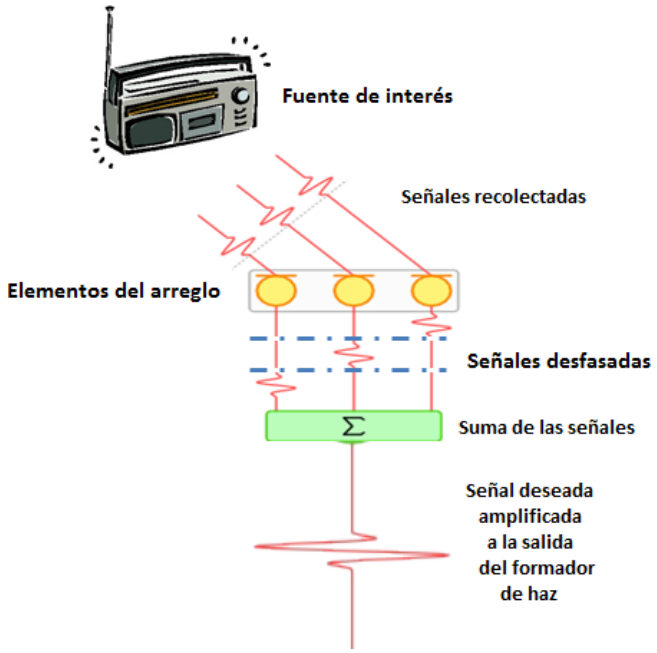

Figura 2.6 Formador de haz de suma y resta.

Matemáticamente, el formador de haz de suma y resta se define como:

$$Y_i(t) = \sum_{i=1}^{n} W_i(t) * S_i(t) \qquad (2.9)$$

En donde:

$Y_i(t) = señal\ resultante\ a\ la\ salida\ del\ arreglo$

$W_i(t) = es\ el\ filtro\ aplicado\ al\ receptor(i)$

$S_i(t) = salida\ del\ receptor$

$n = número\ de\ sensores\ en\ el\ arreglo$

Todas las salidas de los sensores que conforman el arreglo son multiplicadas por un coeficiente para después ser sumadas y producir la señal $Y_i(t)$ deseada. Para lograrlo, cada canal de los sensores es filtrado con un filtro $W_i(t)$.

Resumiendo la explicación del formador de haz de suma y resta, se puede decir que está compuesto por dos etapas. La primera etapa consiste en que, una vez que se conoce la dirección de llegada de la señal de interés, se aplica un desfase a cada señal recibida en cada sensor, éste depende del ángulo de llegada de la fuente de interés al arreglo. La segunda parte del *Beamforming* se trata de sumar las señales retrasadas para así obtener la señal deseada libre de interferencias.

Si se consideran señales como la voz, los coeficientes cambiarán con la frecuencia y esto lleva a considerar la convolución de la señal en el dominio del tiempo.

Se puede filtrar la señal en el dominio del tiempo y en el dominio de la frecuencia. El filtro temporal se vale de filtros FIR (Respuesta Finita al Impulso) los cuales combinan de manera lineal los datos obtenidos temporalmente. El filtrado frecuencial cambia directamente el espectro de la señal por lo tanto se requiere realizar la transformada de Fourier de la señal temporal como primer paso antes del filtrado.

Una de las desventajas que tiene el formador de haz de suma y resta es que requiere de muchos micrófonos para realizar un buen filtrado. Otro de los factores que intervienen con el buen funcionamiento del formador de haz es que si las señales de interferencia tienen una Relación Señal a Ruido mayor que la señal objetivo éstas prevalecerán y minimizarán dicha señal.

2.3.3 Técnicas avanzadas de Beamforming

Además del *Beamforming* convencional ya descrito, existen otros tipos de *Beamforming* que, por ejemplo, intentan minimizar la energía de las interferencias o llevan a cabo un proceso de adaptación en ambientes acústicos dinámicos.

En este apartado se describirán las siguientes técnicas avanzadas de *Beamforming*: Respuesta de Varianza Mínima sin Distorsión (MVDR), Mínima Varianza Limitada Linealmente (LCMV) y Cancelación Generalizada de Lóbulos Laterales (GSC).

El objetivo de atenuar o eliminar de una manera eficiente las señales intrusas se logra gracias a que el sistema ocupa unos atenuadores denominados “pesos” con los cuales el diagrama de radiación se puede enfocar mayormente a la señal de interés y/o se puede complementar con nulos en las direcciones de las interferencias. Adicionalmente, estos pesos se pueden actualizar en línea para adaptarse a cambios en el ambiente.

2.3.3.1 Respuesta sin distorsión con mínima varianza (MVDR)

MVDR es un *Beamforming* que pretende mejorar la relación señal-interferencia en comparación con el formador de haz convencional.

La deficiencia que tiene MDVR es que está diseñado para trabajar en un ambiente con condiciones ideales ya que es muy sensible a cualquier falla ocurrida en el entorno, en las fuentes o en el conjunto de sensores. Un pequeño cambio puede ocasionar un error en el resultado al identificar la dirección de arribo, desajuste en el espacio (más cerca, más lejos) y distorsión en la forma de onda.

MVDR debe calcular un *vector de dirección* que le indica la dirección de la señal objetivo, dicho vector es lo que lleva a MVDR a cometer errores ya que puede ser inexacto, causando degradación en la señal objetivo porque a ésta la considera una señal de interferencia.

Por otra parte, también tiene que elegir un vector de pesos, esto también representa un problema. El vector de pesos se define como:

$$w = \mu R^{-1} a(\theta_0) \qquad (2.10)$$

Donde:

$\mu = Constante\ de\ normalización$

$R = Matriz\ de\ covarianza$

$a(\theta_0) = Vector\ de\ dirección\ de\ la\ señal\ objetivo$

La matriz de covarianza R determina la posición cero que toma la salida del arreglo. R contiene la información de todas las señales incidentes, incluyendo la de la señal objetivo y las señales de interferencia, que no pueden ser separadas por una descomposición normal lo que provoca que en la salida se muestre un cero en la dirección de la señal deseada, anulándola. Por otra parte $a(\theta_0)$ es el vector de dirección de la señal objetivo, este vector es necesario porque ayuda a formar un lóbulo en la dirección de θ_0 para que proteja a la señal objetivo.

En la práctica, R es estimada por ventanas de la señal del arreglo y esto puede conducir a múltiples errores de estimación. Por lo tanto R recopilará información inexacta de la señal de interés y ésta será diferente a la información que contiene $a(\theta_0)$, por lo anterior, la información que proporciona $a(\theta_0)$ muestra un “camino distinto” al de R y la señal objetivo también es suprimida.

También se presentan errores cuando hay un desajuste entre el ángulo de llegada θ_0 de la señal objetivo y el ángulo θ_m (ángulo real de la señal objetivo). Se tiene la siguiente relación:

$$w^H a(\theta_0) = 1 \qquad (2.11)$$

Como θ_0 y θ_m no son iguales, si se sustituye a θ_m en la ecuación el resultado no está obligado a ser 1 debido al desajuste. Por lo tanto la señal de interés es degradada.

Gracias a estos errores en el vector de pesos, MVDR trata a la señal de interés como una señal de interferencia, por lo tanto dicha señal se atenúa.

Estos defectos de MVDR son muy notorios en un ambiente donde las señales de interferencia tienen una relación señal a ruido alta. La relación señal a ruido condiciona qué tanto se degrada el trabajo del conformador de haz. Es posible hacer un mejor *Beamforming* si se reduce la potencia de la señal de interés.

Se han desarrollado distintos métodos para solucionar el problema de MVDR, entre ellos se encuentra el método de la diagonal de carga cuyo principal objetivo es suprimir el ruido blanco en lugar de la interferencia, esto reduce el problema de cancelar también la señal de interés. Una desventaja de este método es que falla al eliminar interferencias grandes ya que pone más esfuerzo en suprimir el ruido blanco. Por lo tanto hay una compensación entre la reducción de la señal y la efectiva supresión de la interferencia.

El principal problema de MVDR es la falta de coincidencia entre el vector de dirección y la señal de interés en el modelo real de la señal contenido en la matriz de covarianza, especialmente en un escenario con una alta relación señal a ruido. Este problema se puede solucionar ajustando el vector de la señal objetivo al modelo de la señal contenida en la matriz de covarianza, eliminando los errores o también reduciendo la potencia de la señal de interés de la matriz de covarianza (disminuyendo el grado mal acondicionado).

2.3.3.2 Mínima varianza limitada linealmente (LCMV)

El formador de haz LCMV es un filtro espacial que permite poner múltiples restricciones con el fin de cancelar las direcciones de interferencias conocidas. Reduce la posibilidad de que la señal objetivo se suprima cuando llegue a un ángulo ligeramente diferente de la dirección deseada.

El objetivo de LCMV es restringir la respuesta del formador de haz con el fin de que la señal deseada pase con una ganancia y una fase específica. En este tipo de filtrado espacial también se tiene que elegir un vector de pesos, estos pesos minimizan la potencia de salida y están sujetos a restricciones en la respuesta. LCMV trata de mantener intacta la señal deseada mientras disminuye el efecto de las señales de interferencia en la salida del formador de haz.

LCMV también depende de la elección de pesos, los cuales minimizan la potencia a la salida del arreglo de sensores. Estos pesos están contenidos en un filtro FIR (Respuesta Finita al Impulso), el cual tiene una respuesta unitaria a señales con frecuencia w_0 . La limitación lineal de los filtros esta dada por:

$$W^H d(\theta, w) = g \qquad (2.12)$$

Donde g representa una constante compleja, d es cualquier señal de ángulo θ y frecuencia w que pasa a la salida con una respuesta W .

La elección de los pesos se puede llevar a cabo de la siguiente manera:

$$min_w W^H R_x W \quad \text{sujeto a} \quad C^H W = f \tag{2.13}$$

Donde la matriz C ($N \times L$) es llamada la matriz de restricción y el vector f de tamaño L se denomina vector de respuesta. Las restricciones son linealmente independientes por lo tanto C tiene un rango L . Cada restricción representa un grado de libertad menos para el filtro en cuanto a la atenuación de la interferencia.

2.3.3.3 Cancelación generalizada de lóbulos laterales (GSC)

GSC es una técnica de *Beamforming* derivada de LCMV y es un conformador de haz adaptativo. Éste modifica el patrón de direccionalidad con relación a las condiciones de ruido y reverberación presentes en el escenario en el que se ejecuta con el fin de minimizar la potencia de las señales que causan interferencia.

GSC se compone de dos etapas. En la primera etapa, la señal deseada y la interferencia son procesadas con el conformador de haz de suma y resta. El objetivo que se debe cumplir en esta primera etapa es separar la señal deseada de la interferencia, sin embargo, dicha señal sigue conservando un poco de ruido.

En la segunda etapa lo que se recolecta son las señales de interferencia. En esta parte se cuenta con una matriz de bloqueo, se llama así porque impide el paso de la señal de interés. El resultado de la matriz de bloqueo pasa a través de una serie de filtros, los cuales determinarán qué tanto ruido deben mandar a la salida, esto lo hacen en función del escenario en el que se están ejecutando; el resultado que arrojen los filtros será sustraído a la señal deseada (ya que tiene ruido) para que de esta manera salga libre de interferencias.

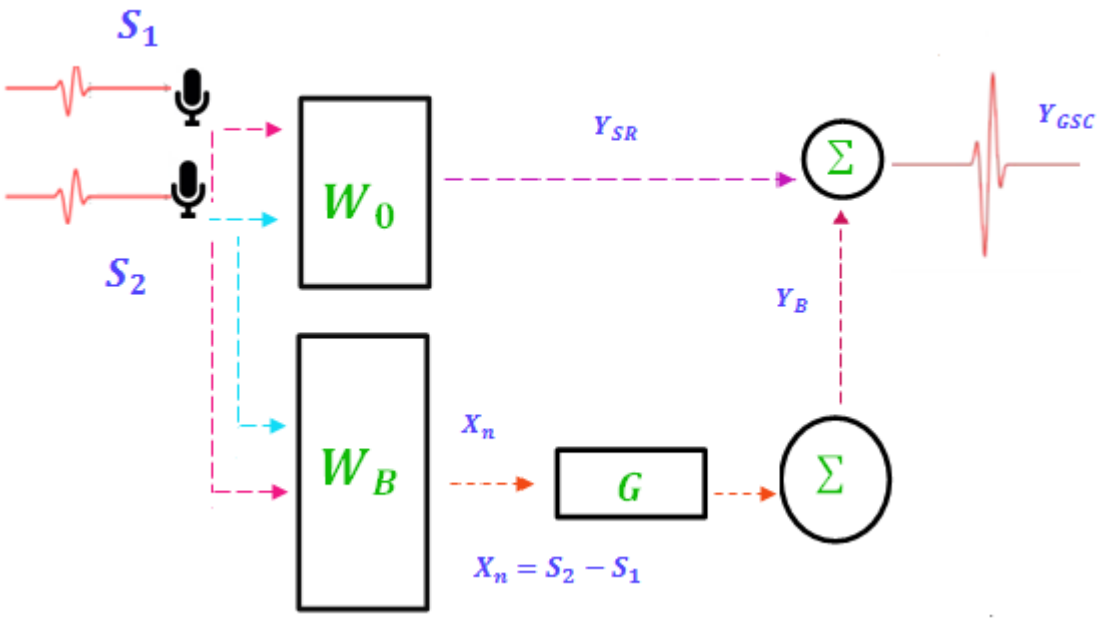

Figura 2.7 Esquema de GSC.

La Figura 2.7 muestra un diagrama de bloques de GSC. W_0 representa el conformador de

haz de suma y resta. W_B representa a la matriz de bloqueo la cual solo captura a las señales de interferencia (también recolecta un poco de la señal deseada, es como un conformador de haz de suma y resta pero en este caso el ruido es la fuente de interés y la señal deseada es considera interferencia). La salida de la matriz de bloqueo es procesada por una serie de filtros adaptativos G los cuales estarán cambiando dinámicamente (idealmente, los filtros se adaptarán de acuerdo al ambiente en el que se está llevando el proceso de filtrado).

Los filtros G se adaptarán con ayuda de un proceso llamado algoritmo de mínimos cuadrados (LMS). LMS lo que hace es adaptar los filtros en función del error estimado en la salida final de GSC.

Una vez que LMS encuentra los filtros óptimos, la salida Y_B restará sólo ruido a la salida Y_{SR} del conformador de haz de suma y resta, de esta manera en la salida final de GSC sólo tendremos la señal Y_{GSC} de interés libre de interferencias.

GSC se representa matemáticamente como una descomposición del vector de pesos $W(p)$ en dos componentes ortogonales W_0 y $W_B G$. La componente W_0 representa la parte en la que se lleva a cabo el conformador de haz de suma y resta. Por otro lado, $W_B G$ es la parte adaptativa.

$$w(p) = w_0 - w_B G \qquad (2.14)$$

W_B es la matriz de bloqueo, las salidas de dicha matriz representa filtros que contienen exclusivamente interferencias por lo tanto G es un vector conformado por una serie de filtros ajustables.

A la salida del formador de haz de suma y resta se tiene la señal Y_{SR} y de la sección adaptativa se obtiene una señal Y_B , a la salida de GSC se tiene la señal Y_{GSC} la cual representa una señal pura, libre de ruido, es decir, Y_{GSC} es la señal objetivo.

$$Y_{GSC} = Y_{SR} - Y_B \qquad (2.15)$$

El ajuste de los filtros es la parte complicada que tiene GSC ya que debe encontrar un filtro adecuado para que, en un tiempo determinado, las interferencias sean minimizadas, o en el mejor de los casos, canceladas. Este proceso de adaptación es importante ya que de él depende la cantidad de ruido existente en Y_B . Es en esta parte de GSC en donde interviene LMS.

Una vez que se conoce una señal de interés y las señales que causan interferencia, LMS se aplica para encontrar un filtro adecuado para minimizar el ruido. LMS tiene la función de minimizar el error entre la señal deseada y la señal estimada Y_{GSC} . El proceso de estimación se realiza en función del cálculo del gradiente que se incline más a la versión de los filtros pasados y los actuales. Esto es, en cada iteración el gradiente será multiplicado por un peso predefinido con el objetivo de encontrar un nuevo conjunto de filtros, los cuales se utilizarán para estimar una nueva señal.

El cálculo del gradiente es el siguiente:

$$\nabla G = Y_{GSC} X_n \qquad (2.16)$$

La actualización de los filtros está dada por:

$$G_{k+1} = G_k + \mu Y_{GSC} X_n \qquad (2.17)$$

Donde μ es la constante de adaptación. El valor que se le da a μ depende de las características del escenario en el que se ejecute GSC. Si dicho escenario está en constante cambio el valor elegido para μ no garantiza que el conformador de haz funcione correctamente a lo largo del tiempo.

Mientras más grande sea μ más grandes serán los pasos de adaptación. La multiplicación se realiza punto a punto para cada X_n .

En este proceso no se puede hacer el cálculo de la señal de una manera simultánea por lo tanto tiene que hacerlo a lo largo del tiempo de muestreo [34].

Mientras las fuentes de ruido no cambien constantemente (aumento de volumen, cambio de posición, aumento del número de fuentes, etc) los filtros G tendrán más posibilidades de adaptarse. Sin embargo, en un ambiente real las fuentes de interferencia cambian constantemente, por lo tanto GSC tendrá dificultades para encontrar el filtro optimo. Una de las soluciones a este problema es utilizar una μ con valores dinámicos y que se adapte a la cantidad de energía de la interferencia todavía presente en la salida. Una forma popular de llevar a cabo esta adaptación es por medio de aplicar la ecuación 2.18:

$$\mu = \begin{cases} \frac{\mu_0}{P_0}, \frac{\mu_0 P_B}{P_0} \wedge ¿ \mu_{max} \\ \frac{\mu_0}{P_B}, \frac{\mu_0 P_B}{P_0} \geq \mu_{max} \end{cases} \qquad (2.18)$$

Donde μ_0 es un valor base con el cual se calcula μ y normalmente es muy pequeño; μ_{max} es el valor máximo de μ ; P_0 y P_B son la potencia de salida de GSC y de la matriz de bloqueo respectivamente. μ_0 y μ_{max} definen una función de valores que puede tomar μ , y está siendo calculado en base a la proporción de potencia entre la salida y la estimación de interferencias [47].

De los conformadores de haz expuestos anteriormente, GSC es el que, por lo menos teóricamente, funciona mejor en ambientes reales ya que es un conformador de haz adaptativo.

Capítulo 3

Metodología de diseño

El objetivo de la tesina es determinar que tan viable es implementar GSC en los auxiliares auditivos. De acuerdo a la explicación dada en la sección anterior de este trabajo, para cada una de las técnicas de filtrado espacial, GSC es el que responde mejor en un ambiente acústico dinámico.

En la siguiente tabla se muestran las desventajas de Delay and Sum, de MVDR y de LCMV de igual manera se exponen las mejoras que ofrece GSC respecto a dichas desventajas.

Filtro espacial	Desventajas	Ventajas de GSC
Delay and Sum	Se necesitan muchos micrófonos para llevar a cabo un buen filtrado de la señal deseada.	Requiere de una cantidad moderada de micrófonos.
MVDR	El cálculo de la matriz de covariancia, para cada frecuencia, puede ser lento por lo tanto no es recomendable para aplicarlo en ambientes reales.	Tiene un intervalo de valores para μ lo que le permite encontrar el valor optimo de una manera menos lenta por lo tanto pude ser útil en un ambiente real.
LCMV	Debe saber no sólo la dirección de la señal deseada sino también la dirección de las interferencias para poder poner ceros en las mismas.	Solo debe conocer la dirección de la señal objetivo y lo demás lo considera interferencia.

Tabla 3. 1 Desventajas de Delay and Sum, de MVDR y de LCMV y mejoras que ofrece GSC respecto a dichas desventajas.

Debido a las características expuestas en la tabla 3.1 se determinó que GSC podría ser un conformador de haz adecuado para ser implementado en los auxiliares auditivos.

El hecho de que GSC sea adaptativo permite capturar una señal de audio deseada en un ambiente en donde haya muchas fuentes de interferencia. En la implementación que se realizó en este trabajo no se tuvo una μ estática sino que se asignó una μ_0 igual a 0.001 y una μ_{max} igual a 0.01.

El código realizado fue implementado en lenguaje C en conjunto con un servidor de procesamiento de señales de audio en tiempo real llamado Jack Audio Connection Kit.

3.1 Jack Audio Connection Kit

Jack Audio Connection Kit es un servidor que procesa señales de audio en tiempo real ya que tiene una latencia baja. Fue desarrollado por Paul Davis. El servidor está licenciado bajo GNU GPL, mientras que las bibliotecas están licenciadas bajo GNU LGPL [35].

Jack se puede instalar y usar en diversos sistemas operativos como Linux, OS X, Solaris, FreeBSD y Windows. Puede conectar varias aplicaciones denominadas clientes a un dispositivo de audio y permitirles interactuar entre sí. Los clientes pueden ejecutarse como procesos independientes, como las aplicaciones normales, o dentro del servidor JACK como "plugins".

El servidor de Jack tiene dos objetivos principales: la ejecución síncrona de todos los clientes y la operación de baja latencia.

Jack realiza distintas actividades:

- Elimina el hardware de interfaz de audio y permite a los programadores concentrarse en la funcionalidad principal del software.

- Admite que las aplicaciones envíen y reciban datos de audio entre sí, al mismo tiempo permite que interactúen con la interfaz de audio. No hay diferencia en cómo una aplicación envía o recibe datos independientemente de si proceden de otro agente o de una interfaz de audio.

El funcionamiento de Jack es el siguiente:

- Llama a **jack_client_open** para conectarse al servidor Jack.

- Registra "puertos" para permitir que los datos sean trasladados hacia y desde su aplicación.

- Registra un "*callback* " que será llamado en el momento adecuado por el servidor Jack.

- Finalmente se le notifica a Jack que la aplicación está lista para comenzar a procesar datos.

Jack es muy sencillo de instalar en Ubuntu de la versión 12.04 en adelante. Desde una terminal sólo se debe dar el siguiente comando:

sudo apt-get install jackd2 libjack-jackd2-dev pulseaudio-module-jack qjackctl

El servidor de Jack requiere de las siguientes herramientas:

- **jackd2** : servidor de JACK
- **qjackctl** : interfaz para configurar y controlar al servidor de JACK
- **libjack-jackd2-dev** : librerías de C, C++ para crear agentes de JACK
- **pulseaudio-module-jack** : módulo de *PulseAudio* para que puedan ejecutarse simultáneamente JACK y *PulseAudio* (La mayoría de los programas de Ubuntu se apoyan en *PulseAudio* para acceder a dispositivos de Audio)

El servidor de Jack trabaja con agentes. Un agente es un programa realizado bajo un lenguaje de programación, en este trabajo utilizamos C. Varios agentes pueden trabajar a la vez sin interferir uno con el otro.

La comunicación del servidor de Jack con los agentes es por medio de escribir y/o leer información en arreglos de datos (estos arreglos son ventanas de audio). Todo lo anterior se logra gracias a una función llamada *callback*.

El servidor de Jack manda llamar al *callback* de todos los agentes activos en cada ciclo de audio, permitiéndoles una cierta duración de tiempo para leer y/o escribir la información (*frames,* muestras que se toman en un intervalo de tiempo) en las ventanas de audio. Dicho tiempo es limitado por la latencia. Si el agente tarda mucho en responder, la información que no haya entregado será desechada en el próximo ciclo de audio. El tamaño de las ventanas de audio está estipulado por el servidor de Jack, el cual puede ser configurado en la interface *QjackCtl*.

Para crear un agente intervienen las siguientes funciones:

- **jack_client_open:** Asigna el nombre que se le dio al agente, en caso de que ese nombre ya exista el bit de *JackNameNotUnique* lo notifica al usuario. Si el agente no se pudo ejecutar, el bit *JackNoStartServer* indica si es que dicho error fue causado por la ausencia de algún servidor Jack corriendo.
- **jack_set_process_callback:** Define cuál función es la que el servidor de Jack va a mandar a llamar en este agente en cada ciclo de audio.
- **jack_on_shutdown:** Le informa al servidor de Jack cuál función debe llamar cuando el servidor es detenido o cuando el agente requiere ser desactivado.
- **jack_port_register:** Registra los puertos de entrada y salida. Se le debe proporcionar la variable del cliente creada por *jack_client_open*, el nombre del puerto y el tipo de audio que se va a utilizar (usualmente *JACK_DEFAULT_AUDIO_TYPE*). También se le debe indicar si el puerto es de entrada o de salida (*JackPortIsInput* si es de entrada, *JackPortIsOutput* si es de salida). El servidor de Jack tiene un máximo de puertos que puede registrar, por lo tanto *jack_port_register* nos notifica si ya se sobrepasó dicho umbral.
- **jack_activate:** Activa al agente. Si la activación se realiza con éxito, el agente es agregado a la lista de agentes del servidor Jack.
- **jack_connect:** Después activar al cliente, se puede conectar sus puertos a otros agentes. Se conectan salidas de agentes a entradas de agentes (no es válido conectar salidas con salidas o entradas con entradas). Esto se puede hacer de manera manual en la interfaz "*Connect*" en *QjackCtl* o con el comando *jack_connect*, entregándole como argumento el nombre de la salida y la entrada a conectar. El nombre de un puerto tiene el siguiente

formato: “<nombre de agente>:<nombre de puerto>”.

- **jack_get_ports:** Nos proporciona los nombres de puertos disponibles. Dichos puertos pueden pertenecer al agente “*System*”, también son conocidos como puertos físicos. Una salida del agente “*System*” es un micrófono, y una entrada es una bocina.
- **jack_port_get_buffer:** Regresa un apuntador a un buffer de datos de audio de algún puerto. Requiere la variable del puerto creada por *jack_port_register* y el tamaño del buffer a regresar.

Las funciones antes mencionadas son las esenciales al momento de crear un agente, lo cual facilita la creación de los mismos ya que se puede realizar una plantilla del código mínimo de un agente de Jack y sólo anexar las variables que se vayan requiriendo.

En cuanto a la configuración de la interfaz *QjackCtl*, se deben seleccionar los parámetros de acuerdo a las necesidades del usuario tales como *Realtime*, *Periods/Buffer*, selección de dispositivos de salida y entrada, etc.

En este trabajo elegimos trabajar con el servidor de Jack porque es accesible y fácil de utilizar.

3.2 Implementación de GSC con Jack Audio Connection Kit

En esta sección se describirán brevemente las partes del código en el que se implementó GSC. El nombre de este agente es *Beamforming*. El código completo se encuentra en el anexo.

En la primera parte se encuentra el *Beamforming Delay and Sum*, el tiempo que tardará en ejecutarse este *Beamforming* dependerá del tamaño de los *frames* (1024).

```
for (K=0; K<nframes; K++) {
    gsc_buffA=(in1[K] + in2[K])/2;
```

Posteriormente se tiene la Matriz de Bloqueo, dicha matriz ejecutará un proceso parecido al Delay and Sum, la diferencia radica en que en esta parte se recolectan las señales que causan interferencia. Otra característica es que se toma el tamaño de N_W (número de frames tomados por cada ciclo de audio, este dato fue modificado manualmente) y no el de nframes. Los valores que tomó N_W fueron 8, 16, 32, 64, 96 y 128.

```
for(i=1;i<Nw;i++)
    gsc_buffB[i-1] = gsc_buffB[i];
gsc_buffB[Nw-1] = (in1[K] - in2[K]);
```

En el siguiente apartado intervienen los filtros. Estos filtros trabajarán con los datos que se obtengan de la matriz de Bloqueo y de N_W .

```
gsc_buffC=0;
for(i=0;i<Nw;i++)
    gsc_buffC += g[i]*gsc_buffB[i];
```

A continuación se tiene la salida de GSC: el resultado de Delay and Sum menos el resultado obtenido del proceso de filtrado.

```
outA[K] = gsc_buffA - gsc_buffC;
outB[K] = outA[K];
```

El siguiente ***for*** almacena los valores de cada N_W . Los valores obtenidos los entrega a la salida A.

```
for(i=1;i<Nw;i++)
        gsc_o[i-1] = gsc_o[i];
gsc_o[Nw-1] = outA[K];
```

En la siguiente parte del código se lleva a cabo el cálculo de las potencias a la salida del GSC (p_o) y de la matriz de Bloqueo (p_buffB).

```
p_o = 0;
p_buffB = 0;
for(i=0;i<Nw;i++){
        p_o += gsc_o[i]*gsc_o[i];
        p_buffB += gsc_buffB[i]*gsc_buffB[i];
```

Con los valores obtenidos de las potencias se determina el valor que tomará μ a lo largo del proceso de adaptación. Estos valores son impresos en pantalla, se graficaron y fueron utilizados para el análisis de resultados. Son importantes porque nos muestran el proceso de adaptación de μ (el valor de μ_0 fue de 0.001 y el de μ_{max} fue de 0.01). Debido a que se involucran las potencias de la salida de GSC y de la matriz de Bloqueo, los resultados obtenidos no están dentro del rango establecido.

```
if(mu0*p_buffB/p_o < mu_max){
        r += mu0*outA[K]/p_o;
}else{
        r += mu0*outA[K]/p_buffB;
}
```

Una vez que se determina cuál es el valor óptimo de μ , se lleva a cabo el algoritmo de mínimos cuadrados.

```
for(i=0;i<Nw;i++){

        if(mu0*p_buffB/p_o < mu_max){
                g[i] += mu0*outA[K]*gsc_buffB[i]/p_o;

        }else{
                g[i] += mu0*outA[K]*gsc_buffB[i]/p_buffB;
        }
```

```
            if(isnan(g[i]))
                    g[i] = 0.0;
    }
```

Este último ***prinf*** nos imprime en pantalla el valor de μ .

```
printf ("%f ",r/nframes);
```

Las partes restantes del código son funciones que requiere el servidor de Jack para llevar a cabo el procesamiento de señales de audio en tiempo real. En él se utilizaron librerías de C y del servidor Jack, se le dio nombre al agente (*Bemaforming*), también se le asignaron nombres a los puertos de entrada y salida, se declararon *Buffers*. Toda la parte de GSC se desarrolla dentro de la función *callback*.

Para que este código nos entregara la salida de GSC en formato WAV se utilizó la función:

```
audio_info.format =  SF_FORMAT_WAV | SF_FORMAT_PCM_16;
audio_file = sf_open ("audio.wav",SFM_WRITE,&audio_info);
```

Está función ayudó mucho a comprender los resultados obtenidos al determinar la Relación Señal-Interferencia (SIR), los cuales se explican en el siguiente apartado.

Capítulo 4

Evaluación y Resultados

A continuación se exponen los resultados de las pruebas realizadas con *Beamforming* GSC. En ellas se observa que en ambientes reales, el desempeño de GSC no es adecuado.

El corpus recolectado por el grupo Golem se utilizó para simular dos escenarios: un ambiente ideal y un ambiente real. La manera en la que se calificó el desempeño del conformador de haz fue mediante el resultado del SIR (Relación Señal Interferencia, la cual mide la distorsión causada por la interferencia de otras fuentes sobre la fuente de interés).

También se obtuvieron gráficas en las que se muestra el comportamiento de μ a lo largo del proceso de adaptación para los distintos valores de N_W (tamaño del filtro aplicado en cada prueba).

4.1 Pruebas realizadas con el Corpus recolectado por el grupo GOLEM

En esta sección se describen las pruebas para evaluar el desempeño de GSC a partir de dos herramientas: el corpus llamado *"Interacciones Acústicas para Audición de Robots"* (AIRA) recolectado por el grupo GOLEM y los archivos WAV que se obtuvieron de GSC.

AIRA es un conjunto de audios que fue recolectado en dos ambientes diferentes: la cámara anecóica del Laboratorio de Acústica y Vibraciones del CCADET-UNAM y en el laboratorio de estudiantes del Departamento de Ciencias de la Computación del IIMAS-UNAM.

La idea de recolectar las grabaciones en dos ambientes diferentes se debe a que en el CCADET-UNAM se simula un ambiente ideal, en este no hay reverberación y las únicas fuentes de interferencia son fuentes sonoras estáticas (personas hablando). El otro escenario, el que se realizó en el IIMAS, es un entorno acústico real, en él la reverberación prevalece y las fuentes de interferencia son muchas (ruidos de los ventiladores de las computadoras, personas hablando y caminando, etc).

Para realizar AIRA se utilizaron de 1 a 4 fuentes sonoras estáticas situadas a 1 [m] del centro del arreglo de micrófonos. Dicho arreglo consta de tres micrófonos omnidireccionales, los cuales formaron un triángulo equilátero. En la cámara anecóica, la distancia que había entre los elementos del arreglo fue de 0,18 [m]. Mientras que en el laboratorio del IIMAS los elementos estaban separados por 0,21 [m] de distancia [38].

Los audios realizados en el CCADET reciben el nombre de "*Clean*", los cuales incluyen el número de personas que intervienen (fuentes) y el ángulo horizontal (o *azimuth*) respecto al arreglo de micrófonos de cada fuente. Por otro lado, las grabaciones recopiladas en el IIMAS reciben el nombre de "*Noisy*", de igual manera incluye el número de personas que intervienen (fuentes) y el ángulo horizontal respecto al arreglo de micrófonos.

Para cada fuente se tiene la grabación antes de ser reproducida, el dialogo empleado y la posición (el ángulo y distancia del centro del arreglo).

GSC se programó con el lenguaje C, cuando dicho programa se ejecutó intervino el servidor *Jack. Jack* emuló las entradas de los micrófonos con las grabaciones de AIRA, para realizar esto se utilizó el agente llamado *ReadMicWavs* el cual reproduce los audios y crea un micrófono para cada uno, una vez que la grabación se acaba, *ReadMicWavs* deja de funcionar.

Gracias a lo descrito anteriormente, el agente *Beamforming* recibe la grabación proporcionada y ejecuta GSC, simultáneamente corre *ReadMicWavs*. Finalmente entrega el resultado de GSC en un archivo WAV que al compararlo con el audio original nos da una idea del desempeño de GSC.

Para evaluar el desempeño de GSC se determinó el valor de la Relación Señal Interferencia (SIR), esta relación indica, en decibeles, la calidad de la señal deseada frente a la interferencia.

Para determinar el SIR se utilizó un código implementado en Octave (programa para realizar cálculos númericos) llamado ***bss_eval_sources***, el cual hace referencia a la Separación Ciega de Fuentes de Audio (Blind Source Separation, BSS por sus siglas en inglés) [39].

Para este trabajo, el objetivo de ***bss_eval_sources*** es medir el SIR por lo tanto necesita tener todas las fuentes en su forma original, es decir, sin interferencia (el número de fuentes depende del número de personas que intervienen en el audio), también requiere la fuente estimada.

Para obtener el SIR se corrió el programa anterior para cada archivo de audio.

Las tablas 4.1 y 4.2 muestran los SIRs obtenidos para cada N_W aplicado a las diferentes capturas de audio, de acuerdo a los resultados del SIR obtenidos, determinamos en que escenario (sin interferencia o con interferencia) GSC se desempeña mejor. El mejor valor del SIR es el que se acerca a los 20 dB.

La primera columna de la tabla contiene el nombre del archivo utilizado del corpus AIRA. La segunda columna nos muestra el valor de N_W , es decir, el tamaño de la ventana que se utilizó durante el proceso de adaptación en GSC. En la tercera columna se muestran los errores que registró el servidor Jack durante la ejecución de GSC, los cuales se presentan cuando el agente no regresa información suficientemente rápido y el servidor Jack los descarta; esta columna es presentada para mostrar si el agente es suficientemente rápido para correr en tiempo real. En la cuarta columna tenemos el valor del SIR. El número de fuentes corresponde al número de personas que se encuentran en la grabación. Por último se tiene el ángulo horizontal respecto al arreglo de micrófonos de cada fuente.

CORPUS	Nw	Errores de Jack	SIR dB	Número de fuentes	Ángulo horizontal [°]
Clean-2source	8	0(0)	15.838	2	-30, 90
	16	0 (0)	17.509	2	-30, 90
	32	0 (0)	16.234	2	-30, 90
	64	0(0)	15.011	2	-30, 90
	96	0(1)	16.179	2	-30, 90
	128	10(34)	14.288	2	-30, 90
Clean-2source090	8	0(0)	11.496	2	0, 90
	16	0(0)	10.419	2	0, 90
	32	0 (0)	13.635	2	0, 90
	64	0(0)	17.215	2	0, 90
	96	0(1)	19.131	2	0, 90
	128	11(33)	19.154	2	0, 90
Clean-2source090_2	8	0(0)	7.7045	2	0, 90
	16	0(0)	8.4238	2	0, 90
	32	0(0)	-1.7714	2	0, 90
	64	0(1)	4.1769	2	0, 90
	96	0(0)	18.737	2	0, 90
	128	0(1)	15.743	2	0, 90
Clean-3source	8	0(0)	6.4630	3	-30, 90, -150
	16	0(1)	3.5602	3	-30, 90, -150
	32	0(0)	2.1297	3	-30, 90, -150
	64	1(1)	-0.052204	3	-30, 90, -150
	96	0(0)	-1.7364	3	-30, 90, -150
	128	10(32)	-2.0245	3	-30, 90, -150
Clean-3source090180	8	0(0)	-1.4665	3	0, 90, 180
	16	0(0)	-5.3515	3	0, 90, 180
	32	0(0)	-1.5616	3	0, 90, 180
	64	0(1)	-1.2748	3	0, 90, 180
	96	0(0)	0.62364	3	0, 90, 180
	128	11(29)	-0.54108	3	0, 90, 180
Clean-4source	8	0(0)	-8.4025	4	0, 90, 180, -90
	16	0(0)	-8.6286	4	0, 90, 180, -90
	32	0(0)	-8.7169	4	0, 90, 180, -90
	64	0(1)	-8.6040	4	0, 90, 180, -90
	96	0(0)	-3.7053	4	0, 90, 180, -90
	128	9(24)	-8.5125	4	0, 90, 180, -90

Tabla 4.1 Resultados de GSC para los distintos valores de N_W con los audios Clean.

Como se puede observar, *Clean-2source090* registró los mejores SIRs para los diferentes N_W, ya que alcanzó valores cercanos a 20 dB. Para analizar estos resultados se debe tomar en cuenta que en este audio solo hay dos fuentes y que la grabación fue realizada en el CCADET-UNAM.

CORPUS	Nw	Errores de Jack	SIR dB	Número de fuentes	Ángulo horizontal [°]
Noisy-2source	8	0(0)	-3.6823	2	-30, 90
	16	0(0)	-7.7240	2	-30, 90
	32	0(0)	-5.3804	2	-30, 90
	64	0(1)	-5.9809	2	-30, 90
	96	7 (10)	-4.3615	2	-30, 90
	128	11(17)	-4.7767	2	-30, 90
Noisy-2source090	8	0(0)	-4.9522	2	0, 90
	16	0(1)	-4.2154	2	0, 90
	32	0(0)	-1.2078	2	0, 90
	64	0(1)	-4.2634	2	0, 90
	96	0(0)	-3.3077	2	0, 90
	128	12(29)	-1.9386	2	0, 90
Noisy-3source	8	0(0)	-0.22824	3	-30, 90, -150
	16	0(0)	-0.34353	3	-30, 90, -150
	32	0(0)	1.2376	3	-30, 90, -150
	64	0(0)	-1.0459	3	-30, 90, -150
	96	9(14)	-0.66226	3	-30, 90, -150
	128	15(40)	-1.0855	3	-30, 90, -150
Noisy-3source090180	8	0(1)	-11.727	3	0, 90, 180
	16	0(0)	-9.9887	3	0, 90, 180
	32	0(0)	-9.9209	3	0, 90, 180
	64	0(0)	-11.401	3	0, 90, 180
	96	5(8)	-10.092	3	0, 90, 180
	128	8(15)	-11.368	3	0, 90, 180
Noisy-4source	8	0(0)	0.49952	4	0, 90, 180, -90
	16	0(0)	0.63366	4	0, 90, 180, -90
	32	0(0)	1.0111	4	0, 90, 180, -90
	64	0(0)	2.1185	4	0, 90, 180, -90
	96	6(7)	1.4793	4	0, 90, 180, -90
	128	7(18)	1.0417	4	0, 90, 180, -90

Tabla 4.2 Resultados de GSC para los distintos valores de N_W con los audios Noisy.

Por otro lado, *Noisy-3source090180* registró los peores SIRs (por debajo de -11 dB). Para este caso se debe considerar que la grabación fue realizada en el Laboratorio de estudiantes del Departamento de Ciencias de la Computación del IIMAS-UNAM.

A continuación se despliegan las figuras 4.2, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 y 4.8, las cuales muestran la adaptación de μ a lo largo de todo el proceso de GSC.

Solo se tomaron los resultados más representativos y las comparaciones se hicieron de la siguiente manera:

- Mejor caso *Clean* comparado con el peor caso *Clean*
- Mejor caso *Noisy* comparado con el peor caso *Noisy*

Primeramente se muestran las gráficas que comparan el mejor y peor caso en *Clean.*

La Figura 4.1 corresponde a *Clean-2source090,* el mejor caso *Clean, y* muestra los valores que tomó μ para los distintos valores de N_W .

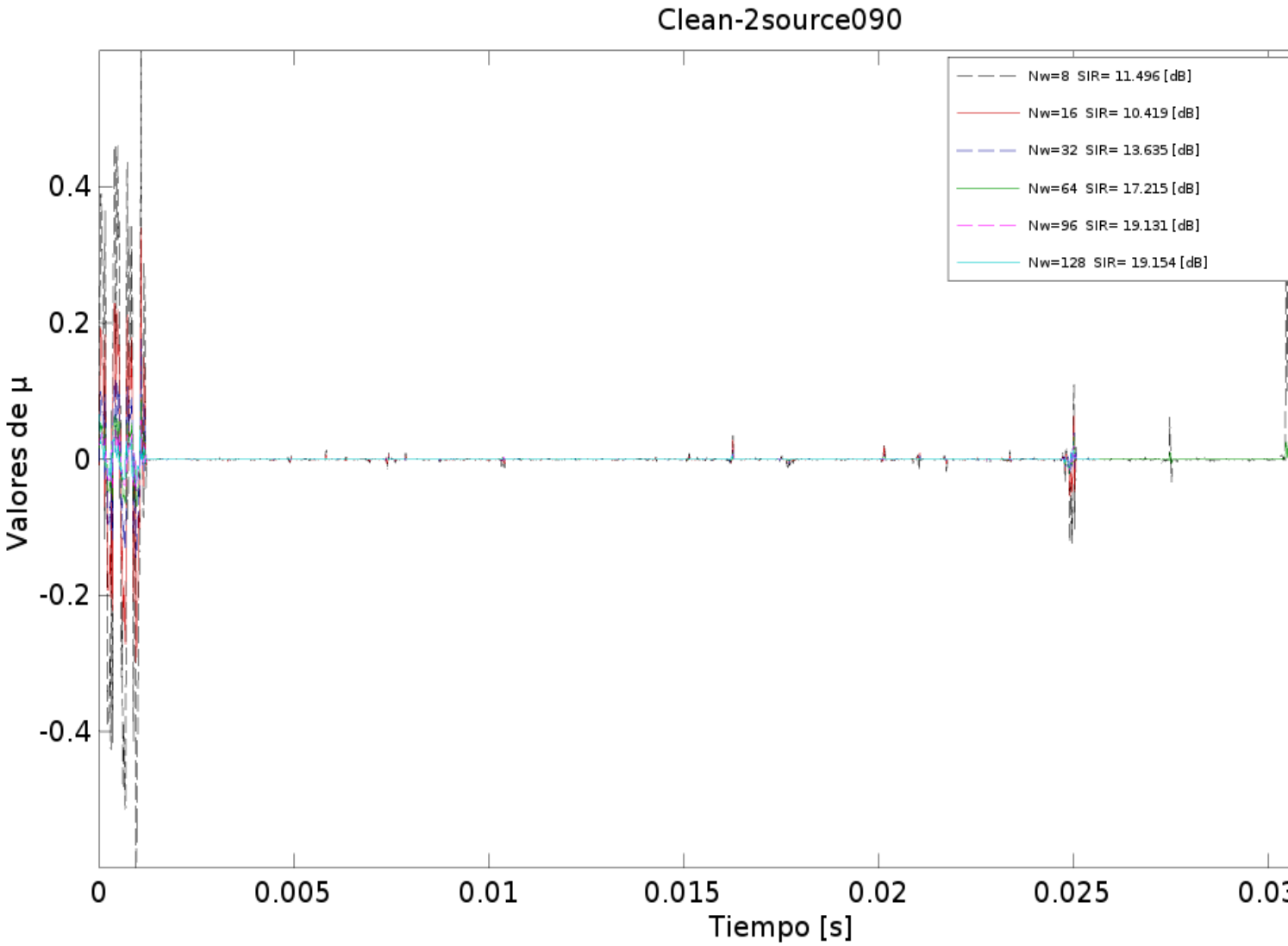

Figura 4.1 Valores de μ para los distintos N_W correspondientes a *Clean-2source090.*

La Figura 4.2 también pertenece a *Clean2-source090* pero tiene aplicado un zoom el cual ayuda a observar con mayor claridad el comportamiento de μ .

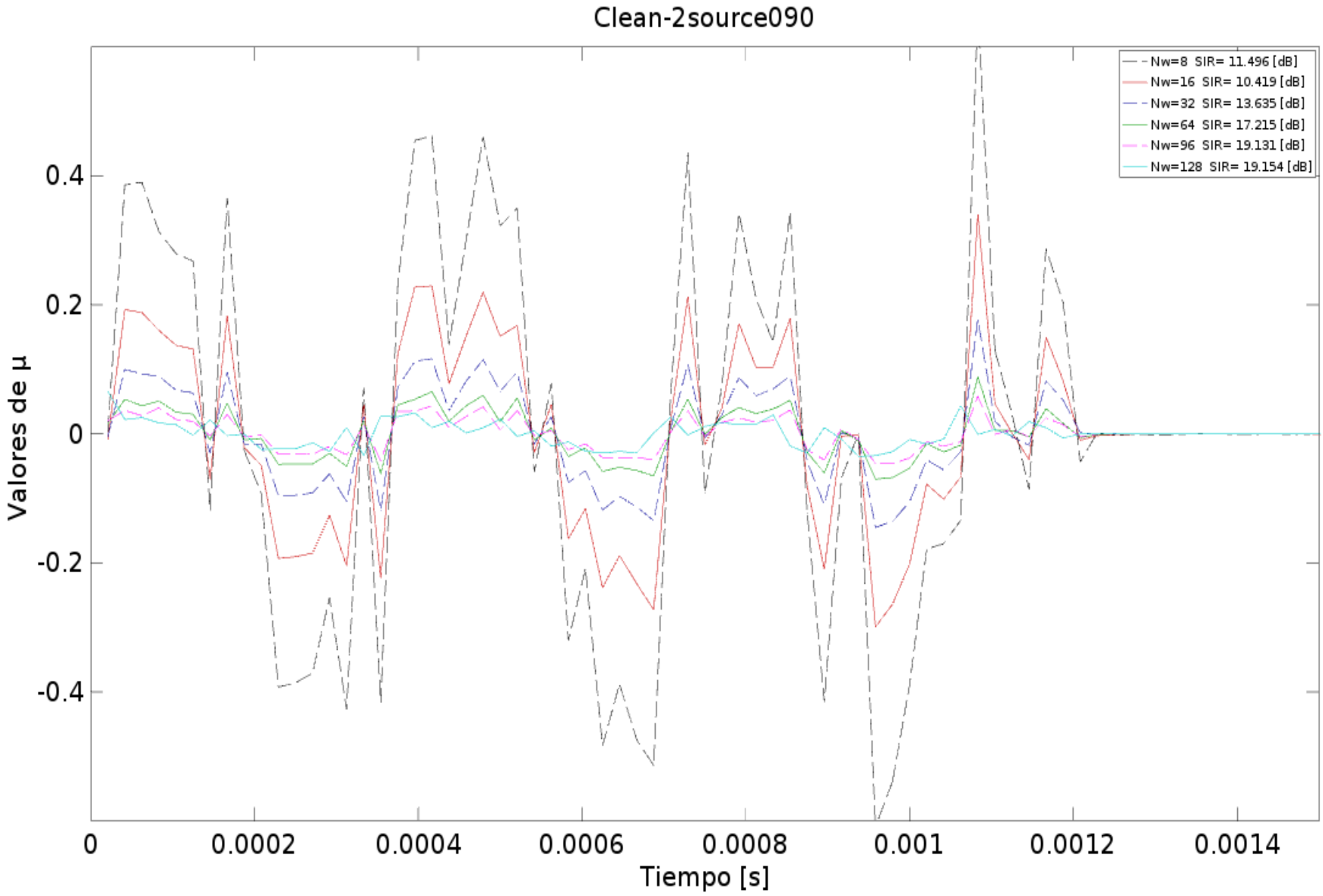

Figura 4.2 Valores de μ para los distintos N_W correspondientes a *Clean-2source090 con un zoom aplicado.*

La Figura muestra el peor caso *Clean*, *Clean-4source.*

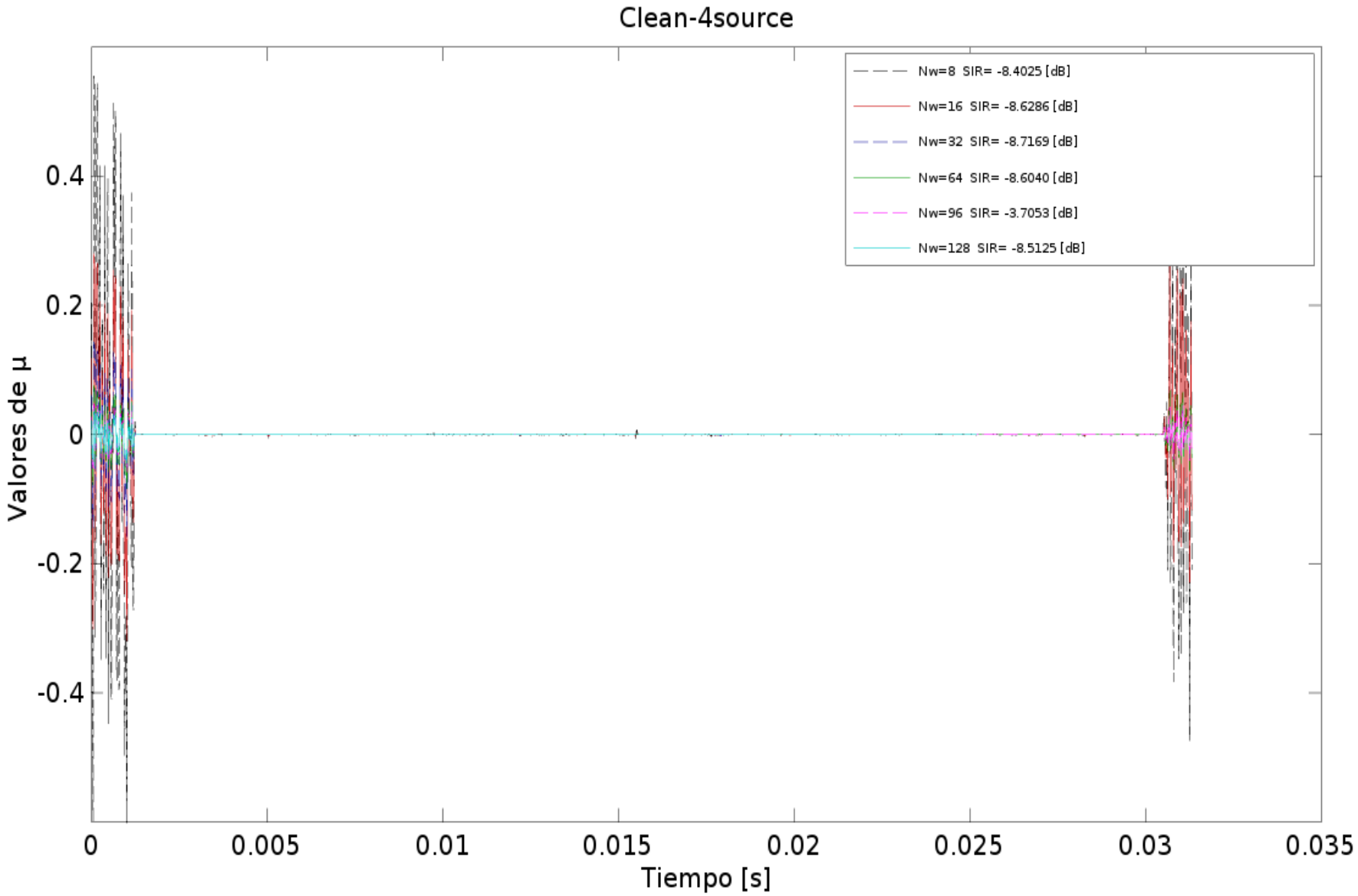

Figura 4.3 Valores de μ para los distintos N_W correspondientes a *Clean-4source.*

La Figura 4.4 muestra el zoom aplicado a la Figura 4.3 para observar más detalladamente el comportamiento de μ .

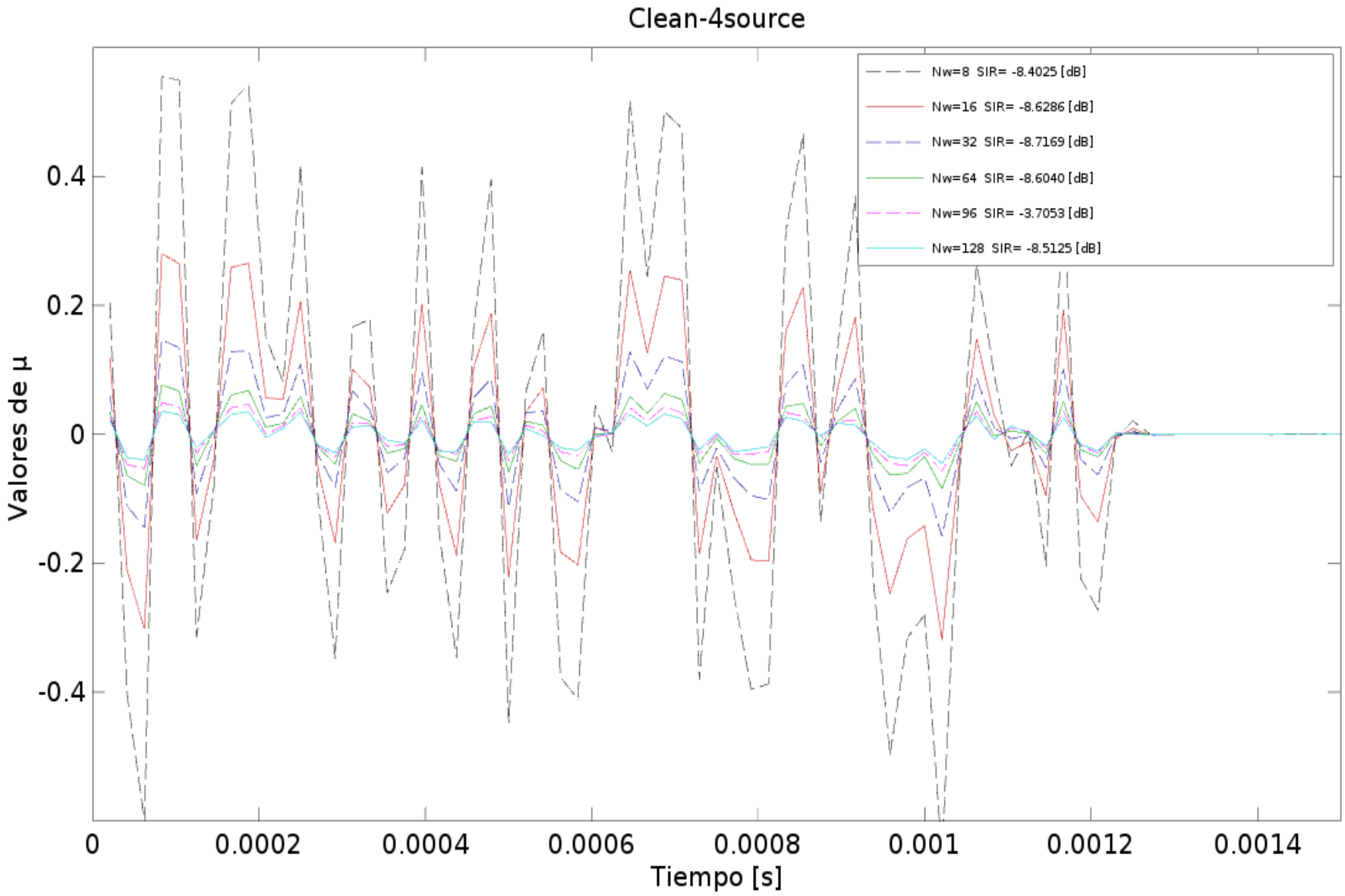

Figura 4.4 Valores de μ para los distintos N_W correspondientes a *Clean-4source con un zoom aplicado.*

Ahora se muestran las comparaciones entre el mejor y peor caso *Noisy.*

La Figura 4.5 muestra el comportamiento de μ para los distintos N_w correspondientes al mejor caso *Noisy*, *Noisy-4source.*

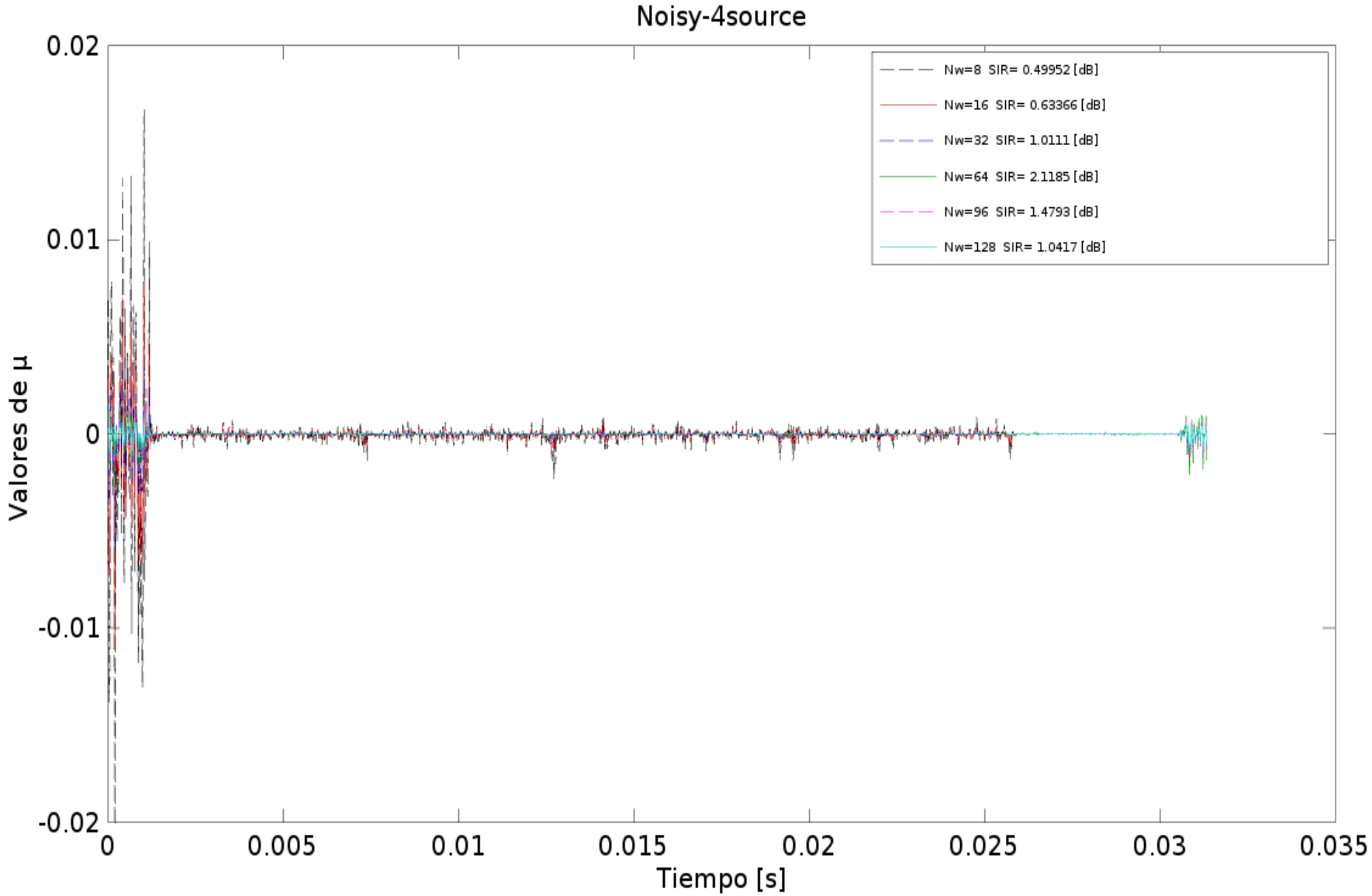

Figura 4.5 Valores de μ para los distintos N_w correspondientes a *Noisy-4source.*

En Figura 4.6 se muestra un zoom aplicado a la Figura 4.5 para observar más claramente el comportamiento de μ .

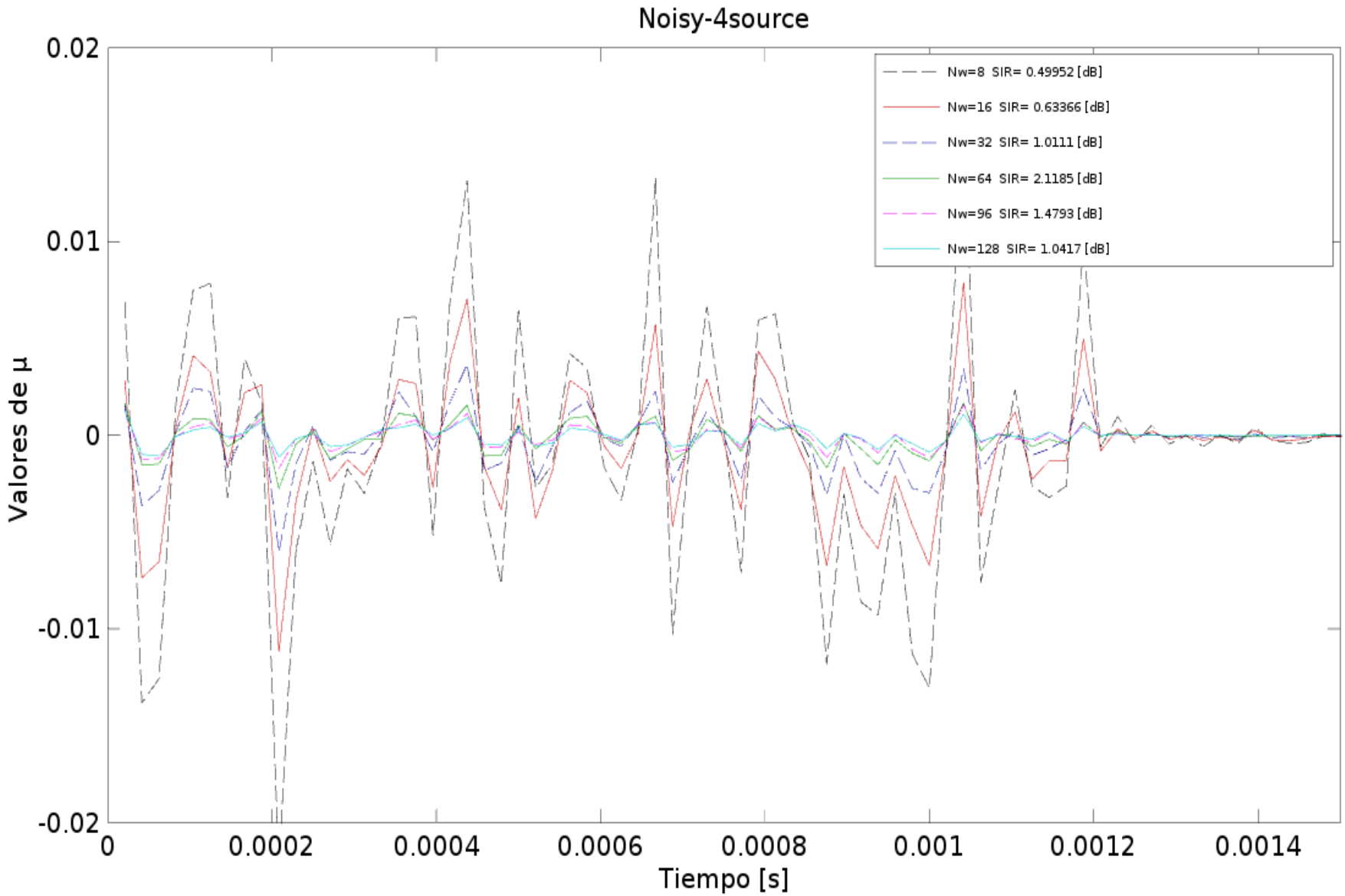

Figura 4.6 Valores de μ para los distintos N_W correspondientes a *Noisy-4source con un zoom aplicado.*

La Figura 4.7 muestra el peor caso *Noisy*, *Noisy-3source09180.*

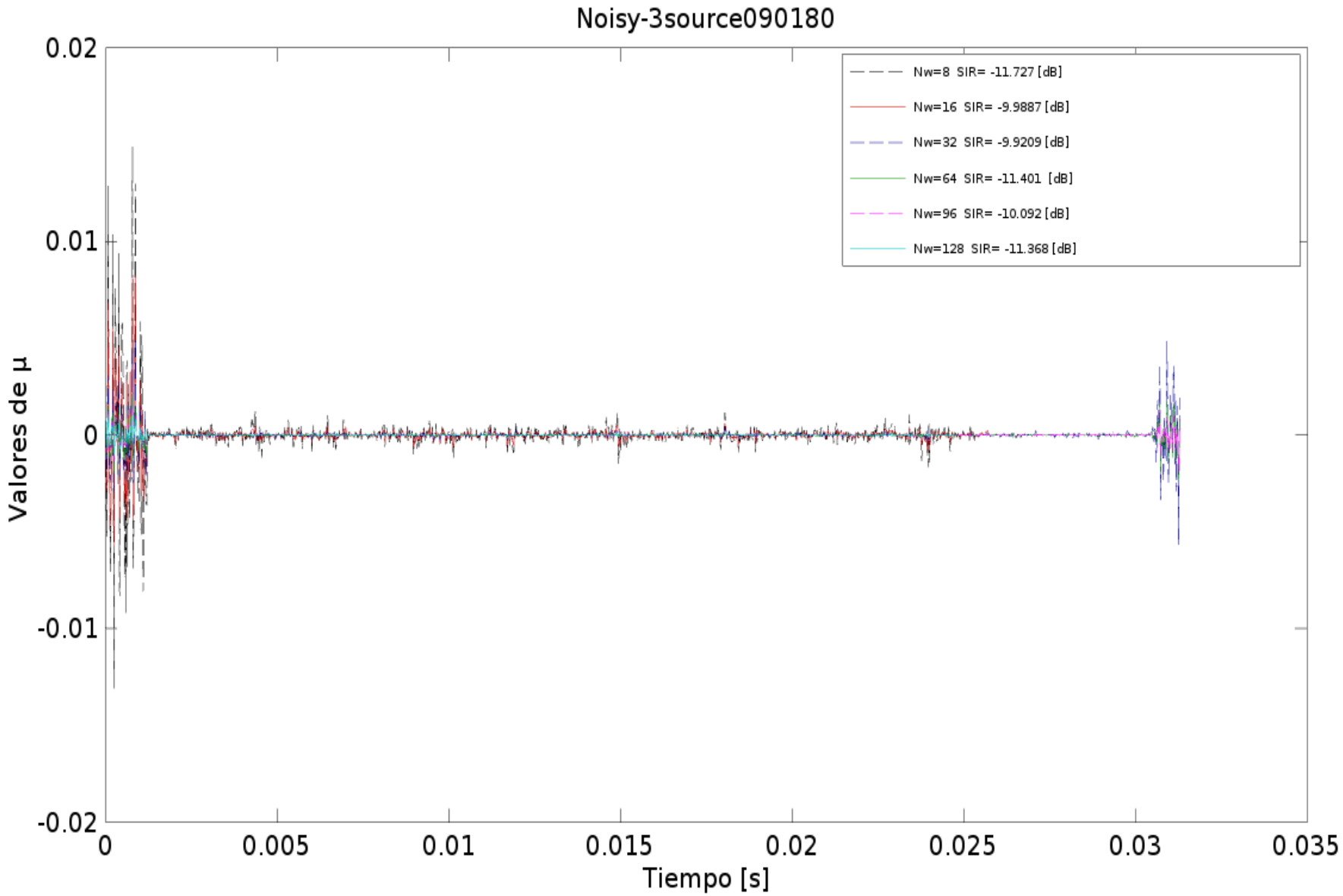

Figura 4.7 Valores de μ para los distintos N_W correspondientes a *Noisy-3source090180.*

La Figura 4.8 muestra un zoom aplicado a la Figura 4.7 con el fin de ver con más detalle el comportamiento de μ .

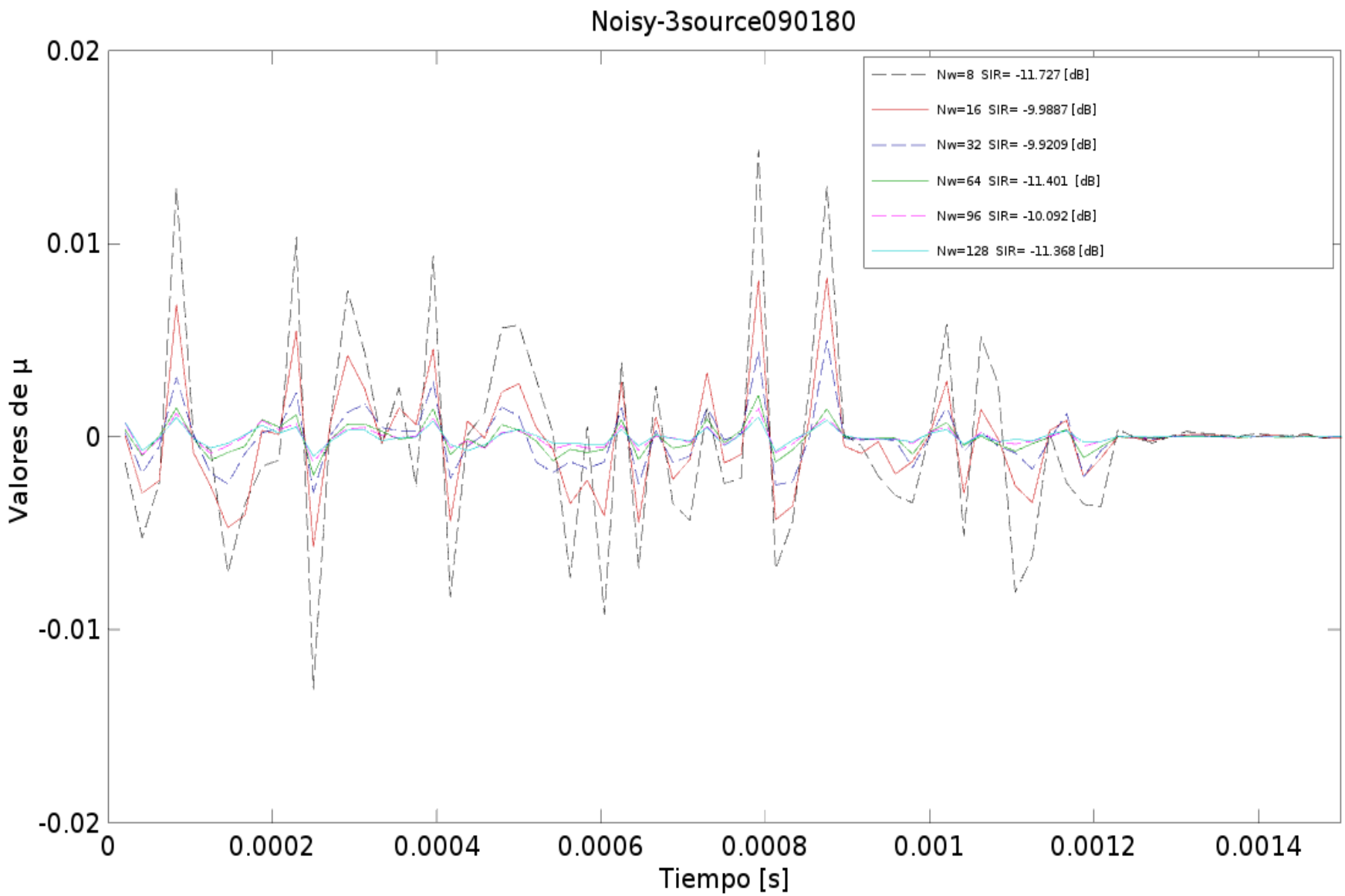

Figura 4.8 Valores de μ para los distintos N_W correspondientes a *Noisy-3source090180 con un zoom aplicado.*

4.2 Análisis de Resultados

La manera en la que se estableció la efectividad de GSC en los auxiliares auditivos fue determinando la relación señal interferencia (SIR), un resultado muy cercano a 20 dB indica que GSC funciona adecuadamente. Por el contrario, un resultado lejano a 20 dB muestra que GSC no tiene un buen desempeño.

En los resultados para los casos *Clean*, que muestra la Tabla 4.1, se observa que GSC funciona de una manera aceptable cuando hay poca reverberación y existen pocas fuentes de interferencia. Sin embargo, no todas las pruebas realizadas en la cámara anecóica fueron exitosas, los resultados muestran que cuando hay dos o más fuentes de interferencia el desempeño de GSC se ve afectado. Ejemplo de esto es *Clean-4source* ya que registró valores para el SIR de hasta -8.7169 [dB].

En la Tabla 4.2 se muestran los resultados de los casos *Noisy*. Estos resultados demuestran que la reverberación afecta significativamente el desempeño de GSC. Por ejemplo para *Noisy-2source090*, el cual sólo tiene una fuente de interferencia, el peor SIR registrado fue de -1.9386 [dB]. Sin embargo, este resultado es mayor que los registrados para *Clean-4source.*

Con los resultados anteriores se puede determinar que, aunque hay pocas fuentes de interferencia, si hay una presencia moderada de reverberación, el desempeño de GSC se ve afectado. Otros factores que influyen son la ubicación de las fuentes de interferencia, el número de individuos que intervienen, lo que dicen, y cómo lo dicen, así como el escenario en donde se ejecute.

Los resultados obtenidos de GSC con el caso *Clean-2source090* fueron los que registraron valores cercanos a los 20 dB con un N_W igual a 128 y 96 (N_W representa el número de muestras tomadas por cada ciclo de audio) gracias a estos resultados pudimos determinar que GSC obtuvo un desempeño aceptable. Para este caso se determinó que entre más grande es N_W , el proceso de filtrado que realiza GSC es mejor.

En el caso de *Clean-2source090* intervienen dos personas. La fuente de interés es la grabación de una mujer, dicha grabación fue colocada a 90 [°] del arreglo de micrófonos mientras que la interferencia es la grabación de un hombre posicionada a 0 [°] del arreglo. Los dos tienen diálogos cortos y el volumen de su voz es el mismo. En esta situación GSC amplifica lo que dice la mujer ya que está en frente del arreglo de micrófonos y discrimina el resto. Por otro lado, también influye el tamaño de N_W , para valores arriba de 64 el desempeño de GSC es bueno, para valores inferiores no hace el filtrado tan bien sin embargo los resultados del SIR no son tan bajos y de todos modos la grabación de la mujer predomina ante la del hombre.

Lo característica primordial de GSC es que es un filtro espacial adaptativo, es decir, el valor de μ se adecua de acuerdo al escenario en el que GSC se desarrolla. Dicha adaptación se observa en las Figuras 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 y 4.8.

El comportamiento de μ para *Clean-2source090* cuando N_W es igual a 8 es muy variado, se puede observar que los valores que toma a lo largo del proceso de GSC no son estables y el filtro funciona de una manera aceptable. En las gráficas posteriores, se observa que mientras

N_w se incrementa, la búsqueda que realiza μ ya no es tan grande, de hecho, cada vez el rango de búsqueda es menor. Cuando N_w es igual a 128, los valores que toma μ son más constantes lo cual no quiere decir que no varíen, pero el rango en el que lo hacen es mínimo (entre 0.0001 y 0.0002). El trabajo que realiza GSC con un N_w igual a 128 y con un escenario en dónde sólo hay una fuente que interfiere es bueno ya que la relación señal interferencia alcanza un valor de casi 20 dB.

El comportamiento de μ , descrito anteriormente, se puede apreciar en la Figura 4.1. La línea azul celeste pertenece a N_w igual a 128, en dicha Figura se observa que se mantiene constante. Al aplicar un zoom, como lo muestra la Figura 4.2, se observa que N_w igual a 96 tiene casi el mismo comportamiento que 128. Entre más pequeño es N_w el rango de búsqueda de μ es más grande, esto queda demostrado con N_w igual a 8 (línea negra punteada) ya que se observa como μ da saltos muy grandes al no lograr encontrar el valor óptimo para llevar acabo el filtrado, a pesar de esto la señal de interés sobresale de la señal de interferencia.

Por otro lado, se tiene un resultado no tan favorable para otro escenario en el que no hay ruido en el ambiente pero el número de fuentes de interferencia es mayor, es el caso de *Clean-4source.*

Clean-4source cuenta con 4 fuentes (dos grabaciones de mujeres y dos de hombres), las cuales están ubicadas a 0 [°], 90 [°], 180 [°] y -90 [°] del arreglo de micrófonos. Los diálogos que se escuchan en estas grabaciones son cortos y el volumen de las voces con las que se dicen es el mismo. En este audio no hay fuentes de interferencia en el ambiente, sin embargo hay cuatro personas hablando al mismo tiempo lo que hace que para GSC sea más difícil ignorar lo que no está en frente del arreglo de micrófonos. Una de las características del arreglo de micrófonos es que los mismos son omnidireccionales y aun así el GSC trata de dar prioridad al audio de la mujer que está a 90 [°] del arreglo.

Para *Clean-4source* los SIR registrados fueron muy bajos, no dieron más de -8.4025 [dB], el único que se acercó a uno fue el que tuvo un N_w igual a 96, éste registró un SIR igual a -3.7053 [dB].

La Figura 4.4 muestra que para todos los valores de N_w , μ sigue el mismo trayecto en el proceso de adaptación sin embargo, el rango de valores que toma no son los mismos, entre más grande es N_w el rango de valores de búsqueda es menor.

Por otro lado, la Figura 4.3 muestra algo interesante para N_w igual a 32 y 128. Alrededor del momento 0.025 [s], μ se queda “estancada”: prácticamente, no cambia. En estos valores de N_w se registraron los peores valores de SIR. De acuerdo a los resultados obtenidos en *Clean-2source090,* se esperaba que mientras más grande fuera N_w el desempeño de GSC sería mejor, sin embargo en este caso se comprobó que no.

Comparando los resultados de los valores de N_w en *Clean-4source* y *Clean2-source90*, se determina que el tamaño de N_w no influye tanto como el ambiente en el que se ejecuta GSC. En el caso *Clean2-source90,* para cada N_w hubo variaciones significativas en cuanto al

valor del SIR, y se puede notar que entre mayor sea el N_w , se obtienen mejores valores del SIR. En cuanto al caso *Clean-4source,* los valores que tomó N_w , no hicieron mucha diferencia en cuanto a la calidad del audio obtenido ni en cuanto al valor del SIR.

Otro de los escenarios en los que se evaluó GSC fue en un ambiente real. Los resultados del SIR obtenidos no son significativos ya que no llegan ni a 3 dB y en los audios obtenidos de GSC no sobresale la grabación que representa la fuente de interés. Estos resultados concuerdan con lo que se expone en el artículo *Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement [41]***,** en donde se determina que GSC, en ambientes reales (moderadamente reverberativos), no alcanza valores mayores a 1 dB. Por lo tanto, el filtrado depende, en mayor parte, del número de micrófonos y de la distancia que hay entre ellos.

En cuanto a un escenario real, el mejor caso fue *Noisy-4source* donde incluso obtuvo mejor desempeño que *Clean-4source*. En esta grabación intervienen cuatro personas (tres hombres y una mujer) las cuales están ubicadas a 0 [°], 90 [°], 180 [°] y -90 [°] del arreglo de micrófonos. Los diálogos de todos los participantes son cortos, todos hablan con el mismo nivel de volumen. La fuente de interés es la mujer.

El mejor valor del SIR registrado para *Noisy-4source* fue de 2.1185 [dB] el cual tuvo un N_w igual a 64, el peor fue 0.49952 [dB] para un N_w igual a 8. En este caso se comprueba que el ambiente en el que se desarrolla GSC influye más que el tamaño del filtro que se aplica.

En cuanto a la adaptación de μ **,** se puede observar en la Figura 4.5, que para valores pequeños de N_w (8, 16, 32) el proceso de búsqueda de μ primero se realiza en rangos de valores grandes, y luego en un rango de valores muy pequeño. Posteriormente, y curiosamente, ocurre lo contrario que con *Clean-4source,* que se “estanca” sólo con un valor de N_w igual a 128, con el resto de los valores no ocurre dicho “estancamiento”. En el caso del *Noisy-4source,* los valores de N_w que produce que se “estanque” μ son la 8, 16 y 32. Por el contrario, valores de 64, 96 y 128 continúan y registran los mejores valores de SIR.

En la Figura 4.6 se puede observar que valores de N_w iguales a 96 y a 128 tienen casi el mismo comportamiento. De hecho, en dicha figura, presentan casi la misma línea. Por otro lado, un valor de N_w igual a 64, aunque sigue casi el mismo camino que las dos anteriores, hay puntos en los que difieren un poco. Dadas estas diferencias es por lo que determinamos que registró el mejor desempeño en cuanto al SIR. Los resultados anteriores nos dicen que GSC también puede ser capaz de funcionar de una manera aceptable en ambientes con distintas fuentes de interferencia.

Los peores resultados que se tuvieron de toda la prueba fueron del caso *Noisy-3source090180*. En este audio intervienen tres personas, dos mujeres y un hombre los cuales están posicionados a 0 [°], 90[°] y 180 [°]. La fuente de interés es el varón. Al igual que en los audios descritos anteriormente, los diálogos de los participantes son cortos y el volumen de sus voces es el mismo.

El mejor valor del SIR registrado para el caso *Noisy-3source090180* fue de -9.9209 [dB] con un N_w igual a 32. El resultado más bajo fue de -11.727 [dB] con un N_w igual a 8. De acuerdo

a los resultados obtenidos se puede observar que mientras se incremente el N_W a 16 y a 32, el SIR mejora. A pesar de esto, en el proceso de adaptación de μ para estos dos valores no hay ninguna relación ya que en la Figura 4.8 se observa como μ no sigue la misma trayectoria para ambos N_W. Hay puntos en los que sus comportamientos se parecen y otros en los que difieren considerablemente.

Respecto a los valores de μ para un valor de N_W igual a 32 en la Figura 4.7 se observa que el proceso de adaptación primero lo hace entre un rango de valores pequeño, posteriormente se va adaptando entre valores poco significativos pero no se mantiene constante. Este comportamiento difiere para valores de N_W mayores a 32, ya que en estos casos el valor de μ se mantiene estático.

Una de las razones por las que el caso *Noisy-4source* tuvo mejores SIRs que *Noisy-3source090180* se debe a que en este último audio intervienen dos mujeres y sólo un hombre. La voz de ellas abarcan frecuencias más altas: las voces masculinas rondan en los 125 [Hz] y las voces femeninas alrededor de los 215 [Hz]. Dado que la reconstrucción de las señales es en base al desfase entre los micrófonos, las señales de baja frecuencia no serán reconstruidas apropiadamente ya que su longitud de onda es mayor a la distancia de los micrófonos. Por lo tanto el conformador de haz GSC le dará “preferencia” a estas señales. Adicionalmente, si el conformador de haz está siendo “apuntado” en una dirección en la cual dichas frecuencias están siendo reflejadas por algún material (la pared, por ejemplo), se les dará prioridad aún sobre la señal de interés que está localizada en esa misma dirección, si ésta es mayormente de bajas frecuencias.

En los resultados descritos anteriormente se determina que GSC funciona bien en ambientes ideales. Aunque estos escenarios no existen en la vida diaria, hay sitios con pocas fuentes de interferencia en los que los usuarios de auxiliares auditivos desarrollan sus actividades. Por ejemplo cuando ven televisión, GSC les permitiría disfrutar de su programa favorito sin necesidad de subir el volumen a niveles exagerados. Otra situación en la que GSC puede ser útil es en una sala del cine, normalmente, mientras la película es proyectada, los espectadores permanecen en silencio.

Por otro lado, en lugares con muchas fuentes de interferencia y mucha reverberación el desempeño de GSC se degrada, por lo menos con dos micrófonos. En este trabajo solo se utilizaron dos micrófonos porque se contempló uno por oído y porque los usuarios requieren de auxiliares cada vez más discretos y pequeños. Para realizar un arreglo de micrófonos con un mayor número de elementos se requiere tomar en cuenta el tamaño físico de dichos micrófonos, así como la distancia que habría entre ellos para así poder determinar la cantidad que se podría utilizar.

Capítulo 5

Conclusiones

El objetivo de este trabajo fue determinar la viabilidad del Cancelador Generalizado de Lóbulos Laterales (GSC) en los auxiliares auditivos.

De acuerdo a los resultados de las pruebas realizadas se determinó que GSC no funciona de manera satisfactoria en ambientes con distintas fuentes de interferencia y reverberación ya que no logra una adaptación que disminuya o elimine las señales no deseadas y amplifique la fuente de interés.

En ambientes ideales, con una fuente de interferencia, GSC lleva a cabo el filtrado espacial de manera exitosa (siempre y cuando la fuente de interés posea frecuencias altas) lo cual indica que los pacientes con pérdida auditiva podrían sentirse cómodos al tener conversaciones uno a uno. Sin embargo, los ambientes en los que se desarrolla la vida cotidiana están llenos de interferencias y reverberación por lo tanto GSC no es viable.

Para las condiciones utilizadas en este trabajo GSC no realizó el proceso de filtrado de una manera satisfactoria. Sin embargo, considerando el caso de *Noisy-4source* que simula un ambiente real con tres fuentes de interferencia, el resultado del SIR obtenido fue de 2 dB, este valor dista mucho de 20 dB pero se asemeja al resultado descrito en *Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement [41]***.** En dicho trabajo la Relación Señal Interferencia es de un 1 dB y se utiliza una μ estática. En esta tesina se utilizó una μ dinámica y el resultado del SIR fue dos veces más grande lo cual indica que es posible tener un mejor desempeño de GSC si aparte de considerar la señal de interés también se toma en cuenta la reverberación.

Es importante mencionar que en este trabajo sólo se hicieron pruebas con un solo par de valores para μ_0 y μ_{max} **.** Por lo tanto cabría la posibilidad de para otros pares de valores (más grandes o más pequeños) GSC podría tener mejores o peores resultados.

Una de las aportaciones importantes de este trabajo fue demostrar que el GSC tradicional no es apto para ser implementado en los auxiliares auditivos. Por lo tanto, los investigadores interesados en mejorar la calidad de vida de las pacientes con discapacidad auditiva pueden inclinarse por otras técnicas de filtrado espacial o considerar los resultados obtenidos en esta tesina para mejorar GSC.

Como trabajo a futuro se puede considerar que hay otras características que se podrían cambiar en las herramientas que se utilizaron para realizar las pruebas, como el número de micrófonos y el tipo de micrófonos. Actualmente existen los micrófonos MEMS (Sistemas Micro Electromecánicos) los cuales ocupan un espacio físico pequeño mientras conceden una alta sensibilidad (otorgando altas razones señal a ruido, SNR), por lo que su principal uso es en

dispositivos pequeños, como teléfonos móviles. Una de las características de estos micrófonos es que la tecnología con la que están hechos (CMOS) permite que en el mismo dispositivo convivan simultáneamente electrónica analogía y digital. El desarrollo de un sistema de adquisición de audio basado en micrófonos MEMS tiene como ventaja contar con micrófonos pequeños que pueden colocarse en una superficie estratégica, por ejemplo dentro de un auxiliar auditivo, sin que la distancia que hay entre los oídos de una persona perturben notoriamente el funcionamiento de los mismos [42].

Otro de los aspectos importantes que se deberían considerar es cuando la fuente de interés no esté frente al usuario de auxiliares auditivos. Dependiendo de la posición de la fuente de interés, cuando la señal de dicha fuente llega a los oídos del consumidor, ciertos intervalos de frecuencia se amplifican mientras que otros se atenúan. Una herramienta que se utiliza para el procesamiento de la señal de audio en la situación antes descrita es la Función de Transferencia Relacionada con la Cabeza (HRTF por sus siglas en inglés). El trabajo de HRTF es registrar las transformaciones (difracción y reflexión) de una onda sonora desde que se propaga de la fuente hasta nuestros oídos, esto lo hace incluyendo la presencia de una cabeza en medio de los oídos, así como las reflexiones ambientales. La forma en la que opera HRTF es generando una señal objetivo desde una cierta posición y se mide la salida a la entrada de los canales de la oreja izquierda y derecha. Estas pruebas se hacen en una habitación anecóica y se utiliza una réplica de una cabeza humana. Las ondas sonoras que llegan a los oídos de la cabeza de prueba presentan alteraciones muy similares en el trayecto hacia los canales auditivos, como si estuvieran llegando a la cabeza de una persona [43].

También sería conveniente considerar una versión más robusta de GSC ya que la que implementamos en esta tesina no resultó conveniente para ser implementada en los auxiliares auditivos por su sensibilidad a la reverberación. La diferencia que existe entre un GSC tradicional y un GSC robusto es que en esta nueva versión no sólo se trata de minimizar las señales de las fuentes de interferencia, sino que también se examina como disminuir la reverberación.

En cuanto al procesamiento de señales de audio hay mucho por hacer, la separación de fuentes es uno de los objetivos primordiales, dicho objetivo es fundamental cuando los resultados que se obtengan son para ayudar a personas con discapacidad auditiva.

Anexo

Código de Cancelación Generalizada de Lóbulos Laterales (GSC)

```
#include <stdio.h>
#include <unistd.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>
#include <jack/jack.h>
#include <sndfile.h>

jack_port_t *input_port1;
jack_port_t *input_port2;
jack_port_t *output_portA;
jack_port_t *output_portB;


jack_default_audio_sample_t gsc_buffA;
jack_default_audio_sample_t *gsc_buffB;
jack_default_audio_sample_t gsc_buffC;
jack_default_audio_sample_t *g;
jack_default_audio_sample_t *gsc_o;


SNDFILE * audio_file;
SF_INFO audio_info;

float r;
int d=0;
int Nw=128;
float mu = 0.0075;

float mu0 = 0.001;
float mu_max = 0.01;

int jack_callback (jack_nframes_t nframes, void *arg){
     int i,K,write_count;

     jack_default_audio_sample_t *outA = (jack_default_audio_sample_t *)
jack_port_get_buffer (output_portA, nframes);
     jack_default_audio_sample_t *outB = (jack_default_audio_sample_t *)
jack_port_get_buffer (output_portB, nframes);

     jack_default_audio_sample_t *in1 = (jack_default_audio_sample_t *)
jack_port_get_buffer (input_port1, nframes);
     jack_default_audio_sample_t *in2 = (jack_default_audio_sample_t *)
jack_port_get_buffer (input_port2, nframes);


     float p_buffB = 0;
```

```
float p_o = 0;

r = 0;

for (K=0; K<nframes; K++) {
      //Delay and sum
      gsc_buffA=(in1[K] + in2[K])/2;

      //Matriz de bloqueo
      for(i=1;i<Nw;i++)
            gsc_buffB[i-1] = gsc_buffB[i];
      gsc_buffB[Nw-1] = (in1[K] - in2[K]);

      gsc_buffC=0;
      for(i=0;i<Nw;i++)
            gsc_buffC += g[i]*gsc_buffB[i];

      outA[K] = gsc_buffA - gsc_buffC;
      outB[K] = outA[K];


      //refrescar o_buff
      for(i=1;i<Nw;i++)
            gsc_o[i-1] = gsc_o[i];
      gsc_o[Nw-1] = outA[K];

      //calcular potencias
      p_o = 0;
      p_buffB = 0;
      for(i=0;i<Nw;i++){
            p_o += gsc_o[i]*gsc_o[i];
            p_buffB += gsc_buffB[i]*gsc_buffB[i];
      }


      if(mu0*p_buffB/p_o < mu_max){
            r += mu0*outA[K]/p_o;
      }else{
            r += mu0*outA[K]/p_buffB;
      }

      for(i=0;i<Nw;i++){

            if(mu0*p_buffB/p_o < mu_max){
                  g[i] += mu0*outA[K]*gsc_buffB[i]/p_o;

            }else{
                  g[i] += mu0*outA[K]*gsc_buffB[i]/p_buffB;
            }
            if(isnan(g[i]))
                  g[i] = 0.0;
      }
```

```
      }

      printf ("%f ",r/nframes);


      write_count = sf_write_float(audio_file,(float *)outA,nframes);

      return 0;
}



void jack_shutdown (void *arg) {
    exit (1);
}

int main (int argc, char argv[]) {
    jack_client_t *client;
    const char **port_names;
    const char *client_name = "Beamforming";
    jack_options_t options = JackNoStartServer;
    jack_status_t status;

    client = jack_client_open (client_name, options, &status);

    if (client == NULL) {
        fprintf (stderr, "jack server not running?\n");
        return 1;
    }

    jack_set_process_callback (client, jack_callback, 0);

    jack_on_shutdown (client, jack_shutdown, 0);

    printf ("engine sample rate: %" PRIu32 "\n", jack_get_sample_rate
(client));

    audio_info.samplerate = jack_get_sample_rate (client);
    audio_info.channels = 1;
    audio_info.format =  SF_FORMAT_WAV | SF_FORMAT_PCM_16;
    audio_file = sf_open ("audio.wav",SFM_WRITE,&audio_info);

    gsc_buffB= (jack_default_audio_sample_t *) malloc (Nw * sizeof
(jack_default_audio_sample_t ));
    g= (jack_default_audio_sample_t *) malloc (Nw * sizeof
(jack_default_audio_sample_t));
    gsc_o= (jack_default_audio_sample_t *) malloc (Nw * sizeof
(jack_default_audio_sample_t));
     int i;
    for(i = 0;i<Nw;i++){
      g[i] = 0.0;
      gsc_o[i] = 0.0;
```

```
    }
    input_port1= jack_port_register (client, "input1",
JACK_DEFAULT_AUDIO_TYPE, JackPortIsInput, 0);
    input_port2= jack_port_register (client, "input2",
JACK_DEFAULT_AUDIO_TYPE, JackPortIsInput, 0);
    output_portA = jack_port_register (client, "outputA",
JACK_DEFAULT_AUDIO_TYPE, JackPortIsOutput, 0);
    output_portB = jack_port_register (client, "outputB",
JACK_DEFAULT_AUDIO_TYPE, JackPortIsOutput, 0);

    if (jack_activate (client)) {
        fprintf (stderr, "cannot activate client");
        return 1;
    }
    printf ("client running\n");


    if ((port_names = jack_get_ports (client, NULL, NULL, JackPortIsPhysical|
JackPortIsOutput)) == NULL) {
        fprintf(stderr, "Cannot find any physical capture ports\n");
        exit(1);
    }
/*
    if (jack_connect (client, port_names[0], jack_port_name (input_port1))) {
        fprintf (stderr, "cannot connect input ports\n");
    }
    if (jack_connect (client, port_names[1], jack_port_name (input_port2))) {
        fprintf (stderr, "cannot connect input ports\n");
    }
    free (port_names);
*/
    if ((port_names = jack_get_ports (client, NULL, NULL, JackPortIsPhysical|
JackPortIsInput)) == NULL) {
        fprintf(stderr, "Cannot find any physical playback ports\n");
        exit(1);
    }
    if (jack_connect (client, jack_port_name (output_portA), port_names[0])) {
        fprintf (stderr, "cannot connect output ports\n");
    }
    if (jack_connect (client, jack_port_name (output_portB), port_names[1])) {
        fprintf (stderr, "cannot connect output ports\n");
    }
    free (port_names);

    sleep (-1);

    jack_client_close (client);
    exit (0);
 }
```

Referencias

[1] Sordera y audición. Organización Mundial de la Salud. Centro de prensa. En http://www.who.int/mediacentre/factsheets/fs300/es/

[2] INEGI. (2015). “Estadísticas a propósito del día internacional de las personas con discapacidad (3 de diciembre)”.

[3] INEGI. Encuesta Nacional de la Dinámica Demográfica 2014. Base de datos.

[4] Kochkin Sergei. (Febrero 2000). “MarkeTrak V:"Why my hearing aids are in the drawer": The consumers’ perspective”. The Hearing Journal, (53).

[5] Cristiani, Horacio. Historia del audífono. Mutualidad Argentina de Hipoacúsicos. En https://www.mah.org.ar/historia-de-los-audifonos.

[6] Viviendo con un instrumento auditivo. Información básica sobre instrumentos auditivos. En https://lat.bestsoundtechnology.com/basic-information/living-with/.

[7] Moreno, Cynthya. (2008). “Diseño y construcción de un auxiliar auditivo con características digitales”

[8] Material y partes importantes de los Auiliares Auditivos. Aparato Auditivo. En https://aparatoauditivo.com.mx/blog/noticias/170-material-y-partes-importantes-de-los-aparatos-auditivos .

[9] Gonzalez Oreste. Otorrinolaringología. Infomed, En http://articulos.sld.cu/otorrino/?tag=protesis-auditiva.

[10] Ergo Compact Power. Información audiológica.

[11] Kochkin Sergei. (Septiembre 2005). “MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age”. The Hearing Journal, (58)

[12] Kochkin Sergei. (Enero 2010). “MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing”. The Hearing Journal, (63)

[13] Van Vliet, Dennis. (Febrero 2012) “When Technology Bites Back”. The Hearing Journal, (65)

[14] Procesamiento digital de señales. En https://es.wikipedia.org/wiki/Procesamiento_digital_de_se%C3%B1ales

[15] Suárez, Luis. (2005). “Visor de señales biomédicas para el fisiógrafo mk-iii-p de narco-scientific biosystems division” Vifibio

[16] Ondas periódicas. Estudiar física. En https://estudiarfisica.com/category/clase-fisica-general/

[17] López, Daniel. Ingeniería de sonido. Ediciones de la U, Colombia, 2011.

[18] Tribaldos, Clemente. Sonido Profesional. Editorial Paraninfo, Madrid. 1993.

[19] Revuelta, Sara. (2016) "Aplicación de las divergencias Alfa-Beta a la Separación Ciega de Fuentes en mezclas de audio"

[20] Broesch, James; Straneneby, Dag; Walker, William. Digital Signal Processing:Instant Access. Elsevier, USA. 2009

[21] Separación Ciega de Fuentes basada en la estructura temporal de las señales de voz. En http://bibing.us.es/proyectos/abreproy/11088/fichero/Proyecto+Fin+de+Carrera%252F8.pdf

[22] Martín, Tomás. (2015) "Transformada de Fourier - Representación de señales de sonido"

[23] ¿Qué es el sonido? Audio digital. En http://informaticaytecnologias.com/audio-digital/

[24] Comparación de audio analógico y digital. Audio digital. En http://cmm.cenart.gob.mx/tallerdeaudio/cursos/cursoardour/Teoria_y_tecnicas/Audiodigital.html

[25] Cuantificación digital. En https://es.wikipedia.org/wiki/Cuantificaci%C3%B3n_digital

[26] Stillo, Gonzalo. "Técnicas adaptativas de filtrado espacial: Beamformers"

[27] Del Valle, Carolina. (2001). "Algoritmos de búsqueda no lineal para diseñar arreglos direccionales de geófonos"

[28] Martínez, David. (2008). "Técnicas de procesado en array para realzado de voz en situaciones adversas"

[29] Dan Li, Qinye Yin, Pengcheng Mu, Wei Guo (2011). "Robust MVDR Beamforming Using the DOA Matrix Decomposition"

[30] S.S, Balasem; Tiong, S.K. (2012). "Beamforming Algorithms Technique by Using MVDR and LCMV"

[31] Pepe, Marcelo. (2004). "Filtrado espacial adaptivo"

[32] Benesty, Jacob; Sondhi, Mohan; Huang, Yiteng. Springer Handbook of Speech Processing. Springer. Berlín. 2008

[33] Sánchez, José. (2004). “Mejora de la señal de voz en condiciones acústicas adversas mediante arrays de micrófonos”

[34] Rascón, Caleb. Audición Robótica. Separación de Fuentes en Línea. Separación de Fuentes por Beamforming. En http://calebrascon.info/AR/Topic7/07.2-Beamforming.pdf

[35] JACK Audio Connection Kit. En https://es.wikipedia.org/wiki/JACK_Audio_Connection_Kit

[36] JACK Audio Connection Kit. En http://www.jackaudio.org/api/

[37] Rascón, Caleb.JACK Audio Connnection Toolkit. Instalación, Configuración y Creación de Agentes de JACK. Audición Robótica. En http://calebrascon.info/AR/

[38] Rascon C., Fuentes G., Meza I. V. (2015).“Lightweight multi-DOA tracking of mobile speech sources”

[39] Vincent, Emmanuel; Gribonval, Rémi; Févotte, Cédric. (2006) “Performance Measurement in Blind Audio Source Separation”

[40] Rascón, Caleb. Audición Robótica. Procesamiento de Varias Señales Concurrentes. Corpus AIRA. En http://calebrascon.info/AR/

[41] Bitzer, Joerg; Uwe, Klaus; Kammeyer, Karla-Dirk. (1999). “Theoretical noise reduction limits of the generelized sidelobe canceller (GSC) for speech enhancement”

[42] Lunati, Valentin; Podlubn, Ariel; González, Fernando. (2013). “Micrófonos mems: análisis y aplicaciones en audición binaural”

[43] Potisk, Tilen. (2015). “Head-Related Transfer Function”

[44] Wang, Xiaofei; Guo, Yanmeng; Yan, Yonghong. (2015) “A reverberation robust target speech detection method using dual-microphone in distant-talking scene”

[45] Audífonos, el precio de perder oído. Tecnología médica para mejorar la audición. En http://revista.consumer.es/web/es/20090401/salud/74682.php

[46] Haykin, Simon; Van Veen, Barry. Señales y sistemas. Limusa, México. 2008.

[47] Bourgeois, Julien. Minker, Wolfgang (2009). Time-Domain Beamforming and Blind Source Separation. Speech Input in the Car Environment