



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA

**Aplicación de tecnologías de sonido binaural para el mejoramiento
de la calidad acústica en sistemas de comunicación móvil.**

T E S I S

Que para obtener el título de:

**DOCTORA EN INGENIERÍA ELÉCTRICA
INSTRUMENTACIÓN**

P R E S E N T A

ANA LAURA PADILLA ORTÍZ

TUTOR:

DR. FELIPE ORDUÑA BUSTAMANTE

2012



**CENTRO DE CIENCIAS APLICADAS Y
DESARROLLO TECNOLÓGICO**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO

PRESIDENTE: DR. GERARDO EUGENIO SIERRA MARTÍNEZ

SECRETARIO: DR. JOSÉ ABEL HERRERA CAMACHO

VOCAL: FELIPE ORDUÑA BUSTAMANTE

1er. SUPLENTE: RAFAEL DE LA GUARDIA GONZÁLEZ

2º. SUPLENTE: SANTIAGO JESÚS PÉREZ RUIZ

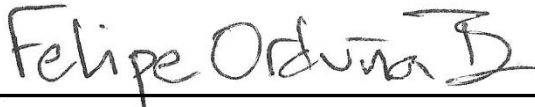
LUGAR DONDE SE REALIZÓ LA TESIS:

**CENTRO DE CIENCIAS APLICADAS Y DESARROLLO TECNOLÓGICO,
UNAM, CIUDAD UNIVERSITARIA.**

MÉXICO, DISTRITO FEDERAL.

TUTOR DE TESIS:

DR. FELIPE ORDUÑA BUSTAMANTE



FIRMA

ÍNDICE

1	INTRODUCCIÓN	1
1.1	DESCRIPCIÓN DEL TEMA DE TESIS	1
1.2	OBJETIVO	3
1.3	ESTRUCTURA DE LA TESIS	5
2	ALGUNOS CONCEPTOS SOBRE AUDICIÓN BIAURAL	6
2.1	TECNOLOGÍA BIAURAL	6
2.1	REFERENCIA GEOMÉTRICA PARA LA DESCRIPCIÓN DEL SONIDO BIAURAL	7
2.2	TEORÍA DUPLEX	8
2.2.1	<i>Diferencias de tiempo interaural</i>	9
2.2.2	<i>Efecto de precedencia o efecto Haas</i>	10
2.2.3	<i>Diferencias de nivel interaural</i>	10
2.2.4	<i>Cono de confusión</i>	11
2.3	FUNCIONES DE TRANSFERENCIA ACÚSTICA REFERIDAS A LA CABEZA (HRTF)	12
2.4	ESPACIALIDAD DEL SONIDO	15
2.5	CORRELACIÓN INTERAURAL (IACC)	15
2.6	EFFECTO DE FIESTA COCTEL O <i>COCKTAIL PARTY EFFECT</i>	16
3	REVISIÓN DE ANTECEDENTES	17
3.1	INTELIGIBILIDAD DE LA PALABRA.	17
3.2	EVALUACIÓN DE LA INTELIGIBILIDAD	18
3.2.1	<i>Métodos objetivos</i>	18
3.2.2	<i>Métodos subjetivos</i>	21
3.2.3	<i>Material de voz en español</i>	22
3.3	PRESENTACIÓN MONOAURAL Y BIAURAL DE LAS SEÑALES	23
3.4	EFFECTO DEL RUIDO Y LA REVERBERACIÓN EN LA INTELIGIBILIDAD	24
3.6	SEPARACIÓN ESPACIAL DE LAS FUENTES	27
3.7	EFFECTO DE FIESTA COCTEL (<i>COCKTAIL PARTY EFFECT</i>)	27
4	PRUEBAS DE INTELIGIBILIDAD	29
4.1	PRUEBA 1: PRUEBAS DE INTELIGIBILIDAD EN PRESENTACIÓN MONOAURAL, MONOFÓNICA, BIAURAL A 0° Y BIAURAL A ±30° CON RUIDO Y REVERBERACIÓN NO CORRELACIONADOS INTERAURALMENTE.	30
4.1.2	<i>Objetivo</i>	30
4.1.3	<i>Participantes</i>	30
4.1.4	<i>Material de voz</i>	31
4.1.3.1	<i>Grabación del material de voz</i>	31
4.1.3.2	<i>Grabación binaural del material de voz</i>	32
4.1.4	<i>Procesamiento del material de voz</i>	33
4.1.5	<i>Contaminación de las señales de voz</i>	35
4.1.5.1	<i>Señal de voz con ruido</i>	35
4.1.5.2	<i>Señal de voz con reverberación</i>	35
4.1.6	<i>Presentación de las señales</i>	37
4.1.7	<i>Orden de las listas de palabras</i>	37
4.1.8	<i>Evaluación de la inteligibilidad</i>	39
4.1.9	<i>Desarrollo de la prueba</i>	39
4.1.10	<i>Resultados</i>	40

4.1.10.1	<i>Análisis estadístico</i>	42
4.1.10.2	<i>Intervalos de confianza</i>	42
4.1.10.3	<i>Pruebas de hipótesis</i>	43
4.1.11	<i>Conclusiones</i>	44
4.2	PRUEBA 2: PRUEBAS DE INTELIGIBILIDAD CON RUIDO Y REVERBERACIÓN CORRELACIONADOS.	45
4.2.1	<i>Objetivo</i>	45
4.2.2	<i>Participantes</i>	45
4.2.3	<i>Procesamiento de las señales biaurales</i>	45
4.2.4	<i>Desarrollo de la prueba</i>	46
4.2.5	<i>Resultados</i>	46
4.2.5.1	<i>Comparación entre señales correlacionadas y no correlacionadas</i>	47
4.2.6	<i>Análisis de los resultados</i>	49
4.2.7	<i>Conclusiones</i>	50
4.3	PRUEBA 3: PRUEBAS DE INTELIGIBILIDAD A DIFERENTES ÁNGULOS ACIMUTALES	50
4.3.1	<i>Objetivo</i>	50
4.3.2	<i>Participantes</i>	50
4.3.3	<i>Grabación binaural de la señal de voz a diferentes ángulos en el plano horizontal</i>	50
4.3.4	<i>Procesamiento de las señales</i>	51
4.3.5	<i>Desarrollo de la prueba</i>	53
4.3.6	<i>Resultados</i>	53
4.3.6.1	<i>Localización de fuentes</i>	55
4.3.7	<i>Análisis de los resultados</i>	55
4.3.8	<i>Conclusiones</i>	56
4.4	PRUEBA 4: PRUEBA DE INTELIGIBILIDAD CON HRTF PERSONALIZADAS Y NO PERSONALIZADAS	57
4.4.1	<i>Objetivo</i>	57
4.4.2	<i>HRTF individualizadas</i>	57
4.4.2.1	<i>Medición de las HRTF</i>	57
4.4.2.2	<i>Participantes</i>	58
4.4.2.3	<i>Procesamiento de las señales de voz</i>	58
4.4.2.4	<i>Desarrollo de la prueba</i>	59
4.4.3	<i>HRTF no individualizadas</i>	60
4.4.3.1	<i>Participantes</i>	61
4.4.3.2	<i>Desarrollo de la prueba</i>	61
4.4.3.3	<i>Resultados</i>	61
4.4.3.3.1	<i>HRTF individualizadas vs. HRTF no individualizadas bajo condiciones de ruido</i>	65
4.4.3.3.2	<i>HRTF individualizadas vs. HRTF no individualizadas bajo condiciones de reverberación</i>	66
4.4.3.3.3	<i>Intervalos de confianza</i>	67
4.4.4	<i>Análisis de los resultados</i>	68
4.4.5	<i>Conclusiones</i>	68
5	CONCLUSIONES	69
	REFERENCIAS	73
	APÉNDICE A	84
	APÉNDICE B	85

AGRADECIMIENTOS

- ❑ A la Universidad Nacional Autónoma de México por abrirme las puertas para realizar mis estudios de posgrado, sin duda de las mejores experiencia de mi vida.
- ❑ A mis padres y mi hermana, Raquel, Guillermo y Raquelita por su amor y apoyo incondicional a lo largo de todos estos años.
- ❑ Al Dr. Felipe Orduña Bustamante, quien con el paso del tiempo no solo ha sido mi guía sino un valiosísimo amigo. Gracias por la enorme paciencia, por el apoyo académico pero principalmente por el apoyo anímico en los momentos complicados.
- ❑ A Arturo Machuca por su ayuda, por compartir conmigo sus conocimientos, y por sus palabras de aliento en todo momento.
- ❑ A la familia Gutiérrez Argüello por su cercanía a pesar de la distancia, por compartir conmigo su crecimiento no sólo profesional sino también sus logros personales.
- ❑ A los profesores del laboratorio: Dra. Rosalba Castañeda, Dr. Felipe Orduña, Dr. Jesús Pérez, M.I. Antonio Pérez, Dr. Pablo Rendón por compartir sus conocimientos dentro y fuera de clases, especialmente por brindarme su amistad.
- ❑ A los amigos del laboratorio: Arturo, Edgar, Héctor, Roberto y Alejandro por su amistad y por hacer amenas las horas en el laboratorio.
- ❑ A mis sinodales por invertir su tiempo en la revisión de esta tesis y sus útiles recomendaciones para mejorar este trabajo.
- ❑ A la Coordinación de Estudios de Posgrado de la UNAM por la ayuda para la realización de mis estudios.

1 Introducción

1.1 Descripción del tema de tesis

Esta tesis estudia los posibles beneficios de la audición binaural en sistemas de comunicación móvil; particularmente en términos de mejorar la inteligibilidad de la voz ante condiciones acústicas adversas en presencia de ruido y reverberación.

Esta investigación doctoral forma parte de un proyecto de investigación de mayor alcance titulado: “*Binaural sound technologies for mobile communication devices*” (Tecnologías de sonido binaural para dispositivos móviles de comunicación), desarrollado en el Grupo de Acústica y Vibraciones del CCADET-UNAM en colaboración con Intel Tecnología de México S.A. de C.V., cuyo objetivo general es aplicar técnicas de reproducción de sonido binaural en dispositivos móviles para mejorar la inteligibilidad de la voz, a través del uso de HRTF (*Head Related Transfer Functions*, o funciones de transferencia acústica referidas a la cabeza) individualizadas para cada usuario mediante procesamiento de imágenes antropométricas. Los resultados obtenidos específicamente en la presente investigación doctoral aportan información útil sobre el efecto que tienen las técnicas de reproducción de sonido binaural en la inteligibilidad de la voz, considerando diferentes aspectos, tales como: condiciones adversas de ruido y reverberación, la correlación interaural entre la señal de voz y el disturbio acústico, así como la influencia del ángulo de incidencia de la señal de voz y el uso de HRTF individualizadas.

El concepto de audición binaural se refiere a la acción de escuchar con ambos oídos, esto representa una ventaja importante sobre el sonido monoaural, ya que permite la percepción espacial acústica de nuestro entorno. Debido a que los oídos están separados físicamente por la cabeza, la señal que recibe un oído generalmente es distinta a la que recibe el otro. Esta diferencia nos permite localizar fuentes sonoras, así como percibir y separar la voz de un orador particular, del ruido de fondo o de muchas conversaciones simultáneas del ambiente que nos rodea, esto se conoce como el efecto de fiesta coctel, o en inglés: *cocktail party effect*.

La aplicación de la audición binaural se denomina “tecnología binaural” y tiene ventajas significativas en cada una de sus aplicaciones, una de ellas es en la aviación. Puesto que el piloto está expuesto a múltiples señales las cuales se originan desde puntos muy diversos, como las provenientes de la

torre de control o bien de otra aeronave, así como señales de voz y/o de alerta provenientes del sistema de audio instalado en la cabina de mando [Begault y Wenzel, 1992], esto sin considerar el ruido en el que se encuentra inmerso (el ruido producido por la misma aeronave o la interacción aerodinámica entre el aire y el fuselaje del avión, las alas, etc.). Por lo anterior, es de vital importancia que el piloto cuente con un sistema que permita mejorar la inteligibilidad de la voz en presencia de ruido de fondo así como ayudar a separar una señal deseada en presencia de múltiples fuentes sonoras, esto ayudaría de manera significativa a minimizar errores humanos. Por ejemplo, Hughes [Hughes, 1989] reportó que al menos 65% de los accidentes aéreos reportados en el período de 1977 a 1987 fueron ocasionados por errores humanos.

Otra aplicación es en equipos diseñados para personas invidentes: se equipa al sujeto con una brújula electrónica y un transmisor receptor que se comunica con los satélites de posicionamiento global GPS. Las señales de los satélites determinan la posición del invidente en un radio de 1 metro en un mapa programado en una computadora que indica la localización de los objetos (por ejemplo, casetas telefónicas, paradas de autobuses, etc.) del lugar. Con esta información, la computadora genera mensajes sobre la ubicación de dichos objetos, que el invidente escucha a través de los audífonos. Los mensajes se presentan de modo que el volumen y la sincronización de los sonidos que llegan a los oídos corresponden a la forma en que serían percibidos si provinieran de los objetos [Loomis *et al*, 1990; Loomis *et al*, 1994; Loomis *et al*, 1998].

Otra aplicación importante y novedosa de la tecnología binaural es en los dispositivos móviles, tales como teléfonos celulares ó computadoras de mano (PDA, por las siglas en inglés de *Personal Digital Assistant*) donde el objetivo es mejorar la experiencia del usuario [AES Staff, 2006]. Una aplicación más es en los sistemas de entretenimiento, como videojuegos, en los cuales la principal aportación es mostrar ambientes virtuales acústicos más reales, donde sea posible escuchar un sonido en movimiento y además ubicar la dirección de un sonido determinado [Hamidi y Kapralos, 2009].

De reciente desarrollo, se encuentran las teleconferencias, donde la principal ventaja consiste en mejorar la inteligibilidad de la voz y generar una experiencia auditiva más real y al mismo tiempo menos estresante [Harikrishnan *et al.*, 2011], así como videoconferencias, en donde es posible interactuar con más de dos participantes al mismo tiempo [Kanada, 2004, 2005].

Existe una investigación doctoral que complementa este trabajo de tesis, titulada “*Fotoantropometría¹ automatizada para personalizar funciones de transferencia (HRTF) con aplicaciones en sonido binaural de dispositivos móviles de audio y comunicación*”, mismo que se desarrolla actualmente en el CCADET-UNAM por el alumno Edgar Augusto Torres Gallegos (Doctorado en Ingeniería Eléctrica, ingreso 2010-1). En dicho trabajo se propone generar o relacionar las funciones de transferencia acústica referidas a la cabeza (HRTF) individuales a partir de la antropometría del usuario del dispositivo móvil. Esta determinación individualizada de las HRTF ayudará a alcanzar el objetivo práctico de la reproducción binaural del sonido para los usuarios de dispositivos móviles. Por otra parte, la presente investigación doctoral se enfoca al objetivo complementario de estudiar los posibles beneficios de la reproducción binaural en la inteligibilidad de la voz ante condiciones acústicas adversas en presencia de ruido y reverberación.

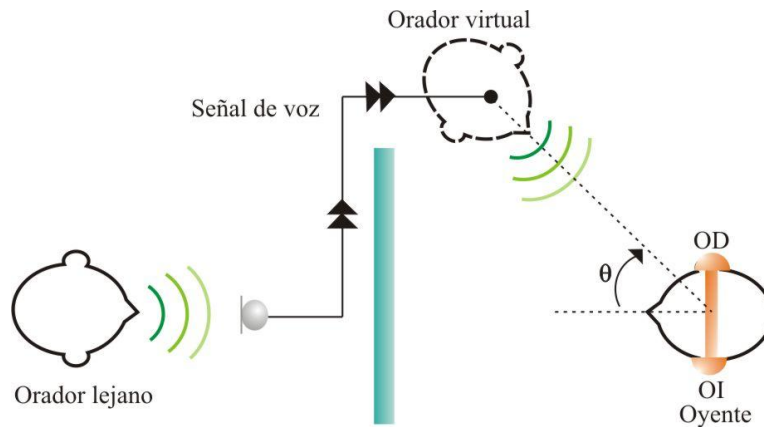
1.2 Objetivo

Los objetivos que persigue esta investigación se enuncian a continuación:

- Desarrollar criterios y métodos de evaluación de la calidad acústica en sistemas de comunicación móviles, considerando aspectos de inteligibilidad de la voz, índices de calidad acústica, inmunidad al ruido, a la reverberación, etc.
- Estudiar el impacto de reproducción binaural de sonido espacial a través de audífonos para aplicaciones en dispositivos móviles de comunicación y entretenimiento.

Lo anterior se ilustra en la Figura 1-1:

¹ Relativo a la medición de las dimensiones y rasgos de las personas sobre una fotografía.



**Figura 1-1: Sistema de comunicación móvil biaural. El orador que está ubicado en algún punto remoto, emite un mensaje, el oyente que está físicamente separado del orador tiene la habilidad de ubicar al orador de manera virtual, en la posición donde recibe de forma más eficiente el mensaje.
OD: Oído Derecho, OI: Oído Izquierdo.**

En la presente tesis se realizaron diferentes pruebas de inteligibilidad, cada una con un objetivo particular. La primera prueba permitió cuantificar la ventaja que se obtiene al emplear técnicas de grabación y/o reproducción biaural, comparado con una señal monoaural o bien monofónica, considerando también ambientes acústicos desfavorables, en presencia de ruido y reverberación. La segunda prueba estudió el efecto que tiene la correlación interaural (IACC) en la inteligibilidad, por lo que considera señales idénticas de ruido y reverberación para ambos oídos y se compararon con las pruebas realizadas previamente. La tercera prueba tuvo como objetivo evaluar la inteligibilidad a diferentes ángulos acimutales (plano horizontal), para identificar en qué ángulos se recibe de forma más efectiva el mensaje. En la cuarta y última prueba se midieron funciones de transferencia referidas a la cabeza (HRTF) para cuantificar el beneficio aportado por la reproducción de sonido biaural a través de HRTF personalizadas en la inteligibilidad. Todas las pruebas se realizaron a través de audífonos. En general, los resultados muestran que la inteligibilidad se ve más afectada por la reverberación que por el ruido, además de obtener una mejor inteligibilidad bajo condiciones de ruido interauralmente correlacionado; mientras que la inteligibilidad más baja se observa en condiciones de reverberación interauralmente correlacionada.

Los resultados muestran que las técnicas de reproducción biaural tienen ventaja significativa sobre técnicas de reproducción monoaural, además que la correlación interaural y el ángulo de incidencia en el plano horizontal (ángulo acimutal) tiene una clara influencia en la inteligibilidad de la voz. De manera general, los resultados presentados en esta tesis permiten concluir que la escucha lateralizada de un orador, real o virtual, tiende a mejorar la inteligibilidad de la voz.

Durante el desarrollo de esta tesis se realizaron algunas estancias de investigación en el Systems Research Center-México de Intel Corporation, en Guadalajara, Jal., por lo que algunas grabaciones y pruebas de inteligibilidad se llevaron a cabo en el Laboratorio de Acústica de Intel SRC-M.

1.3 Estructura de la tesis

Esta tesis consta de cinco capítulos los cuales están ordenados como se detalla a continuación. En la introducción se hace una descripción del tema, se establecen los objetivos generales y se describe el contenido de la tesis.

En el capítulo “Algunos conceptos sobre audición binaural” se describen los fundamentos conceptuales del sonido binaural, se explica la importancia de las funciones de transferencia referidas a la cabeza (HRTF) en la audición binaural así como la importancia del efecto fiesta cocktail.

En el siguiente capítulo “Revisión de antecedentes” se muestra una revisión de otros trabajos relacionados con las pruebas de inteligibilidad, indicando lo que han hecho otros investigadores y los resultados que han obtenido.

En el capítulo “Pruebas de inteligibilidad” se explica detalladamente cada una de las pruebas realizadas en la presente investigación, estableciendo los objetivos que persigue cada una ellas y mostrando los resultados que se obtuvieron.

En el último capítulo “Conclusiones” se plantean las conclusiones del proyecto, las aportaciones originales de este trabajo así como las recomendaciones para trabajo futuro en esta línea de investigación.

2 Algunos conceptos sobre audición biaural

En este capítulo se describen los fundamentos conceptuales del sonido biaural, la forma en que se divide para su análisis en diferentes elementos o componentes, sus formas de reproducción, así como las ventajas y desventajas que presentan cada una de ellas. Se explican también las claves o indicios perceptuales de las que hace uso el sistema auditivo humano para identificar la posición de una fuente sonora ubicada en algún punto, tal como se explica en la *teoría Duplex*.

También se detalla la importancia que tienen las funciones de transferencia referidas a la cabeza (HRTF) en el sonido biaural y se muestra un breve resumen de algunos estudios que se han realizado sobre las HRTF.

Por último, se habla del efecto fiesta coctel (*cocktail party effect*) que es muy común cuando las personas se encuentran en lugares públicos, expuestas a diferentes sonidos que provienen de muy diversos puntos, lo que hace que tengan que esforzarse en discriminar los sonidos que interfieren con la señal de interés para ellos.

2.1 Tecnología biaural

La tecnología biaural normalmente es utilizada para presentar un ambiente acústico virtual a un oyente. El principio de esta tecnología consiste en emular un campo sonoro real a través de la reproducción de dos señales, una para cada oído, con la finalidad de que el oyente tenga la experiencia auditiva como si estuviera físicamente presente en un campo sonoro [Møller, 1992; Takeuchi *et al.*, 2001].

Es posible que la reproducción de sonido biaural se realice por medio de altavoces o audífonos, con ventajas y desventajas en cada caso, las cuales se detallan a continuación. Las ventajas que presentan los altavoces es que existe una condición de audición libre a través de las orejas, además de una sensación de espacio acústico, es decir, exteriorización de fuentes virtuales; mientras que la desventaja que presenta la reproducción por este medio es que se genera el efecto de cruce de canales (*crosstalk*), en el cual la señal que es emitida por el altavoz derecho alcanza también al oído izquierdo y viceversa, por lo que la zona de reproducción biaural que se puede controlar es pequeña debido a que ligeros movimientos de la cabeza provocan señales erróneas [Takeuchi *et al.*, 2001].

En el caso de los audífonos la ventaja es que hay un mayor control e independencia de las señales que se reproducen en cada oído, de este modo se asegura que el sonido que llega a un oído sea sólo reproducido en ese oído. La desventaja es que se pierde *espacialización*, por lo que el sonido es escuchado como si estuviera en el interior de la cabeza. Puesto que hay forma de soslayar esta desventaja, la reproducción de sonido baural generalmente se realiza por medio de audífonos [Møller, 1992].

2.2 Referencia geométrica para la descripción del sonido baural

Como ya se explicó, el sonido baural consiste en la audición con ambos oídos, el cual emplean los seres humanos para la localización de fuentes. Para su análisis se hace uso de un sistema de coordenadas esféricas con referencia a una cabeza humana ubicada en el origen de coordenadas, por lo que se establecen tres planos característicos, como se muestra en la Figura 2-1 [Blauert, 1997]:

- Plano horizontal (o acimutal): define la distinción entre arriba y abajo.
- Plano medio: define la distinción entre derecha e izquierda.
- Plano frontal: define la distinción entre adelante y atrás.

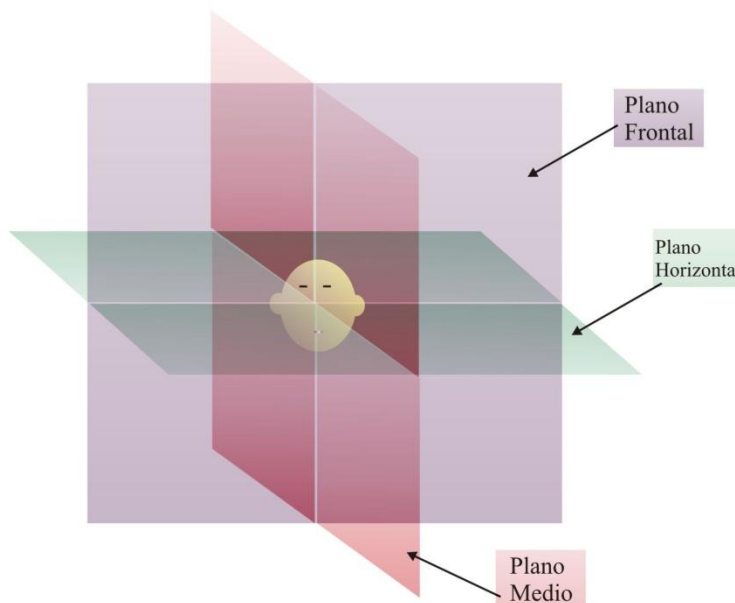


Figura 2-1: Referencias del sistema de coordenadas.

En la Figura 2-2 se observa una fuente sonora radiando en campo libre, las ondas generadas por la fuente llegan al oyente. La distancia del oyente a la fuente sonora se denota por la letra r , el ángulo

de incidencia es caracterizado por el ángulo acimutal θ y el ángulo de elevación se denota como ϕ . Los valores positivos de θ indican que la fuente está ubicada a la derecha del oyente. Los ángulos positivos de elevación denotan que la fuente está arriba del plano horizontal; mientras que para valores negativos la posición de la fuente es a la izquierda del escucha y por debajo del plano horizontal [Møller, 1992].

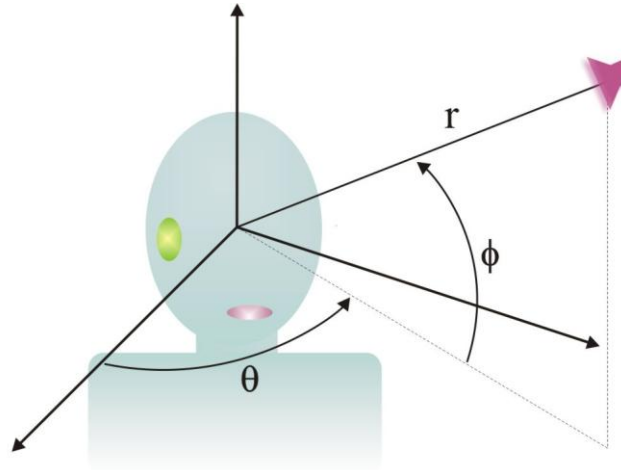


Figura 2-2: Convención de las variables que indican distancia (r), ángulo de incidencia (θ) y elevación (ϕ)[Møller, 1992].

2.3 Teoría Duplex

Para la localización lateral de una fuente sonora, el sistema auditivo emplea indicios perceptuales provenientes principalmente de las diferencias de intensidad interaural (ILD) y las diferencias de tiempo interaural (ITD) con que las ondas sonoras llegan a cada uno de los oídos. La teoría de que la localización de una fuente sonora está dada por las ITD y las ILD, fue propuesta por John William Strutt (Lord Rayleigh) en 1907 [Strutt, 1907].

Rayleigh observó que el sonido recibido en el oído opuesto al que se encuentra la fuente sonora, llamado *contralateral*, sería sombreado de forma muy pronunciada por la cabeza, dando lugar a una diferencia en el nivel del sonido que alcanzaba a ambos oídos. También observó que la diferencia en la intensidad sería insignificante para las frecuencias inferiores a 1000 Hz, dado que la longitud de onda es similar o mayor al diámetro de la cabeza [Strutt, 1907].

2.3.1 Diferencias de tiempo interaural

Los sonidos que provienen desde diferentes puntos llegan a un oído antes que al otro. A esta falta de coincidencia temporal, se le denomina diferencias de tiempo interaural (ITD, Interaural Time Difference) [Bronkhorst y Plomp, 1988; Begault y Wenzel, 1992; Brungart y Rabinowitz, 1999; Cheng y Wakefield, 2001; Ramos *et al.*, 2007].

Los sonidos de baja frecuencia tienen una longitud de onda grande comparado con el diámetro de la cabeza, aproximadamente 18 cm, y se reciben en cada oído con diferencias de nivel muy pequeñas debido a la poca dispersión acústica que ocurre para bajas frecuencias, así que las diferencias interaurales de tiempo son perceptualmente más útiles por debajo de los 1000 Hz.

En la Figura 2-3 se observa que el sonido de la fuente sonora A (exactamente frente al oyente), llega al mismo tiempo a ambos oídos; mientras que en el caso de la fuente sonora B, que se localiza a un costado del oyente, el sonido llega primero al oído derecho que al izquierdo.

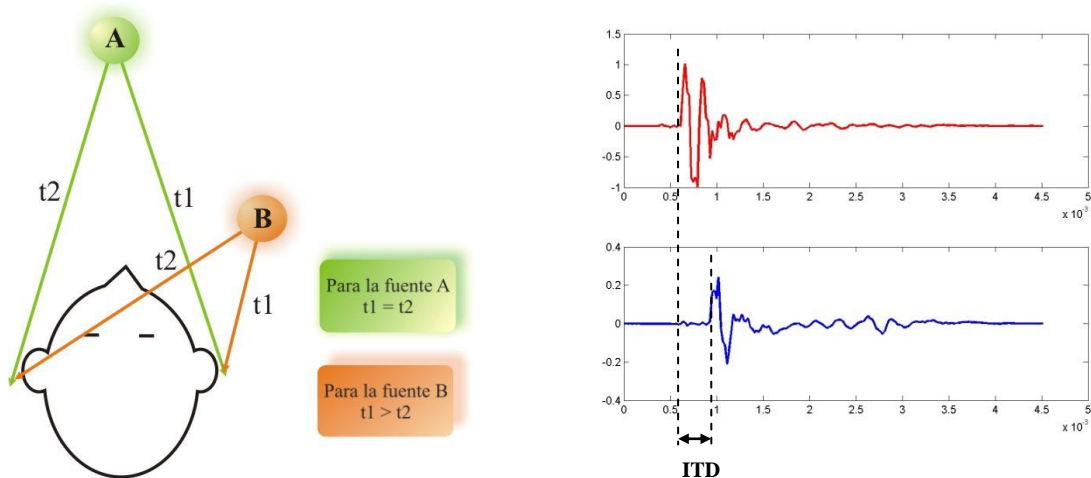


Figura 2-3: Diferencia de tiempo interaural (ITD). El sonido de la fuente sonora A llega al mismo tiempo a ambos oídos; mientras que en el caso de la fuente B el sonido llega primero al oído derecho que al izquierdo. Las respuestas acústicas que se muestran en la parte derecha corresponden respectivamente a las señales que llegan a los oídos derecho e izquierdo desde la fuente B.

Para medir la forma en que la ITD cambia según la ubicación de la fuente sonora, se colocan micrófonos en los oídos del oyente y se determina el tiempo de llegada de los estímulos sonoros provenientes de diversos sitios. Los resultados de estas mediciones indican que la ITD es cero cuando la fuente sonora se encuentra justo enfrente o detrás (o más en general, en el plano medio) del oyente [Bronkhorst y Plomp, 1988; Blauert, 1997; Brungart y Rabinowitz, 1999; Ramos *et al.*, 2007]; mientras que es máxima cuando está ubicada a un lado, frente a alguno de los oídos.

2.3.2 Efecto de precedencia o efecto Haas

Cuando recibimos dos sonidos de corta duración, provenientes de fuentes sonoras ubicadas espacialmente en puntos distintos, con una diferencia de tiempo muy corta, el sistema auditivo parece que elimina el segundo sonido y da precedencia al primero. Lo que genera que la localización de las fuentes sonoras dependa básicamente del primer sonido que se reciba. Esto se conoce como efecto de precedencia, efecto Haas o bien Ley del primer frente de onda [Blauert, 1971].

Para que el efecto de precedencia ocurra debe existir una diferencia mínima de 1 ms entre la llegada de uno y otro sonido. Si el sonido retardado es considerablemente más intenso que el primero (más de unos 10 ó 15 dB) el efecto de precedencia se cancela [Morfey, 2001].

2.3.3 Diferencias de nivel interaural

Las diferencias de nivel interaural (ILD, Interaural Level Difference), se definen como la diferencia en el nivel de presión sonora que llega a ambos oídos, la cual se debe a que la cabeza, debido al efecto de la dispersión acústica, crea una barrera que produce una sombra acústica, que atenúa los sonidos de altas que frecuencias llegan al oído que está más lejos [Strutt, 1907].

Las ondas de baja frecuencia (longitudes de onda grande) no son tan afectadas por la cabeza, mientras que las frecuencias altas (longitudes de onda pequeña) inciden en la cabeza creando una sombra acústica que atenúa el sonido transmitido del otro lado de la misma [Begault y Wenzel, 1992; Ramos *et al.*, 2007; Bronkhorst y Plomp, 1988; Brungart y Rabinowitz, 1999; Cheng y Wakefield, 2001; Hartmann y Constan, 2002].

Se ha medido este efecto de la frecuencia sobre la ILD por medio de dos micrófonos pequeños que registran la intensidad del sonido que llega a cada uno de los oídos en respuesta a una fuente sonora. Los resultados muestran que existe poca diferencia de nivel para frecuencias inferiores a 1000 Hz, pero es mucho mayor para frecuencias superiores [Brungart y Rabinowitz, 1999].

Para determinar experimentalmente la ILD se emplea la siguiente ecuación,

$$ILD = \left| 10 \log_{10} \left(\frac{\sum_n h_l^2[n]}{\sum_n h_r^2[n]} \right) \right| \quad [dB] \quad (2-1)$$

Donde h_l y h_r son la respuesta impulso medidas en el oído izquierdo y derecho, respectivamente.

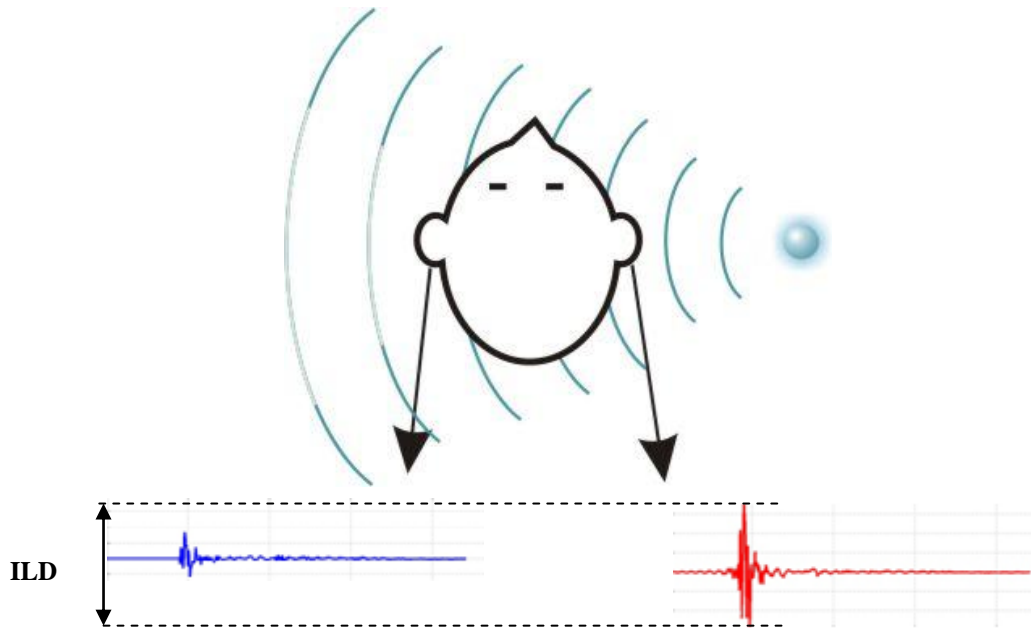


Figura 2-4: La señal emitida por la fuente sonora, ubicada a la derecha del oyente llega con mayor intensidad al oído derecho que es el más cercano; mientras que la señal que recibe el oído izquierdo es de menor intensidad debido a la sombra acústica provocada por la cabeza.

2.3.4 Cono de confusión

La teoría Duplex tiene algunas limitaciones, ya que sólo explica la localización de fuentes para diferentes ángulos en el plano horizontal o acimutal, es decir, si la fuente se encuentra a la derecha o a la izquierda del oyente [Brungart y Rabinowitz, 1999], no en el plano medio. Tampoco proporciona información alguna en cuanto al ángulo de elevación y la distancia de la fuente.

Cuando la fuente está ubicada en el plano medio las diferencias de nivel y de tiempo que se generan tienen un valor igual a cero o muy cercano a él, existen algunos casos en los cuales las diferencias interaurales son perceptualmente indistinguibles para distintas posiciones. Este fenómeno se define como *cono de confusión*, debido a la ambigüedad de las diferencias interaurales [Moore, 2004]. Lo anterior se ilustra en la Figura 2-5.

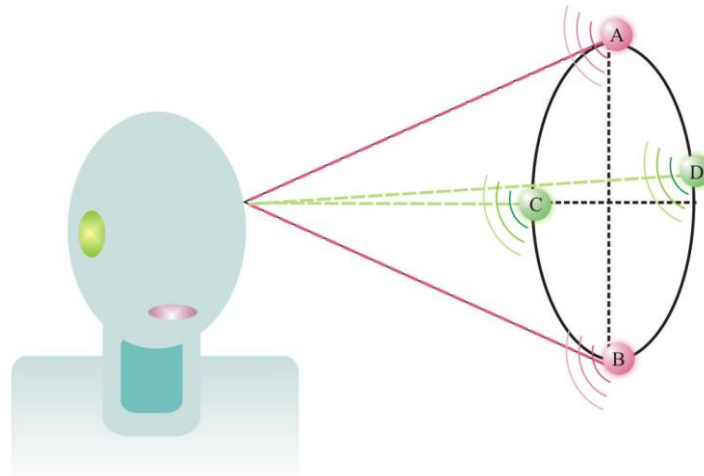


Figura 2-5: Representación gráfica del cono de confusión: para el oyente tanto las fuentes sonoras A y B como las fuentes sonoras C y D generan ILD e ITD que son perceptualmente indistinguibles.

2.4 Funciones de transferencia acústica referidas a la cabeza (HRTF)

Una señal que es emitida por una fuente sonora, ubicada en algún punto, antes de alcanzar los tímpanos de un oyente, es filtrada (modificada) acústicamente por la cabeza, los pabellones auriculares (orejas) y el torso. Esta serie de transformaciones que sufre la señal son generalmente diferentes para cada oído, y permiten que el sistema neuronal identifique la posición de la fuente. En términos de un sistema lineal, se puede definir la función de transferencia como la razón del espectro de la señal emitida por la fuente y la recibida en cada uno de los dos oídos. Las HRTF son diferentes para cada oído y dependen del ángulo de incidencia de la onda [Blauert, 1997; Nishino *et al.*, 2007], como se ilustra en la Figura 2-6.

Las funciones de transferencia acústica referida a la cabeza se definen como las respuestas en frecuencia entre fuentes sonoras en diferentes posiciones y los oídos de un oyente, dichas respuestas engloban todas las transformaciones antes mencionadas, las cuales están asociadas con los indicios perceptuales determinantes para la percepción auditiva del entorno [Duda, 1993; Wenzel *et al.*, 1993; Brungart y Rabinowitz, 1999; Cheng y Wakefield, 2001; Kazuhiro *et al.*, 2007; Ramos *et al.*, 2007].

Dado que existen diferencias anatómicas entre los seres humanos, las funciones de transferencia son únicas para cada persona [Møller *et al.*, 1995]. En general, las HRTF se miden en ambientes anecóicos con la finalidad de eliminar la influencia del recinto.

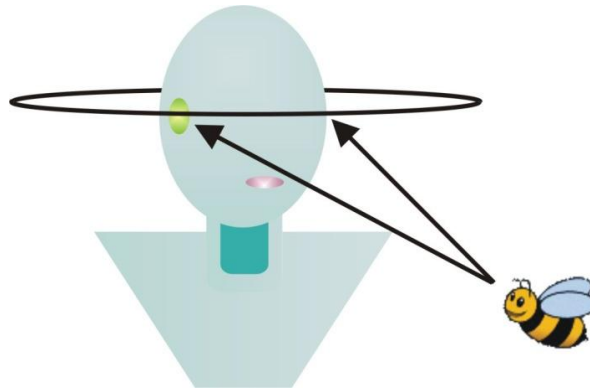


Figura 2-6: Función de Transferencia Referencia a la Cabeza (HRTF).

Existen muchos estudios que miden las HRTF directamente en sujetos [Møller *et al.*, 1995; Algazi *et al.*, 2001] o bien en maniqués acústicos, como el KEMAR² [Gardner y Martin, 1994; Nakado *et al.*, 2008], estos estudios tienen algunas diferencias en cuanto a las técnicas de medición, tales como el tipo de señal de excitación utilizada, el punto de referencia (punto del canal auditivo donde se realiza la medición), condición del conducto auditivo (abierto o bloqueado), etc. Otra diferencia es en cuanto a las posiciones angulares de la fuente, algunos trabajos sólo reportan ángulos en el hemisferio superior [Takane *et al.*, 2002], otros sólo en el plano horizontal [Nishino *et al.*, 2007], algunos otros miden en el plano horizontal considerando también la elevación de la fuente [Møller *et al.*, 1995; Algazi *et al.*, 2001]. Algunos trabajos incluyen también medidas antropométricas de los sujetos [Algazi *et al.*, 2001; Nishino *et al.*, 2007]. En internet se encuentran bases de datos que contienen HRTF las cuales se pueden consultar de forma libre [Gardner y Martin, 1994; Algazi *et al.*, 2001; Grassi *et al.*, 2003; IRCAM, 2003; Nakado *et al.*, 2008].

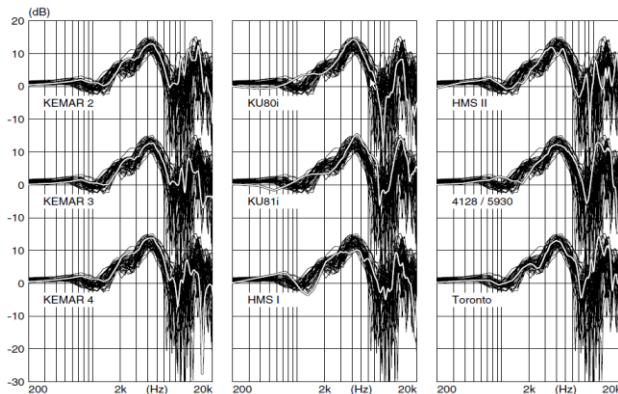


Figura 2-7: Estudio que muestra HRTF medidas en 40 sujetos (líneas oscuras) con el conducto auditivo bloqueado y con incidencia frontal, comparadas con las HRTF medidas con diferentes maniqués (líneas blancas) [Møller *et al.*, 1995].

² KEMAR (Knowles Electronics Manikin for Acoustics Research). Maniqué acústico que fue construido como una herramienta de prueba y evaluación para investigaciones acústicas relacionadas con la audición.

En la Tabla 2-1 se aprecian las diferencias que se encuentran en los trabajos que reportan mediciones de HRTF. Debido a que en este trabajo se realizan mediciones de HRTF y además se hace uso de una base de datos de mediciones de HRTF [Algazi *et al.*, 2001], la revisión de este tema resulta indispensable, ya que permite establecer el contexto de lo que se ha realizado en cuanto a técnicas medición, posiciones angulares, entre otras.

Tabla 2-1: Resumen de algunos trabajos que han medido HRTF.

Autores	Número de Sujetos / Maniquí acústico	Ángulos medidos (°)	Mic	Condición conducto auditivo (CA)	Señal empleada
Gardner y Martin (1994) ^(*)	KEMAR	ϕ de -40 a 90, para cada ángulo de elevación se midieron todo los ángulos en el plano horizontal (360°) en intervalos del mismo tamaño.	Etymotic ER-11 Preamplificador ER- 11		MLS
Møller <i>et al.</i> , (1995).	40	$-180 < \theta \leq 180$ $-90 < \phi \leq 90$ Ambos en int. de 22.5°.	B&K 4182	Bloqueado	MLS
Algazi <i>et al.</i> , (2001) ^{(*)(**)}	45	$\theta = \pm[80, 65, 55, 45, 40, 35, 30, 25, 20, 15, 10, 5, 0]$. $-45 \leq \phi \leq 230.625$, int. 5.625°.	ER-7C	Bloqueado	MLS
Takane <i>et al.</i> , (2002) ^(*)	3	$0 \leq \phi \leq 90$, int. 10° θ en intervalos de 5°, para cada ángulo de elevación.	Electret	Bloqueado	Pulso de tiempo extendido (OATPS)
IRCAM (2003) ^(*)	51	$-45 < \phi \leq 90$, int. 15°. θ en intervalos de 15°, para cada ángulo de elevación.	Knowless FG3329. Preamplificador de 40 dB de ganacia.	Bloqueado	<i>Chirp</i> logarítmica
Nishino <i>et al.</i> , 2007 ^{(*)(**)}	86	72 ángulos acimutales en int. de 5°.	SONY ECM-77B	Parcialmente bloqueado	TPS
Nakado <i>et al.</i> , (2008) ^(*)	B&K 4128 KEMAR		SONY ECM-77B	Bloqueado	<i>Chirp</i>

θ : ángulo acimutal; ϕ : ángulo de elevación. *Mediciones disponibles en la red. **Incluye medidas antropométricas.

2.5 Espacialidad del sonido

Según Marshall [Marshall *et al.*, 1981], la amplitud aparente de la fuente (ASW, Amplitude Source Width) aumenta con el número de las primeras reflexiones que inciden lateralmente sobre el oyente. Esto debido a que estas reflexiones se suman con el sonido directo, por lo que la localización de la fuente tiende a ser ambigua. Por otra parte, la existencia de primeras reflexiones laterales hace que aumente el grado de disimilitud entre los sonidos que llegan a los dos oídos, tanto por la diferencia entre los caminos que recorren las señales para llegar a ambos oídos, como por el obstáculo que representa la presencia de la cabeza. Cuanto mayor sea la diferencia entre ambos sonidos, mayor será la ASW.

Existen dos parámetros que permiten cuantificar el grado de ASW. El primero es la eficiencia lateral (LF), relacionada con la energía de las primeras reflexiones laterales, y el otro es la correlación interaural (IACC), asociada al grado de disimilitud entre las primeras reflexiones que llegan a ambos oídos. En esta tesis ocuparemos el segundo parámetro, es decir la correlación interaural (IACC).

2.6 Correlación interaural (IACC)

La correlación interaural (IACC, InterAural Cross-Correlation) es una medida de la similitud del sonido que llega a ambos oídos. Si las señales son iguales, el valor de este parámetro es 1; mientras que si las señales son distintas, el IACC toma un valor de 0. Matemáticamente se define como:

$$IACF_{(t_1 t_2)} = \frac{\int_{t_1}^{t_2} h_L(t) \times h_R(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} h_L^2(t) dt \times \int_{t_1}^{t_2} h_R^2(t) dt}} \quad (2-1)$$

(2-2)

$$IACC = \max |IACF_{t_1 t_2}|$$

Donde t_1 y t_2 definen el intervalo de tiempo de la respuesta impulso, h_L y h_R son las respuestas a impulso en el oído izquierdo y derecho, respectivamente. Normalmente, el intervalo de tiempo τ se elige entre -1 y +1 ms, que es el tiempo que toma, aproximadamente, el sonido en viajar de un oído al otro [Fletcher, 2007].

2.7 Efecto de fiesta coctel o *cocktail party effect*

Los oyentes que se encuentran en lugares públicos, como restaurantes, etc., están expuestos a diferentes sonidos que provienen de muy diversos puntos, sin embargo los seres humanos tienen la capacidad de enfocar la atención en escuchar a un orador en particular entre una mezcla de conversaciones y ruido de fondo, ignorando otras conversaciones. Esta capacidad de discriminar los sonidos se define como efecto de fiesta coctel (*cocktail party effect*) [Cherry, 1953; Pollack y Pickett, 1958].

Escuchar con los dos oídos tiene un efecto significativo en los sonidos que escuchamos, incluso si estamos inmersos en ambientes acústicos poco favorables [French y Steinberg, 1947; Cherry, 1953; Nordlund y Fritzell, 1963; Hawley *et al.*, 2004]. Otra ventaja significativa de la audición binaural es la posibilidad de localizar fuentes sonoras ubicadas en el espacio, principalmente cuando éstas se encuentran separadas espacialmente [Kock, 1950; Koenig, 1950; Dirks y Wilson, 1969a; MacKeith y Coles, 1971; Nábelek y Pickett, 1974; Bronkhorst y Plomp, 1988; Hawley *et al.*, 1999].

3 Revisión de antecedentes

En esta parte del trabajo se presentan los resultados publicados en artículos sobre el claro entendimiento de un mensaje y los factores que lo afectan, como por ejemplo, el incremento de la inteligibilidad cuando un mensaje es escuchado de forma binaural, además de qué forma se ve afectada por el ruido de fondo o bien por ambientes reverberantes. Además, las ventajas que se obtienen a través de las diferencias interaurales de tiempo e intensidad. Se muestran resultados de algunos trabajos que analizan la separación espacial de fuentes, así como también, los resultados de algunos trabajos respecto al efecto fiesta cocktail.

3.1 Inteligibilidad de la palabra.

Se define la inteligibilidad de la palabra, como la capacidad de entender lo que se dice en el interior de un recinto, o a través de algún otro sistema de comunicación acústica. Dos factores que afectan la inteligibilidad son la reverberación y el ruido de fondo [Steeneken y Houtgast, 1980; Houtgast y Steeneken, 1985a, 1985b].

Las pruebas subjetivas de inteligibilidad son herramientas útiles para evaluar la calidad de un canal de transmisión destinado a la transmisión de voz. En general, una prueba subjetiva de inteligibilidad, consiste en que un sujeto escucha un material de voz, y escribe en una hoja destinada para este propósito, o bien repite de forma oral, la palabra que acaba de escuchar. El porcentaje de aciertos de las palabras reportadas por el sujeto respecto al total de palabras emitidas, es un número que caracteriza la inteligibilidad del canal de transmisión bajo prueba.

En el idioma inglés, este tipo de pruebas tienen un amplio desarrollo, por lo que los estímulos auditivos utilizados como material de voz son muy variados. Dichos estímulos pueden ser palabras (monosílabas o bisílabas), enunciados, incluso un breve discurso, cada una de las cuales presenta ciertas ventajas [Miller *et al.*, 1951; Houtgast y Steeneken, 1984].

En lo que se refiere a la forma de presentar la respuesta, existen pruebas en las cuales el oyente tiene la libertad de escribir lo que entendió, a estas pruebas se les conoce como pruebas de respuesta abierta; mientras que las pruebas donde el oyente tiene que elegir entre un conjunto limitado de palabras, las cuales son posibles respuestas, se les denomina pruebas de respuesta cerrada.

3.2 Evaluación de la inteligibilidad

3.2.1 Métodos objetivos

Para evaluar la inteligibilidad de la voz de forma objetiva existen varios índices que permiten cuantificar este parámetro, a continuación se describen alguno de ellos.

- **Índice de articulación (AI)**

Este índice parte de la suposición de que la inteligibilidad de una señal de voz está basada en la suma de las contribuciones individuales de bandas de frecuencia, fue propuesto por French y Steinberg [French y Steinberg, 1947], y retomado más tarde por Kryter [Kryter, 1962; Steeneken y Houtgast, 1980]. Ellos describieron que la información contenida en la señal de voz no está distribuida de forma uniforme a lo largo del intervalo de frecuencia de dicha señal de voz, por lo que desarrollaron un modelo considerando 20 bandas de frecuencias contiguas, las cuales tienen una contribución igual a un índice definido. Los valores que toma el índice de articulación oscilan entre 0 (inteligibilidad nula) y 1 (inteligibilidad perfecta).

Este fue el comienzo del desarrollo y la aplicación de mediciones objetivas para predecir la inteligibilidad para diferentes canales de transmisión [Kryter, 1962].

- **Índice de transmisión de palabra (STI)**

Este índice se basa en la idea que de la señal de voz se considera como una señal modulada en amplitud, en donde el grado de modulación contiene la información útil de la voz. Si en la transmisión de la señal de voz a través del recinto se agrega ruido y/o reverberación, el grado de modulación en la señal se verá afectado, por lo que la inteligibilidad se reducirá.

La pérdida de modulación se cuantifica emitiendo ruido en siete bandas de frecuencias portadoras en el intervalo de 125 a 8,000 Hz (en bandas de octava), cada una modulada con 14 frecuencias de modulación de 0.63 a 12.5 Hz (en pasos de tercios de octava). A continuación, se calcula la relación entre el grado de modulación de la señal original y la señal recibida, se obtiene entonces el factor de reducción de modulación, en cada una de las 98 combinaciones (7 frecuencias portadoras y 14 frecuencias de modulación).

Después se obtiene el promedio ponderado del factor de reducción de modulación, lo que resulta en un número entre 0 y 1 que corresponde a condiciones de inteligibilidad pobre y excelente, respectivamente [Steeneken y Houtgast, 1980; Houtgast y Steeneken, 1985a].

▪ **Índice rápido de transmisión de palabra (RASTI)**

Con el objetivo de disminuir el tiempo de medición del STI surge otro parámetro como una simplificación del anterior, llamado RASTI. Esta simplificación consiste en reducir el número de bandas de frecuencias portadoras a dos bandas (500 y 2,000 Hz), y un subconjunto de 9 frecuencias de modulación (0.7, 1.0, 1.4, 2.0, 2.8, 4.0, 5.6, 8.0 y 11.2 Hz).

Algunas investigaciones [Steeneken y Houtgast, 1980; Houtgast y Steeneken, 1984 y Houtgast y Steeneken, 1985a,1985b], han establecido una escala cualitativa de inteligibilidad y su relación con los índices objetivos y los porcentajes de inteligibilidad (Figura 3-1). Dicha escala está basada en muchas pruebas experimentales realizadas en diferentes lenguajes y diferentes tipos de pruebas subjetivas de inteligibilidad (palabras bisílabas, frases cortas, números, etc).

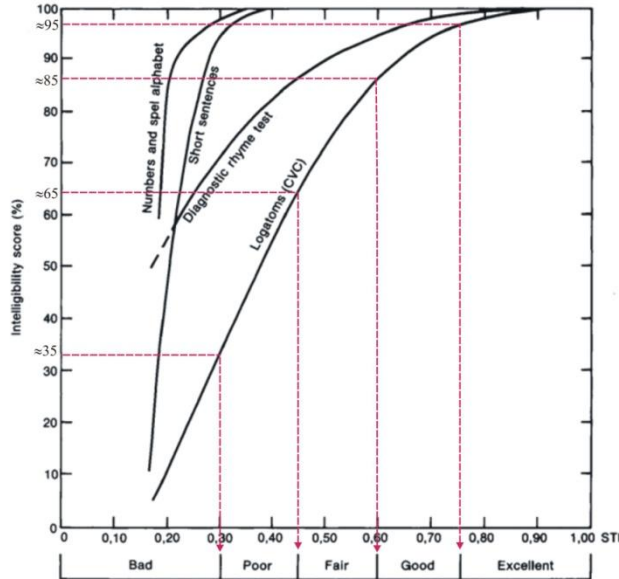


Figura 3-1: Relación entre la inteligibilidad de la palabra para diferentes tipos de pruebas [Houtgast y Steeneken 1985b].

Tabla 3-1: Relación entre la inteligibilidad de la palabra y los índices STI/RASTI [Houtgast y Steeneken 1985b].

Inteligibilidad	Mala	Pobre	Regular	Buena	Excelente
STI / RASTI	0 - 30	30 - 45	45 - 60	60 - 75	75 - 100
% de Aciertos ³	0 - 35	35 - 65	65 - 85	85 - 95	95 - 100

*Correspondencia entre índice STI/RASTI y prueba de inteligibilidad de *logátomos*³.

A partir de la Figura 3-1, es posible establecer una correspondencia entre la escala subjetiva de inteligibilidad en términos de los índices STI/RASTI, y el porcentaje de aciertos (Inteligibility Score) para una prueba con logátomos. En la Tabla 3-1 se muestra la relación entre los índices STI/RASTI y la escala cualitativa de la inteligibilidad de la palabra.

Todos los índices mencionados anteriormente son básicamente índices monoaurales, es decir, solo investigan la inteligibilidad considerando un solo oído, sin averiguar la interacción binaural. Se han propuesto algunos modelos objetivos para la evaluación de la inteligibilidad binaural. Un modelo comúnmente aceptado es el que se detalla a continuación:

- **Teoría de Ecuilización – Cancelación (EC)**

Este modelo propuesto por Durlach [Durlach, 1963], como su nombre lo indica, se basa principalmente en dos etapas: Ecuilización (E) y Cancelación (C). Este modelo, aprovecha el hecho de que las señales que llegan desde diferentes direcciones originan diferencias interaurales de tiempo y de nivel entre ambos oídos. En la etapa de ecuilización (E) la señal de un oído se atenúa y retrasa respecto al otro, de modo que las señales del oído derecho e izquierdo se ajustan hasta igualarse en amplitud y fase, lo anterior con el objetivo de que la señal no deseada sea la misma para ambos oídos. Mientras que en la etapa de cancelación (C), la señal de un oído se resta con la señal del otro oído con el objetivo de maximizar la relación señal-a-ruido SNR. Para que el proceso de EC se lleve a cabo con éxito, es necesario que la relación interaural de la señal no deseada sea diferente de la relación interaural de la señal deseada.

Existen modelos computacionales que implementan el modelo de EC para evaluar la inteligibilidad de forma binaural [Beutelmann y Brand, 2006; Beutelmann *et al.*, 2010; Wan *et al.*, 2010].

³ Palabras monosílabas CVC (consonante-vocal-consonante) sin sentido.

3.2.2 Métodos subjetivos

Las primeras pruebas de inteligibilidad fueron descritas en 1910, aunque fue a partir de 1930 y 1940 cuando se diseñan y utilizan con mayor frecuencia en investigaciones sobre comunicaciones telefónicas y comunicaciones de equipos militares, respectivamente.

En algunas pruebas subjetivas de inteligibilidad se utilizan oraciones portadoras, las cuales tienen la finalidad de centrar la atención del escucha, estas frases también permiten al orador modular su voz con el objetivo de mantener constante el nivel de su voz, a lo largo de toda la prueba [Egan, 1948]. Un ejemplo de una oración portadora es: “Usted escuchará la siguiente palabra: _____”.

Existen diferentes tipos de pruebas de inteligibilidad en idioma inglés, algunas de ellas son:

- **Prueba de rima modificada (MRT, Modified Rhyme Test):** Estas pruebas utilizan cincuenta listas de seis rimas o palabras homófonas (palabras que suenan de igual modo, pero que difiere en el significado) en inglés. Cada palabra está construida a partir de una secuencia de sonidos consonante-vocal-consonante (CVC), las seis palabras de cada lista difieren únicamente en el sonido de la consonante inicial o final (por ejemplo: hold, cold, told, fold, sold, gold). A los escuchas se les muestra una de las listas de seis palabras y su tarea es identificar cual de las seis es emitida por el orador. Usualmente se usa oración portadora.
- **Prueba de diagnóstico por rima (DRT, Diagnostic Rhyme Test):** Utilizan palabras monosílabas construidas a partir de una secuencia de sonidos CVC. Se ordenan ciento noventa y dos palabras, en noventa y seis parejas de rimas que difieren únicamente en sus consonantes iniciales (por ejemplo: meat-beat, weed-reed, yield-wield, hit-fit). A los escuchas se les muestra una pareja de palabras, y después se les pide que identifiquen las palabras que son presentadas por el orador. No se utilizan oraciones portadoras.
- **Palabras fonéticamente balanceada (PB):** Estas listas están formadas por 50 palabras monosílabas, las cuales se eligen para cubrir la frecuencia de ocurrencia de los fonemas de cada lenguaje, utilizado de forma cotidiana. Una característica de estas listas es que las palabras tienen la misma proporción de sonidos fundamentales del lenguaje común, además del mismo nivel de dificultad.

En el año 1948, Egan [Egan, 1948] desarrolló veinte listas de cincuenta palabras monosílabas fonéticamente balanceadas que inicialmente estaban dirigidas para determinar la inteligibilidad en sistemas de comunicación. Los criterios que utilizaron para construir las listas fueron:

- Tener en promedio la misma dificultad.
 - Deben tener la misma composición fonética.
 - Tener una composición fonética representativa del idioma.
 - Ser palabras de uso común.
-
- **Prueba de diagnóstico por consonantes centrales (DMCT):** Esta prueba se deriva de la prueba DRT. Utiliza una lista de noventa y seis parejas de palabras de dos sílabas que difieren únicamente en la consonante central (por ejemplo, bobble-bottle). Estas diferencias están organizadas en seis categorías, y la evaluación de cada categoría puede ser utilizada para identificar problemas específicos en un sistema de comunicación. A los escuchas se les muestra una pareja de palabras, y después se les pide que identifiquen las palabras que son presentadas por el orador. No se utilizan oraciones portadoras.
 - **Umbral de recepción de la voz (SRT, Speech Reception Threshold):** Se define como el nivel en el cual el oyente, que está sometido a la prueba, es capaz de identificar y reproducir el 50% del material que le es presentado. Estas pruebas se realizan en presencia de ruido de fondo o bien con fuentes separadas espacialmente. El método para obtener el SRT es mediante la aplicación de un método adaptivo el cual consiste en incrementar y/o disminuir el nivel del estímulo presentado, tal como se detalla a continuación: se presenta el estímulo auditivo, si el oyente da una respuesta incorrecta el nivel del estímulo se incrementa, se presenta nuevamente el estímulo si la respuesta es correcta el nivel del estímulo se disminuye, y así sucesivamente. Finalmente, se obtiene el promedio del nivel de presentación.

3.2.3 Material de voz en español

En este trabajo de tesis se utilizaron palabras bisílabas con sentido en español como material de voz. Este material consistió en palabras graves, cuya característica es que el acento está ubicado en la penúltima sílaba. Las palabras graves representan el tipo más común de palabras bisílabas en español [Zubick *et al*, 1983].

Es importante señalar que el desarrollo de material de voz en español para pruebas de inteligibilidad es muy limitado; sin embargo hay algunos trabajos en este campo [Tato, 1949; Ferrer, 1960; Cancel, 1965; Berruecos, 1967; Benitez, 1968]. Algunas de estas investigaciones han señalado la dificultad de diseñar listas de palabras monosílabas en español, debido a que existe un número muy limitado de palabras monosílabas con sentido [Tato, 1949; Ferrer, 1960]

3.3 Presentación monoaural y binaural de las señales

Uno de los primeros trabajos realizados en este tema fue hecho por Hirsh [Hirsh, 1950] en el que compara la audición monoaural con la audición binaural, estudia también los movimientos de la cabeza y la influencia de diferentes ambientes acústicos (anechoico y reverberante) en la inteligibilidad. Sin embargo, además de la conclusión básica (que fue demostrada tiempo después por muchos otros estudios, entre los cuales se pueden citar: [Dirks y Wilson, 1969a; MacKeith y Coles, 1971; Bronkhorst y Plomp, 1989] de que la inteligibilidad se incrementa cuando las fuentes sonoras (voz y ruido) se separan espacialmente. Algunos resultados son incompatibles con los encontrados en otros estudios realizados tiempo después, por ejemplo, reporta que la audición monoaural es superior a la audición binaural, sin embargo numerosas investigaciones han demostrado la superioridad de la audición binaural [Cherry, 1953; Pollack y Pickett, 1958; Nordlund y Fritzell, 1963; Moncur y Dirks, 1967; Olsen y Carhart, 1967; Dirks y Wilson, 1969a; MacKeith y Coles, 1971; Nabelek y Mason, 1981].

En la literatura existen pocos estudios que investigan el efecto del ángulo acimutal en el patrón de la señal de voz o bien en la inteligibilidad. Por ejemplo, Flanagan [Flanagan, 1960], estudió las características direccionales del sonido emitido por la boca, sin embargo no realizó mediciones de inteligibilidad. Nordlund y Fritzell [Nordlund y Fritzell, 1963] realizaron un estudio cuyo objetivo era investigar la influencia del ángulo acimutal en las señales de voz recibidas por un oyente. Realizaron grabaciones con un maniquí acústico a diferentes ángulos (0° , $\pm 30^\circ$, $\pm 60^\circ$, $\pm 90^\circ$, $\pm 135^\circ$ y $\pm 180^\circ$) y las presentaron de forma monoaural y binaural. Los resultados muestran que el ángulo acimutal tiene una clara influencia en la inteligibilidad, además también demostraron que bajo condiciones de audición binaural la inteligibilidad es significativamente mejor que la audición monoaural para todos los ángulos estudiados. Otros estudios en esa misma dirección son los realizados por [Nordlund, 1962; Nordlund y Lidén, 1963].

Drullman y Bronkhorst [Drullman y Bronkhorst , 2000] estudian la inteligibilidad, en relación con diferentes aspectos, que se enumeran a continuación: 1) distintos materiales de prueba, como palabras y enunciados. 2) múltiples oradores (de 1 hasta 4), hablando simultáneamente. 3) diferentes presentaciones: monoaural (monótica), binaural (dicótica) y 3D (con HRTF individualizadas y no individualizadas). 4) reconocimiento y localización de oradores. Los resultados de este trabajo muestran que el porcentaje de inteligibilidad que se obtiene con los enunciados, es mayor que el porcentaje obtenido con las palabras, esto se observa para todas las formas de presentación del sonido. También se concluye que a medida que aumenta el número de oradores hablando simultáneamente, la inteligibilidad disminuye considerablemente, esto se observa para ambos materiales de voz y para todas las formas de presentación.

En el caso de la presentación 3D y en presencia de dos o más oradores, el porcentaje de inteligibilidad es mayor comparado con las presentación monoaural y binaural, en particular cuando el material de voz son enunciados. Los resultados de las pruebas de inteligibilidad, reconocimiento y localización, utilizando HRTF individualizadas y no individualizadas, no mostraron diferencias significativas.

Nábelek y Robinson [Nábelek y Robinson , 1982] investigaron la percepción de la voz presentando el material de voz de forma monoaural y binaural en grupos de diferentes edades. También estudian el efecto que tienen los ambientes reverberantes en la inteligibilidad y cómo se relaciona con la edad de los oyentes. Las edades promedio de estos grupos son: 10, 27, 42, 54, 64 y 72 y los tiempo de reverberación que emplean son $T_{60} = 0.4, 0.8$ y $1.2s$. Los resultados muestran una ventaja significativa en la audición binaural sobre la audición monoaural incluso en condiciones reverberantes, así como también una disminución en la percepción del habla para tiempos de reverberación mayor. El grupo con mayor ventaja fue el de 27 años con 1.8, 4.2 y 5.3%, respectivamente para los diferentes tiempos de reverberación estudiados, en presentación binaural.

3.4 Efecto del ruido y la reverberación en la inteligibilidad

En muchas investigaciones se ha demostrado que la audición binaural tiene ventajas sobre la audición monoaural, incluso si el mensaje está inmerso en un ambiente ruidoso o reverberante. Koenig [Koenig, 1950] afirma que una fuente sonora puede ser escuchada si se desea; pero cuando no se escucha de manera consciente, tiende a desaparecer. Con un sistema binaural es posible

entender un mensaje de voz bajo condiciones de ruido extremo en un recinto, incluso con relaciones señal a ruido negativas [Koenig, 1950].

Un trabajo que estudia los efectos del ruido y la reverberación en presentación binaural y monoaural en la recepción de consonantes, es el realizado por Nabelek y Pickett [Nabelek y Pickett, 1974]. Para su estudio utilizaron dos diferentes tiempos de reverberación ($T_{60} = 0.3$ y 0.6 s), así como diferentes tipos de ruido: impulsivo y cuasi-estacionario (murmullo de ocho voces hablando simultáneamente). La señal de voz y la señal de ruido estaban espacialmente separadas por un ángulo de 60° . Las pruebas las realizaron en personas con audición normal y en personas con aparatos de ayuda auditiva. Los resultados muestran que escuchar de forma binaural presenta ventajas comparado con la escucha monoaural, considerando los efectos del sombreado que genera la cabeza, esto se puede explicar de la siguiente manera: al escuchar de forma binaural, el segundo oído obtiene información adicional que combinada con la información del otro oído mejora la recepción.

También observaron que para un tiempo de reverberación menor y sin ayuda auditiva, se obtiene una ventaja en la presentación binaural de 5 dB en presencia de ruido cuasi-estacionario y de 4 dB en presencia de ruido impulsivo. Con la introducción de la ayuda auditiva, y considerando mayores tiempos de reverberación, la ventaja en presentación binaural tiene un decremento de 3 dB. Además, concluyeron que la audición binaural no solamente mejora la percepción de la voz, sino también proporciona una mejor localización.

Algunos trabajos [Nordlund y Fritzell, 1963; Moncur y Dirks, 1967; Olsen y Carhart, 1967] definen la ventaja binaural como la diferencia en porcentaje de palabras correctas entre la presentación binaural y la monoaural.

El trabajo de Arsenault y Punch [Arsenault y Punch, 1999] estudia diferentes condiciones experimentales de presentación de las señales de voz y ruido. Realizaron grabaciones con un maniquí KEMAR, ubicaron a 0° (justo frente al maniquí) una fuente que emitía la señal de voz, mientras que a $+270^\circ$ (oído izquierdo) ubicaron una fuente que emitía una señal de ruido (grabada en una cafetería). Debido a la ubicación espacial de las fuentes sonoras, se observa que tanto el oído derecho como el izquierdo reciben la misma señal de voz; mientras que la señal de ruido es distinta para ambos oídos. El oído izquierdo recibe directamente la señal proveniente de la fuente de ruido; mientras que el oído derecho recibe la señal de ruido con una menor intensidad debido a la sombra generada por la cabeza.

A partir de la grabación biaural realizada con el maniquí acústico se estudiaron diferentes condiciones experimentales las cuales se detallan a continuación: 1) grabación biaural (realizada con el maniquí acústico), 2) biaural favorable, donde la grabación del oído derecho (ruido sombreado) se presentó en ambos oídos, 3) monoaural favorable, misma condición que la anterior, pero ahora solamente en el oído derecho, 4) monoaural no favorable, la grabación del oído izquierdo (ruido directo) se presentó solamente en el oído izquierdo, 5) ayuda simulada monoaural, en este caso la grabación biaural se presentó en ambos oídos, pero atenuada 20 dB en relación con la grabación biaural original.

Los resultados demuestran que la grabación biaural presenta una ventaja de 12.3, 4.9 y 2.3 dB sobre las presentaciones monoaural no favorable, monoaural favorable y biaural favorable, respectivamente. Estos resultados demuestran las ventajas que presenta la escucha biaural.

Se prueba además que los sujetos con audición normal presentan una mejora de 3.3 dB debido al efecto de sombreado que genera la cabeza, respecto a los sujetos con audición deteriorada, y 3.2 debido a la audición biaural.

El trabajo de Moncur y Dirks [Moncur y Dirks, 1967] analiza solamente los efectos de la reverberación en la inteligibilidad. Compararon los resultados del oído cercano (el oído que se ubica más cerca de la fuente de sonido), el oído lejano (el oído que se encuentra lejos de la fuente) y con ambos oídos para distintos tiempos de reverberación: $T_{60} = 0.9, 1.6$ y 2.3 s. Sus resultados muestran una ventaja en la inteligibilidad en la escucha biaural de 7.1, 10.2 y 9.5% respectivamente, lo anterior comparado con la audición monoaural. Además, observaron que las diferencias interaurales de tiempo (ITD) contribuyen de forma significativa en la inteligibilidad.

Otros trabajos analizan los efectos de la reverberación en personas con deficiencias auditivas como lo hacen Gelfand y Hochberg [Gelfand y Hochberg, 1976] quienes comparan un grupo de treinta personas con audición normal con un grupo de otras treinta personas con pérdida auditiva neurosensorial bilateral⁴. El material de voz fue la Prueba por Rima Modificada (MRT). Trabajaron con diferentes tiempos de reverberación, $T_{60} = 0, 1, 2$ y 3 s., las señales fueron presentadas de forma monoaural y biaural a través de audífonos. Los resultados muestran que en general, el grupo con audición normal obtuvo porcentajes de inteligibilidad significativamente más altos que el grupo con pérdida auditiva. En la presentación biaural los porcentajes de inteligibilidad fueron mayores que en la presentación monoaural, esto se observó incluso en los sujetos con deficiencia auditiva.

⁴ La pérdida de audición neurosensorial se produce por daños en las células ciliadas de la cóclea en el oído interno.

3.6 Separación espacial de las fuentes

Algunos trabajos se han realizado con fuentes de voz y ruido separadas espacialmente, como el de Kock [Kock, 1950] que llevó a cabo experimentos con una señal de voz ubicada a -45° (izquierda) y una fuente de ruido colocada a 135° (a la derecha y atrás del sujeto), es decir, las fuentes se encontraban diametralmente opuestas, respecto al oyente. Los resultados demuestran que cuando la señal de voz y la señal de ruido llegan a ambos oídos del escucha con el mismo tiempo de retardo, existe una disminución en la capacidad de concentrarse en la señal de voz en esos ángulos.

Uno de los experimentos realizados por Kock consistió en presentarle al oyente la misma señal de voz (esta señal de voz estaba en fase) y una señal de ruido distinta para cada uno de los oídos. La habilidad para detectar la señal deseada bajo estas condiciones se comparó con la habilidad de detectar la misma señal deseada en presencia de la misma señal de ruido para ambos oídos. Los resultados demuestran una mejoría entre un 5 y 10 dB del primer caso sobre el segundo, puesto que cuando la señal de voz y de ruido son idénticas en ambos oídos no existe un beneficio binaural.

MacKeith y Coles [MacKeith y Coles, 1971] realizaron experimentos colocando 2 altavoces alrededor del sujeto cuya posición se fue modificando, uno de los altavoces emitía la señal de ruido mientras que el otro emitía la voz. El material de voz empleado fue un conjunto de listas de palabras fonéticamente balanceadas (PB), el ruido empleado era de banda ancha de espectro similar al de la voz. Los resultados muestran una ventaja en la escucha binaural cuando la voz y el ruido provienen de fuentes separadas, cuando ambas fuentes se ubicaron frente al sujeto (0°), no se obtuvo ventaja. Se observó una mayor ventaja de la audición binaural cuando el ruido y la voz eran diametralmente opuestos; es decir, el altavoz que emitía la voz colocado a -90° (a la izquierda) y el de ruido a 90° (a la derecha) del sujeto.

Dirks y Wilson [Dirks y Wilson, 1969a] observaron que una separación de 10° entre la fuente de voz y de ruido, es suficiente para detectar un cambio medible en la inteligibilidad. La inteligibilidad binaural es superior a la monoaural en presencia de diferencias interaurales de tiempo (ITD).

3.7 Efecto de fiesta coctel (cocktail party effect)

Cherry [Cherry, 1953] realizó dos pruebas distintas de reconocimiento de voz. En la primera, presentó simultáneamente dos mensajes de voz distintos en ambos oídos del sujeto, emitidos por el mismo orador, la tarea del sujeto consistía en repetir de forma oral el mensaje presentado en uno de los oídos. Los resultados demuestran que es posible separar casi en su totalidad un mensaje en presencia de otro. En una variante de este mismo experimento se le solicitó al sujeto que escribiera

en una hoja las palabras o frases que identificaba. Subjetivamente hablando, esta tarea le pareció mucho más sencilla, además el tiempo que duraba la prueba se redujo considerablemente.

En la segunda prueba, presentó un mensaje en el oído izquierdo; mientras que el oído derecho se aplicó uno distinto, al igual que en la prueba anterior, la tarea del escucha era repetir el mensaje del oído izquierdo de forma oral. Los resultados indican que el sujeto no tiene dificultad en repetir el mensaje del oído izquierdo, sin embargo, una característica muy marcada es la monotonía y el poco contenido emocional de la voz en la repetición, aunque, subjetivamente, el sujeto no está consciente de este hecho. Por otro lado, es posible que el sujeto tenga poca claridad del mensaje que ha repetido, especialmente si el tema es difícil. El sujeto fue capaz de repetir algunas palabras presentes en el mensaje del oído derecho.

Se llevaron a cabo experimentos para identificar qué atributos son reconocidos del mensaje presentado en el oído derecho, es decir, el mensaje que debe ser rechazado. Los resultados de estos experimentos son: los cambios de idioma que ocurren en el transcurso de la prueba (por ejemplo, del idioma inglés al alemán) no fueron identificados, tampoco identificaron palabras aisladas ni el contenido semántico del mensaje. Sin embargo, del mensaje rechazado es posible reconocer ciertos detalles como por ejemplo, cuando existe un cambio de orador (de hombre a mujer) o bien la presencia de un tono puro (400 Hz).

El caso más desfavorable ocurre cuando la misma voz es utilizada como “señal deseada” y “señal no deseada”. Esta condición también fue estudiada por Cherry, y observó que para los sujetos es una tarea muy complicada descifrar dos mensajes emitidos por la misma persona.

Existen trabajos que investigan los efectos de la edad en la percepción de la voz, así como la separación espacial de fuentes sonoras, por ejemplo Kim y sus colaboradores [Kim *et al.*, 2006] trabajaron con 54 voluntarios divididos en grupos de acuerdo a su edad. El primero de ellos estaba formado por personas con edades entre 18 y 37 años, el segundo era de 38 a 57, y el tercero con personas mayores a los 58 años. Los sujetos se sentaron a un metro de distancia de tres altavoces ubicados a 0, 90 y 270° desde los cuales se emitió una señal de voz y una de ruido. La voz fue emitida por el altavoz ubicado a 0°, y para el ruido se emplearon los 3 altavoces, uno a la vez. Los resultados muestran que la edad juega un papel importante en la percepción del habla, debido a la pérdida de audición neurosensorial, la cual afecta principalmente altas frecuencias, esto se observa incluso en ambientes silenciosos. También concluyeron que cuando la voz y el ruido están separados 90° se obtiene una mejor inteligibilidad para todas las edades.

4 Pruebas de inteligibilidad

En la presente investigación, se llevaron a cabo cuatro pruebas subjetivas de inteligibilidad con la finalidad de cuantificar los beneficios que ofrece la escucha binaural. El objetivo de la primera prueba es investigar la inteligibilidad bajo diferentes formas de presentación de la señal de voz: monoaural (señal en un solo oído), monofónico (misma señal de voz en ambos oídos), binaural (dos señales distintas en cada uno de los oídos) y binaural a $+30^\circ$ (a la derecha del oyente) ó a -30° (a la izquierda del oyente). La señal de voz se presentó limpia, y bajo condiciones extremas de ruido y reverberación no correlacionadas interauralmente.

La segunda prueba estudia la influencia de la correlación interaural (IACC) en la inteligibilidad de la voz, es decir, se presentó la misma señal de ruido y reverberación en ambos oídos (máxima correlación interaural) y se compararon con los resultados obtenidos en la primera prueba.

La tercera prueba investiga la inteligibilidad a diferentes ángulos acimutales, para identificar los ángulos en los cuales un oyente recibe de manera más efectiva un mensaje. La orientación de los ángulos estudiados van desde -60° hasta $+60^\circ$ en intervalos de 15° .

La cuarta prueba cuantifica la ventaja que se obtiene al utilizar HRTF individualizadas contra HRTF no individualizadas.

En total participaron 79 voluntarios, todos alumnos universitarios a nivel licenciatura o maestría. Se descartaron algunos resultados en el caso de sujetos que mostraron pérdidas auditivas importantes (claramente manifestadas en su audiometría), o bien con *tinnitus*⁵. Para reducir la posibilidad de incluir sujetos con problemas de presbiacusia⁶, se limitó el intervalo de edad a personas no mayores de 50 años [CHABA, 1988; Mazelova *et al.*, 2003]. El intervalo de edad de los participantes fue de 19 a 38 con un promedio de 26.05 años.

⁵ Percepción del sonido en ausencia de un estímulo acústico [Manual Merck, 1978].

⁶ Pérdida de la capacidad para escuchar las frecuencias altas y, que se manifiesta a medida que las personas envejecen [Manual Merck, 1978].

En cada una de las pruebas se utilizó un grupo distinto de voluntarios para evitar el efecto de aprendizaje que ocurre cuando los oyentes se familiarizan con las palabras que escuchan [Neovius y Raghavendra, 1993].

En el caso del presente trabajo, orientado a la evaluación de sistemas de comunicación ante condiciones extremas de ruido y reverberación, el material de voz empleado en las pruebas de inteligibilidad consistió de palabras bisílabas con significado en español, balanceadas fonéticamente [Castañeda, 1991; Salinas, 2007]. Esta elección de palabras se justifica por la consideración de no someter a los sujetos a una tarea excesivamente complicada. La pertinencia de esto se verifica en función de que los resultados obtenidos no marcan porcentajes de inteligibilidad demasiado bajos (ver secciones siguientes). El empleo del material de voz utilizado en este trabajo, se justifica por el hecho de estar orientado a dispositivos móviles donde la comunicación principalmente es a través de palabras.

4.1 Prueba 1: Pruebas de inteligibilidad en presentación monoaural, monofónica, binaural a 0° y binaural a ±30° con ruido y reverberación no correlacionados interauralmente.

4.1.2 Objetivo

- Cuantificar el beneficio que se obtiene al utilizar técnicas de grabación y reproducción binaural contra técnicas estándares mono o estéreo.

4.1.3 Participantes

Participaron 30 sujetos (10 mujeres y 20 hombres), todos reportaron una audición normal. La edad de los participantes oscila entre los 20 y 36 años de edad, con una edad promedio de 25.8 años. Los sujetos no estaban familiarizados con las palabras de la prueba. Para todos los escuchas el idioma castellano es su lengua materna. Ninguno de los sujetos tenía experiencia en pruebas psicoacústicas. Todos los participantes eran estudiantes universitarios a nivel licenciatura o maestría.

4.1.4 Material de voz

El material de voz empleado en estas pruebas consistió de 4 listas diferentes de 50 palabras bisílabas con sentido en español balanceadas fonéticamente (PB, por sus siglas en inglés) [Castañeda, 1991; Salinas, 2007], emitidas por una mujer originaria de la Cd. de México. Las palabras tienen una frase introductoria con la finalidad de conseguir la atención del oyente, por ejemplo: “Usted escuchará la siguiente palabra: ____”. Las listas de palabras utilizadas se muestran en el Apéndice A.

El diseño del material de voz está construido a partir de las propiedades acústicas y fonéticas del idioma español; basados en estas características, el material de voz fue seleccionado y diseñado poniendo énfasis principalmente en el balance fonético contenido en cada lista. Se dice que el material está fonéticamente balanceado porque en cada lista están representados, aproximadamente, todos los sonidos o fonemas del idioma español característico de la Ciudad de México. Para el análisis fonético de las palabras se consideraron 22 fonemas descartando el fonema “X”, por ser el de menor incidencia en el español [Berruecos, 1967].

4.1.3.1 Grabación del material de voz

La grabación del material de voz se realizó en la Cámara de Transmisión del CCADET, que anteriormente estaba acondicionada como un pequeño estudio de grabación, porque cumplía con las condiciones adecuadas para realizar grabaciones, como bajo nivel del ruido de fondo, etc. Para la grabación de voz se utilizó una frecuencia de muestreo de $f_s = 44.1 \text{ kHz}$ [Salinas, 2007].

El equipo empleado para la grabación del material se detalla a continuación:

- Micrófono de condensador, patrón cardioide, marca Audio-Technica, modelo AT3525
- Consola digital marca Yamaha, modelo 03D.
- Interfaz de audio multicanal Echo Layla.
- Computadora Personal Pentium 4.

En este tipo de pruebas es muy importante la forma en que los participantes reciben el estímulo auditivo, es por eso que el uso de material grabado presenta algunas ventajas:

- El escucha no tiene referencia de lectura de labios durante el desarrollo de la prueba.
- La grabación tiene un nivel sonoro constante.

- La misma voz aparece en todas las grabaciones.

Al inicio de cada pista de voz grabada se incluyeron las instrucciones de la prueba de manera explícita. De modo que para el escucha quedaba claro el inicio y el final de la pista. Además, cada pista contaba con silencios, entre cada palabra, aproximadamente de 3 segundos con el propósito de dar tiempo suficiente al escucha de escribir la palabra.

Es preciso mencionar que ni la compilación del material de voz ni la grabación del mismo, se desarrollaron dentro de esta tesis [Castañeda, 1991; Salinas 2007].

4.1.3.2 Grabación biaural del material de voz

Es posible realizar una grabación biaural a través de maniqués acústicos los cuales simulan la forma en que un humano recibe las ondas sonoras. En el mercado existe equipo diseñado especialmente para el análisis de la audición biaural, que consiste en una cabeza, un par de pabellones auriculares, en los cuales se encuentran dos micrófonos, que emulan los dos tímpanos, un cuello y un torso, el diseño de este equipo está basado en las dimensiones promedio de un adulto [Burkhard y Sachs, 1975].

Algunos ejemplos de maniqués acústicos son el HMS (Head Measurement System) III de Head Acoustics (ver Figura 4-1), el maniqué KEMAR 45BA, HATS (Head and Torso Simulator) 4128 de Brüel & Kjaer (Figura 4-10a). Este equipo se utiliza en mediciones de ruido al interior de vehículos, diagnósticos de ruido, pruebas a protectores auditivos, entre otros.

Las grabaciones biaurales de voz, se realizaron en el interior de la Cámara Semi-Anecoica del SRCM (Systems Research Center-Mexico) de Intel Guadalajara. Se utilizaron dos maniqués de la marca Head Acoustics (Figura 4-1 izquierda). Con el maniqué HMS II.3 en modo parlante, se llevó a cabo la reproducción de las pistas de voz grabada; mientras que con el HMS III se realizó la grabación de las mismas. Los maniqués se colocaron a 1 m de distancia uno de otro. Las grabaciones se realizaron para diferentes ángulos acimutales, $\theta = 0, +30$ y -30° . El signo positivo se refiere a fuentes ubicadas a la derecha del oyente; mientras que el signo negativo indica fuentes ubicadas a la izquierda.



Figura 4-1: Maniqués con los que se realizó la grabación binaural. El maniqué HSM II.3 (izquierda) reprodujo la señal de voz, mientras que el maniqué HSM III (derecha) grabó la señal.

4.1.4 Procesamiento del material de voz

Existen varios factores que pueden disminuir la inteligibilidad de la palabra. La degradación de la voz puede ocurrir, por el ruido de fondo o la reverberación, si hablamos de una señal transmitida en el interior de un recinto, o bien por el ancho de banda y ruido, si hablamos de una señal en un sistema de comunicación. Generalmente, la señal recibida por el oyente no es exactamente la misma que emitió el orador, la reducción en el índice de modulación se cuantifica por la función de transferencia de modulación (MTF, Modulation Transfer Function).

El concepto de reducción de la amplitud de modulación propuesto por [Houtgast y Steeneken, 1985a; 1985b] fue empleado como guía para contaminar las señales de voz. Con el objetivo de obtener la misma degradación de la señal, considerando el efecto del ruido y la reverberación de forma independiente, se propuso que el valor del factor de reducción de modulación fuera: $m = 0.1$, que corresponde a un valor de índice de transmisión del habla, $STI=10\%$, el cual es asociado cualitativamente con una inteligibilidad mala (Ver Tabla 3-1) [Houtgast y Steeneken, 1985]. El factor de reducción de modulación se calcula de la siguiente manera:

$$m(F) = \frac{1}{\sqrt{1 + \left[2\pi F \frac{T_{60}}{13.8}\right]^2}} \times \frac{1}{1 + 10^{(-SNR)/10}} \quad (4-1)$$

Donde F es la frecuencia de modulación, T_{60} es el tiempo de reverberación y SNR es la relación señal a ruido.

De acuerdo a [Steeneken y Houtgast, 1980] y [Houtgast y Steeneken, 1985a, 1985b] es posible considerar los efectos del ruido y la reverberación de forma independiente. Para calcular los índices

de reducción de modulación de la señal afectada por el tiempo de reverberación, se emplea el primer término de la ecuación 4-1, como se muestra a continuación:

$$m(F) = \frac{1}{\sqrt{1 + \left[2\pi F \frac{T_{60}}{13.8}\right]^2}} \quad (4-2)$$

Se considera una frecuencia de modulación $F = 2 \text{ Hz}$, que es consistente con el ritmo de la producción del habla en las grabaciones del material de prueba utilizado en este trabajo (lo que equivale, aproximadamente dos sílabas pronunciadas por segundo).

Despejando T de la ecuación 4-2, se obtiene:

$$T = \frac{13.8\sqrt{1 - m^2}}{2\pi F m} \quad (4-3)$$

$$\therefore T = 10.92$$

Ahora bien, si se considera sólo el efecto del ruido los índices de reducción de modulación se calculan de la siguiente manera:

$$m = \frac{1}{1 + 10^{\left(\frac{-SNR}{10}\right)}} \quad (4-4)$$

Si de la ecuación anterior despejamos la SNR , tenemos:

$$-SNR = 10 \log \left(\frac{1 - m}{m} \right) \quad (4-5)$$

Ahora, sí $m = 0.1$ se tiene:

$$\begin{aligned} -SNR &= 9.542 \\ \therefore SNR &= -9.542 \approx -10 \text{ dB} \end{aligned} \quad (4-6)$$

Se observa entonces que, con el valor propuesto del factor de reducción de modulación de $m = 0.1$, se obtiene un tiempo de reverberación $T_{60} \approx 10 \text{ s.}$ y una relación señal a ruido $SNR \approx -10 \text{ dB.}$

Estos valores de reverberación y relación señal a ruido se utilizaron para contaminar de forma artificial las señales de voz utilizadas.

4.1.5 Contaminación de las señales de voz

Una vez grabadas las pistas binaurales de voz, se procesaron para contaminar las señales de voz con ruido y reverberación. El objetivo de presentar una señal de voz bajo condiciones acústicas de audición excesivamente nocivas, fue para garantizar una pérdida significativa de la inteligibilidad, por esta razón se simularon ambientes acústicos adversos. Para la señal de ruido se consideró una relación señal a ruido de $SNR = 10 \text{ dB}$; mientras que el tiempo de reverberación fue de $T_{60} = 10 \text{ s}$.

4.1.5.1 Señal de voz con ruido

Para obtener la señal de voz con ruido, se desarrolló un programa en Matlab versión 7.0 que genera una señal de ruido blanco, la cual se suma con la señal de voz con una relación señal a ruido de -10dB, según se describe en las siguientes ecuaciones:

$$S_{LN} = S_L(t) + N_L(t) \quad (4-7)$$

$$S_{RN} = S_R(t) + N_R(t) \quad (4-8)$$

Donde $S_L(t), S_R(t)$ son las señales binaurales grabadas en el oído izquierdo y derecho, respectivamente, $N_L(t), N_R(t)$ son las señales de ruido con distribución uniforme y promedio cero. Para contaminar las pistas de voz con ruido se utilizó la instrucción `rand(N,1)-0.5` de Matlab, la cual genera un vector columna de N números aleatorios con una distribución uniforme en el intervalo (-0.5, 0.5).

4.1.5.2 Señal de voz con reverberación

Para obtener una señal de voz con reverberación (Ver Figura 4-2), se desarrolló un programa en Matlab versión 7.0 que genera una respuesta impulso reverberante artificial, y se convoluciona con la señal de voz limpia. La convolución es una operación matemática que permite construir a partir de dos señales una tercera. El símbolo que se utiliza para denotar esta operación es: *. La respuesta impulso reverberante se muestra en la siguiente ecuación:

$$h(t) = h_0 u(t) \exp\left(\frac{-13.8t}{T_{60}}\right) \quad (4-9)$$

Donde h_0 es un factor de escala, $u(t)$ es una señal aleatoria con distribución uniforme, promedio cero y varianza unitaria. Esta señal se generó con la instrucción **rand** y se multiplicó con una señal exponencial decreciente cuya constante de decaimiento tiene un valor de 13.8, como se muestra en la ecuación (4-9). T_{60} es el tiempo de reverberación en segundos.

La señal de voz reverberada se obtiene por convolución, como se muestra en las siguientes ecuaciones:

$$S_{LT} = S_L(t) * h_L(t) \quad (4-10)$$

$$S_{RT} = S_R(t) * h_R(t) \quad (4-11)$$

Donde $h_L(t), h_R(t)$ es la respuesta impulso reverberante en el oído izquierdo y derecho, respectivamente.

Existen diversos métodos acústicos computacionales que permiten generar respuestas impulso que si bien son artificiales [Kuttruff, 2009], representan características acústicas realistas de alguna sala en específico. Sin embargo, nuestro interés es utilizar una respuesta a impulso cuyas características acústicas no sean particulares de alguna sala; sino que pueda ser generalizada a un número amplio de recintos, siendo el tiempo de reverberación su única característica. La ecuación 4-9 permite generar respuestas a impulso sin un patrón particular de reflexiones, pero sí asegura un decaimiento exponencial en amplitud, el cual está prescrito por el tiempo de reverberación dado.

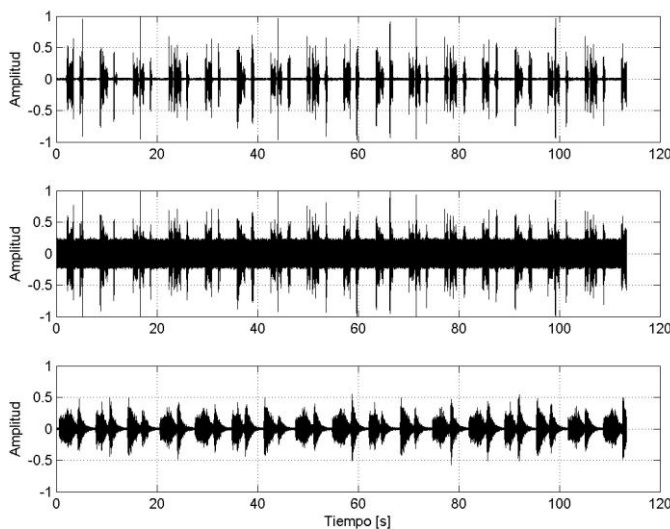


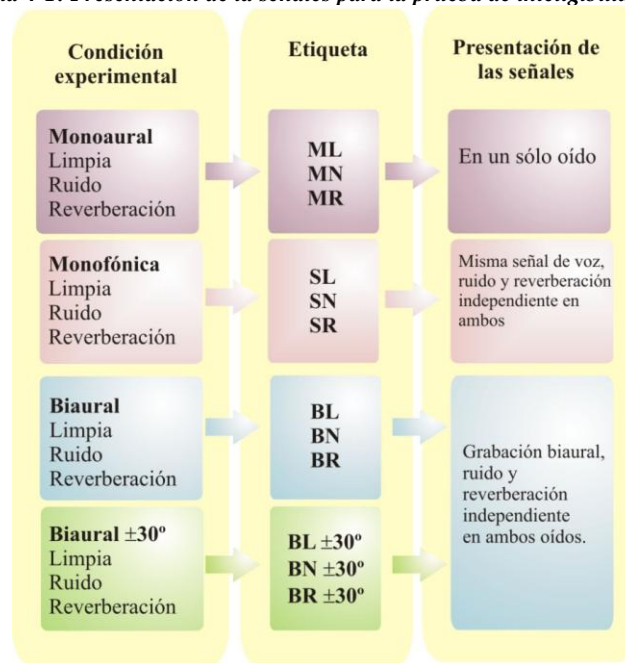
Figura 4-2: Forma de onda de las señales utilizadas en las pruebas de inteligibilidad. La primera figura muestra el espectro de la señal de voz limpia, en la del centro se muestra el espectro de la señal de voz contaminada con ruido, la última muestra el espectro de la señal de voz contaminada con una señal artificial reverberante.

4.1.6 Presentación de las señales

Las señales grabadas con el maniquí se presentaron a los sujetos en tres formas diferentes: monoaural, monofónica, biaural y biaural a $\pm 30^\circ$. La Tabla 4-1 muestra las condiciones experimentales y la presentación de las señales.

- **Monoaural (A):** Los dos canales de la grabación biaural se promediaron y la señal resultante se reprodujo en un solo oído (derecho o izquierdo, según la elección del sujeto).
- **Monofónica (M):** La señal monoaural, pero reproducida en ambos oídos.
- **Biaural (B):** La grabación biaural se reprodujo en ambos oídos.

Tabla 4-1: Presentación de la señales para la prueba de inteligibilidad 1.



4.1.7 Orden de las listas de palabras

El orden y condiciones de reproducción de las listas de palabras obedecen a una secuencia arbitraria que presenta en forma desordenada y evita que se repitan las listas de forma consecutiva bajo las mismas condiciones (limpia, ruido o reverberación). La secuencia de presentación de las listas de palabras es: 4, 3, 2, 1.

Por otro lado, la Figura 4-3 ilustra la forma en que se prepararon y presentaron las listas de palabras a los oyentes. Por motivos de conveniencia práctica, se prepararon de antemano pistas de sonido almacenadas digitalmente, de las listas de palabras para cada tipo de condición acústica: voz limpia, voz con ruido o con reverberación.

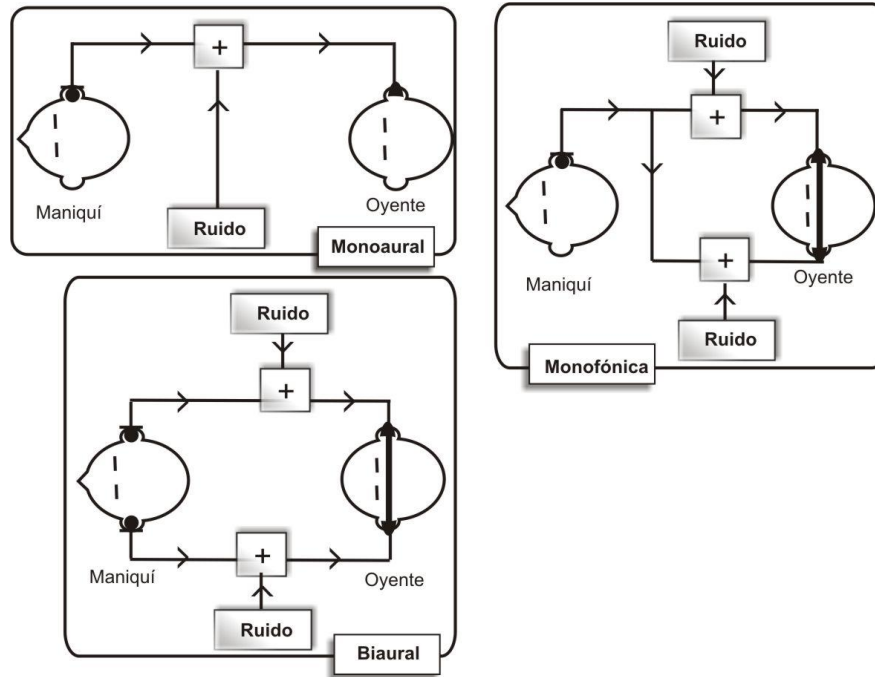


Figura 4-3: Presentación de las señales: Monoaural: señal de voz, ruido y reverberación en alguno de los 2 oídos. Monofónica: misma señal de voz, ruido y reverberación independiente en ambos oídos. Biaural: señal de voz distinta en ambos oídos.

Para medir el nivel de presión sonora de las pistas de voz se utilizó un maniquí B&K modelo 4128 y un analizador de espectros Brüel & Kjaer modelo 2133. Al maniquí se le colocaron un par de audífonos de la marca Sennheiser modelo HD 600. Las pistas se reprodujeron al nivel que se reproducen normalmente durante una prueba. La Figura 4-4 muestra la forma en que se realizaron las mediciones.

El nivel de presión sonora de las pistas de voz limpia fue de $L_{eq} = 70 \text{ dB}$. Sin modificar el nivel de la presentación, el nivel de las pistas de voz con ruido fue de $L_{eq} = 80 \text{ dB}$, esto se debe a que se contaminaron con una SNR fue -10 dB . El nivel de las pistas de voz con reverberación fue de $L_{eq} = 69.7 \text{ dB}$.

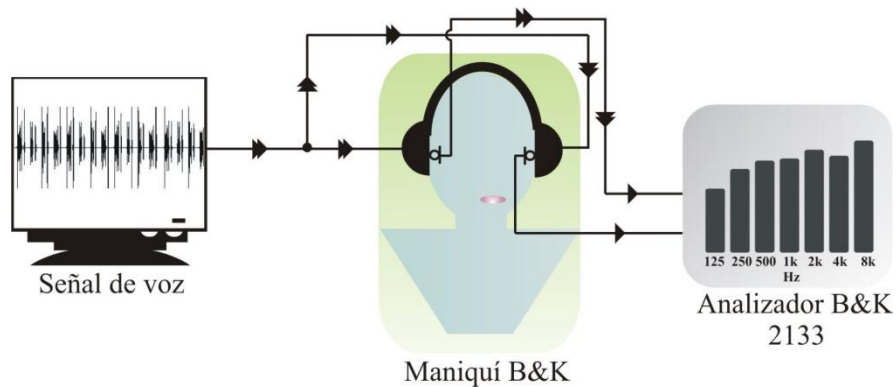


Figura 4-4: Diagrama experimental que ilustra la forma en que se realizaron las mediciones de nivel de presión sonora de las pistas de voz utilizadas en las pruebas.

4.1.8 Evaluación de la inteligibilidad

Para evaluar las pruebas de inteligibilidad se cuantificaron las palabras que el sujeto escribió correctamente (sin importar la ortografía) y según los aciertos se calculó el porcentaje de inteligibilidad. En cada prueba se considera que 50 aciertos equivalen a 100% de inteligibilidad. La evaluación se realiza como se muestra en la siguiente ecuación:

$$\% \text{ de inteligibilidad} = \frac{\text{número de aciertos} \times 100}{50 \text{ aciertos}}$$

Para la evaluación de las palabras no se consideraron los errores de ortografía, es decir, se tomaba como correcta si escribían “jedor” en lugar de “gestor”, etc.; también se consideraron las palabras homófonas como hacia/Asia (en el español hablado en México), tomando como correcta cualquiera de las dos palabras.

4.1.9 Desarrollo de la prueba

Los sujetos realizaron la prueba en el interior de la cámara semi-anechoica con unos audífonos marca Sennheiser modelo HD 600. La prueba se dividió en dos sesiones de 45 minutos cada una. En cada una de las sesiones los participantes escucharon en total 6 pistas de sonido con las grabaciones de las listas de palabras en diferentes condiciones acústicas, es decir, se repetían 2 de las 4 listas de palabras en cada sesión, sin embargo, no escucharon la misma lista de palabras bajo la misma condición.

Se les preguntó a los participantes cual era el oído de su preferencia (para hablar por teléfono, etc.) y de acuerdo a su respuesta se les presentaba la pista de sonido monoaural en el oído que habían elegido. Diecinueve personas escogieron el oído derecho y once el oído izquierdo.

A todos los participantes se les explicó de forma oral las instrucciones generales de la prueba, como la duración de la sesión, el número de sesiones en total, etc. La siguiente sesión era programada en un lapso no menor a dos días para evitar los efectos de memorización.

Después de un primer análisis de los resultados, se observó que algunas palabras de las listas eran consistentemente difíciles de entender por la mayoría de los sujetos; especialmente bajo condiciones severas de reverberación. El número de estas palabras fue diferente en cada lista, lo que genera resultados inconsistentes al comparar la evaluación de la inteligibilidad bajo diferentes condiciones las cuales fueron probadas con diferentes listas de palabras. Por esta razón, algunas palabras fueron eliminadas de algunas listas, teniendo en cuenta para su posterior análisis sólo aquellas que fueran entendidas por al menos el 33% de los sujetos en cada una de las pruebas. En algunos casos, esto lleva a reducir las listas a menos de 50 palabras, como se muestra en la Tabla 4-2.

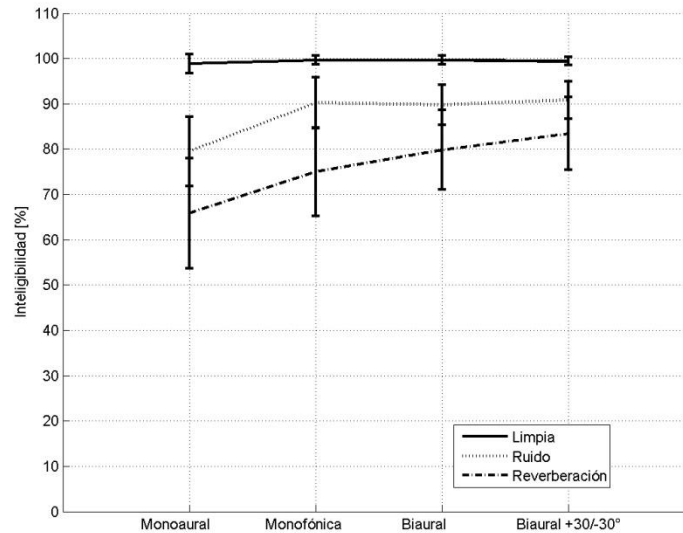
4.1.10 Resultados

Los resultados muestran que cuando la señal está limpia la inteligibilidad tiene casi 100% para todas las presentaciones (monoaural, monofónica, binaural y binaural $\pm 30^\circ$), lo que demuestra que el material de voz (sin disturbios) empleado en las pruebas, es una referencia adecuada ya que los sujetos obtienen la puntuación máxima que es posible obtener.

También se observa una ventaja significativa en la presentación binaural bajo condiciones acústicas adversas (ver Figura 4-5), lo anterior confirma los resultados de Moncur y Dirks, 1967, así como los de Libbey y Rogers, 2004.

Bajo condiciones de ruido, en la presentación monofónica (90.2%) la inteligibilidad mejora un 10.8% sobre la presentación monoaural (79.4%). Como se esperaba, en la presentación monofónica (90.2%) y binaural a 0° (89.7%) se obtienen resultados similares, puesto que en ambas presentaciones las señales en el oído derecho y el oído izquierdo son casi idénticas. En la presentación binaural a $\pm 30^\circ$ (90.8%), se obtiene una ventaja de 11.4% en la inteligibilidad sobre la presentación monoaural (79.4%).

En condiciones de reverberación y para la presentación monofónica (74.9%), se obtiene una ventaja de 9.1% en la inteligibilidad, sobre la presentación monoaural (65.8%); mientras que para la presentación biaural a 0° (79.8%) se obtiene un ventaja de 14% y de 17.6% cuando el oyente no está directamente frente al orador, sobre la presentación monoaural, respectivamente.



Limpia	98.8 (2.08)	99.6 (0.96)	99.6 (0.96)	99.4 (0.93)
Ruido	79.4 (7.63)	90.2 (5.60)	89.7 (4.45)	90.8 (4.10)
Reverberación	65.8 (12.20)	74.9 (9.73)	79.8 (8.80)	83.4 (8.04)

Figura 4-5: Porcentajes correctos de inteligibilidad para distintas presentaciones y considerando diferentes ambientes acústicos. En la tabla se muestra el promedio y la desviación estándar entre paréntesis.

Tabla 4-2: Presentación de las señales bajo condiciones de ruido y reverberación no correlacionadas y número de palabras utilizadas en el análisis de los resultados.

Presentación	Número de palabras
Monoaural	
Ruido	44
Reverberación	31
Monofónica	
Ruido	49
Reverberación	41
Biaural	
Ruido	44
Reverberación	40
Biaural ±30°	
Ruido	49
Reverberación	46

4.1.10.1 Análisis estadístico

Se realizó un análisis estadístico, con la finalidad determinar la certeza estadística de los resultados obtenidos. Se determinaron los intervalos de confianza para las diferentes presentaciones y considerando las diferentes condiciones acústicas, además se establecieron pruebas de hipótesis, en distintos casos (Ver Tabla 4-3 y Figura 4-6), también se determinó la certeza estadística de las pruebas de hipótesis en función del número de sujetos involucrados, con el objetivo de optimizar este número en pruebas futuras. Los análisis se basaron en la distribución t de Student, que se aplica en casos en los que el tamaño de la muestra es limitado (típicamente < 30) [Spiegel *et al.*, 2009].

4.1.10.2 Intervalos de confianza

Se denomina intervalo de confianza a un par de números entre los cuales se estima que estará cierto valor desconocido, con una determinada probabilidad de acierto. Formalmente, estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. Generalmente se construyen intervalos con una confianza estadística de 95%. Los límites del intervalo de confianza están dados por:

$$L_{SUP} = \bar{X} + Z_C \frac{\sigma}{\sqrt{n}} \qquad L_{INF} = \bar{X} - Z_C \frac{\sigma}{\sqrt{n}} \qquad (4-12)$$

Donde \bar{X} es la media, σ es la desviación estándar, Z_C es el valor crítico de la distribución t considerando $n - 1$ grados de libertad. El valor crítico Z_C depende del nivel de confianza que se desee, en este caso el nivel de confianza empleado es de 95%. El valor crítico se obtiene directamente de tablas (ver Tabla 4-3) y está en función del número de muestras, en este caso $n = 30$, por lo que el valor crítico es 1.96.

Tabla 4-3: Valores de Z_C utilizados en la práctica para distintos niveles de confianza [Spiegel *et al.*, 2009].

Nivel de confianza	99.73	99	98	96	95.45	95	90	80	68.27
Z_C	3	2.58	2.33	2.05	2	1.96	1.645	1.28	1

En la Tabla 4-4 se muestran la media y los intervalos de confianza calculados:

Tabla 4-4: Media e intervalos de confianza (entre paréntesis) con un nivel de confianza del 95%, del porcentaje de inteligibilidad para las distintas presentaciones y considerando todos los ambientes acústicos.

	Limpia	Ruido	Reverberación
Monoaural	98.8 (99.6, 98.1)	79.4 (82.2, 76.7)	65.8 (70.1, 61.4)
Monofónica	99.6 (99.9, 99.3)	90.2 (92.2, 88.1)	74.9 (78.4, 71.4)
Biaural	99.6 (99.9, 99.3)	89.7 (91.3, 88.1)	79.8 (82.9, 76.7)
Biaural +30°/-30°	99.4 (99.7, 99.1)	90.8 (92.2, 89.3)	83.4 (86.3, 80.5)

4.1.10.3 Pruebas de hipótesis

En estadística, una prueba de hipótesis es cualquier afirmación acerca de una población o bien de sus parámetros. Este tipo de pruebas contrastan dos hipótesis estadísticas. Tal contraste involucra la toma de decisión acerca de las hipótesis. La decisión consiste en rechazar o no una hipótesis en favor de la otra. Una hipótesis estadística se denota por “H”, de las cuales se proponen dos:

- H_0 hipótesis nula.
- H_1 hipótesis alternativa.

Se establecieron pruebas de hipótesis en las cuales se compararon los porcentajes de inteligibilidad obtenidos bajo diferentes condiciones (considerando diferentes ambientes acústicos, diferentes presentaciones, etc.).

La Figura 4-6, muestra la certeza estadística de las pruebas de hipótesis con un nivel de significancia del 95%. Se observa que bajo condiciones de ruido y reverberación, la inteligibilidad biaural a 0° será mayor que para la presentación monoaural. Los resultados también muestran que entre 10 y 15 pruebas son suficientes para tener un resultado con porcentaje muy cercano al 100% de probabilidad, por esta razón se decidió incluir solamente 15 sujetos en la segunda prueba. Lo anterior permite concluir que, incluso si se realizaran un número mayor de pruebas, los resultados mostrarían la misma tendencia.

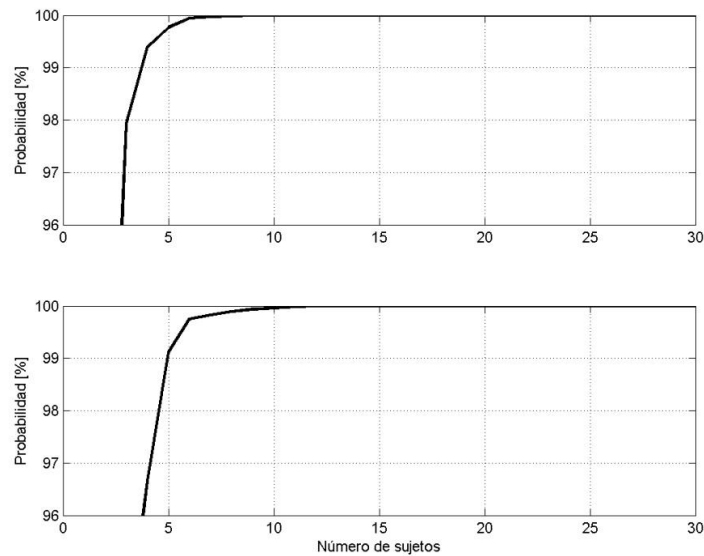


Figura 4-6: Certeza estadística de las pruebas de hipótesis para diferentes casos. Arriba) *Inteligibilidad biaural a 0° > Inteligibilidad monoaural con ruido*. Abajo) *Inteligibilidad biaural > Inteligibilidad monoaural con reverberación*.

4.1.11 Conclusiones

Se llevaron a cabo pruebas subjetivas de inteligibilidad bajo condiciones extremas de ruido y reverberación. Se consideró una relación señal a ruido de $SNR = 10 \text{ dB}$ y un tiempo de reverberación de $T_{60} = 10 \text{ s}$. Se estudiaron diferentes formas de presentación de la señal: monoaural, monofónica, biaural a 0° y biaural a $\pm 30^\circ$, con disturbios no correlacionados interauralmente. Los resultados de estas pruebas demuestran que la inteligibilidad se ve afectada por el ruido y la reverberación; sin embargo bajo estas condiciones acústicas adversas, es posible mejorar la inteligibilidad a través de la audición biaural, comparado con la presentación monoaural y monofónica.

4.2 Prueba 2: Pruebas de inteligibilidad con ruido y reverberación correlacionados.

4.2.1 Objetivo

- Investigar el efecto de la correlación interaural en la inteligibilidad.

4.2.2 Participantes

Participaron 15 sujetos (6 mujeres y 9 hombres), sin problemas de audición; que no habían participado en las pruebas anteriores. La edad de los participantes oscila entre los 21 y 38 años de edad, con una edad promedio de 26.9 años. Al igual que en la prueba anterior, los sujetos no estaban familiarizados con las palabras de la prueba y para todos los escuchas el idioma castellano es su lengua materna.

4.2.3 Procesamiento de las señales biaurales

Debido a que en esta prueba se estudia la inteligibilidad considerando la misma señal de ruido y reverberación en ambos oídos, la señal monoaural ya no se contempla en este experimento. Por lo que sólo se trabaja con la señal monofónica, baural a 0° y baural a $\pm 30^\circ$.

Las grabaciones biaurales se contaminaron con la misma señal de ruido y reverberación en ambos oídos, como se muestra en la Figura 4-7. La señal baural $\pm 30^\circ$ se procesó de la misma manera que la señal baural a 0° .

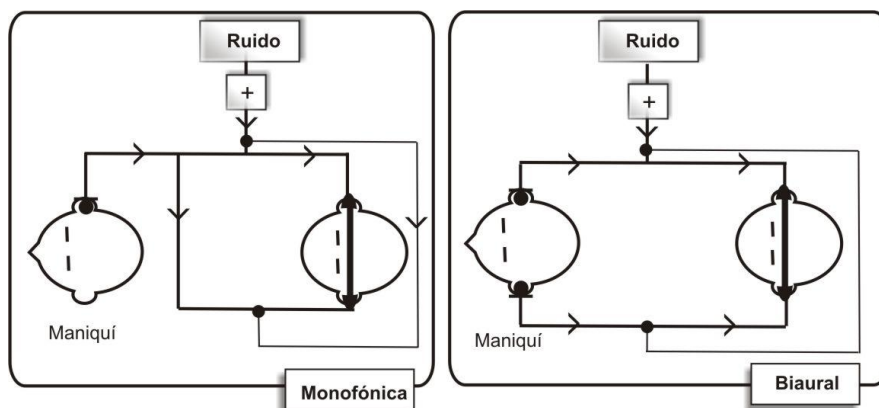


Figura 4-7: Presentación de las señales. Monofónica: misma señal de voz, ruido y reverberación en ambos oídos. Baural: señal de voz distinta en ambos oídos, misma señal de ruido y reverberación.

4.2.4 Desarrollo de la prueba

El procedimiento para esta prueba fue el mismo que para la prueba anterior, pero en este caso se omite la señal limpia, por esta razón la prueba se realiza en una sola sesión y toma aproximadamente 45 minutos.

Para investigar el efecto de la correlación interaural se realizaron pruebas similares a las anteriores, pero ahora con señales de ruido y reverberación completamente correlacionadas en ambos oídos. La Tabla 4-5 muestra el orden y la condición experimental de las pistas de voz para la prueba 2.

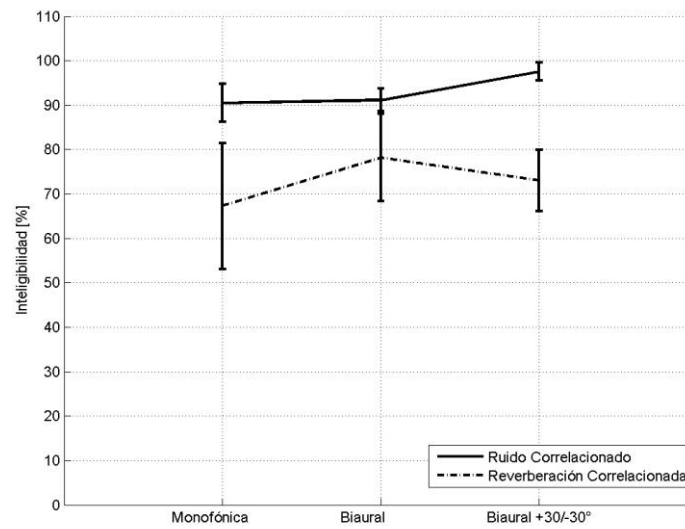
Tabla 4-5: Orden de las pistas de voz utilizadas en la prueba 2.



4.2.5 Resultados

Según los resultados obtenidos del análisis estadístico de la prueba 1, es suficiente trabajar con 15 personas para obtener 95% de confianza estadística en los resultados, por lo que en esta prueba se trabajó con 15 personas.

Los resultados se muestran en la Figura 4-8, se observa que para la presentación biaural a $\pm 30^\circ$ (97.5%) y bajo condiciones de ruido correlacionado, se obtiene una ventaja de 7.1% y 6.4% sobre las presentaciones monofónica (90.4%) y biaural a 0° (91.1%), respectivamente. Bajo condiciones de reverberación correlacionada, la presentación biaural a 0° (78.1%) presenta una ventaja de 10.9% sobre la presentación monofónica (67.2%).



Ruido	90.4 (4.27)	91.1 (2.56)	97.5 (1.98)
Reverberación	67.2 (14.1)	78.1 (9.80)	73.0 (6.87)

Figura 4-8: Porcentajes de inteligibilidad para distintas presentaciones considerando señales de ruido y reverberación correlacionadas. En la tabla se muestra el promedio y la desviación estándar entre paréntesis.

La Tabla 4-6 muestra el número de palabras que fueron entendidas por al menos el 33% de los sujetos, y las cuales fueron utilizadas para el análisis de los resultados.

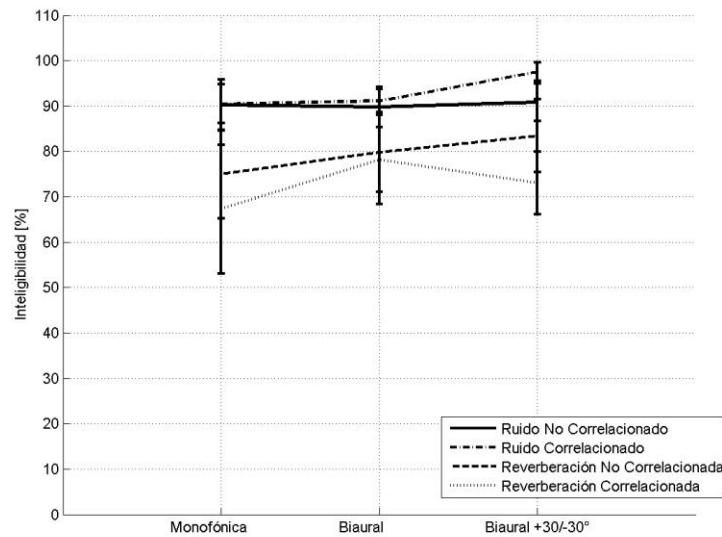
Tabla 4-6: Presentación de las señales bajo condiciones de ruido y reverberación correlacionadas y número de palabras utilizadas en el análisis de los resultados.

Presentación	Número de palabras
Monofónica	
Ruido	42
Reverberación	24
Biaural	
Ruido	50
Reverberación	37
Biaural ±30°	
Ruido	49
Reverberación	40

4.2.5.1 Comparación entre señales correlacionadas y no correlacionadas

La comparación entre la evaluación de la inteligibilidad en presencia de señales de ruido y reverberación correlacionadas y no correlacionadas se muestra en la Figura 4-9. Se observa que en condiciones de ruido y para la presentación biaural a $\pm 30^\circ$, la inteligibilidad mejora un 6.7% bajo

condiciones correlacionadas (97.5%), comparado con ruido no correlacionado (90.8%). Por otro lado, en presencia de reverberación y bajo condiciones no correlacionadas, la inteligibilidad mejora para todas las presentaciones (monofónica, biaural 0° y biaural ±30°), comparada con condiciones reverberantes correlacionadas. Para la presentación monofónica y en presencia de reverberación, la inteligibilidad tiene una ventaja de 7.7% para condiciones no correlacionadas (74.9%) comparado con reverberación correlacionada (67.2%). Estos resultados coinciden con el estudio realizado por Durlach (1963), que demostró que la inteligibilidad tiende a ser más pobre para la presentación monofónica, es decir, cuando las señales (deseada y no deseada) son las mismas en ambos oídos (IACC = 1).



RUIDO	Correlacionado	90.4 (4.27)	91.1 (2.56)	97.5 (1.98)
	No Correlacionado	90.2 (5.60)	89.7 (4.45)	90.8 (4.10)
REVERBERACIÓN	Correlacionado	67.2 (14.14)	78.1 (9.80)	73.0 (6.87)
	No Correlacionada	74.9 (9.73)	79.8 (8.80)	83.4 (8.04)

Figura 4-9: Porcentajes de inteligibilidad para distintas presentaciones y considerando señales de ruido y reverberación correlacionadas y no correlacionadas. La desviación estándar se muestra entre paréntesis.

En la Tabla 4-7 se presentan los valores de la correlación interaural para las pistas de voz empeladas en las pruebas 1 y 2.

Tabla 4-7: Correlación Interaural (IACC) de las pistas de voz empleadas en las pruebas.

Pista	Correlación Interaural	
	No correlacionado	Correlacionado
MN	0.2955	1
MR	0.0860	1
BN	0.2840	0.9898
BR	0.0780	0.9769
BN +30°	0.2549	0.7324
BR +30°	0.1049	0.9094
BN -30°	0.2209	0.7334
BR -30°	0.0748	0.9113

4.2.6 Análisis de los resultados

En presencia de ruido y cuando el orador está ubicado a $\pm 30^\circ$ respecto al oyente, la inteligibilidad es mayor bajo condiciones correlacionadas que no correlacionadas, lo cual es consistente con los resultados de otro estudio [Licklider, 1948]. Estos resultados también son consistentes con la teoría de Ecuación – Cancelación (EC) [Durlach, 1963], de acuerdo con este modelo, el hecho de que la señal de ruido sea la misma para ambos oídos (ruido correlacionado) permite la cancelación de esta señal; mientras que, bajo condiciones de ruido no correlacionado, la cancelación de la señal de ruido es muy difícil o bien imposible de lograr. Otro factor importante es la lateralización de la señal de voz, debido a que genera diferencias interaurales que son aprovechadas por el sistema auditivo para enfatizar la relación señal a ruido mejorando la inteligibilidad de la voz.

Por otro lado, los resultados obtenidos bajo condiciones de reverberación correlacionada y no correlacionada muestran que la inteligibilidad tiende a ser mayor para condiciones de reverberación no correlacionada. Este resultado también puede ser explicado con base a la teoría de Ecuación – Cancelación, puesto que bajo condiciones de reverberación no correlacionada, la respuesta a impulso reverberante es diferente para cada uno de los oídos, sin embargo la relación de fase interaural se mantiene constante para ambos oídos, por esta razón, el sistema auditivo es capaz de ecualizar y cancelar la parte reverberante de la señal, mejorando así la inteligibilidad de la voz.

En el caso de ruido no correlacionado, la relación de fase interaural varía de forma aleatoria, de modo que es distinta para cada uno de los oídos, por lo que no es posible cancelar la señal de ruido.

4.2.7 Conclusiones

Se llevaron a cabo pruebas subjetivas de inteligibilidad bajo condiciones extremas de ruido y reverberación. Se estudiaron señales de voz contaminadas con disturbios correlacionados y no correlacionados interauralmente. Los resultados de estas pruebas demuestran que la inteligibilidad biaural mejora bajo condiciones de ruido correlacionado y reverberación no correlacionada.

4.3 Prueba 3: Pruebas de inteligibilidad a diferentes ángulos acimutales

4.3.1 Objetivo

- Evaluar la inteligibilidad a diferentes ángulos sobre el plano horizontal, para determinar la posición en la cual el oyente recibe de manera más efectiva el mensaje considerando condiciones acústicas adversas.

4.3.2 Participantes

En estas pruebas participaron 15 sujetos (7 mujeres, 8 hombres) ninguno de ellos participaron en las pruebas anteriores. El rango de edad fue de 19-34 años, con un promedio de edad de 24.5 años; las características de los participantes son similares a las pruebas anteriores. A todos los participantes se les realizó una audiometría (previa a la prueba de inteligibilidad), descartando a todos aquellos que presentaran una pérdida auditiva mayor a los 15 dB HL. El examen auditivo se realizó con un audiómetro marca Brüel & Kjaer modelo 1800, que analiza las frecuencias centrales de las bandas de octava de 500 a 8 000 Hz.

4.3.3 Grabación biaural de la señal de voz a diferentes ángulos en el plano horizontal

Se llevaron a cabo grabaciones biaurales a diferentes ángulos en el plano horizontal, desde -60° a $+60^\circ$ en intervalos de 15° . Estas grabaciones se realizaron en la cámara anecoica del Laboratorio de Acústica y Vibraciones del CCADET. La grabación se realizó con un maniquí de la marca Brüel & Kjaer, Head and Torso Simulator Modelo 4128 (Figura 4-10 derecha), la señal de voz se tenía almacenada en una computadora personal y se reproduce a través de un altavoz auto-amplificado marca Event Electronics ALP5. La reproducción y grabación se realizó a través de una interfaz de audio M-Audio modelo Pro Fire 610. La distancia entre el maniquí y el altavoz fue aproximadamente 1 m.

La primera grabación biaural se realizó para el ángulo de $+60^\circ$, ya que en este punto es donde el altavoz se encuentra más cerca y más orientado hacia los oídos del maniquí, por lo que en esta

posición se obtiene el nivel más alto de grabación. Lo anterior con el objetivo de establecer la misma ganancia de la señal registrada por el micrófono para todos los ángulos, evitando así variaciones en el nivel debido al cambio de posición del altavoz (por cuestiones de facilidad, la posición del altavoz se fue modificando según el ángulo a grabar; mientras que la posición del maniquí se mantuvo fija durante la grabación) y además evitar que la señal registrada por los oídos del maniquí se sature.

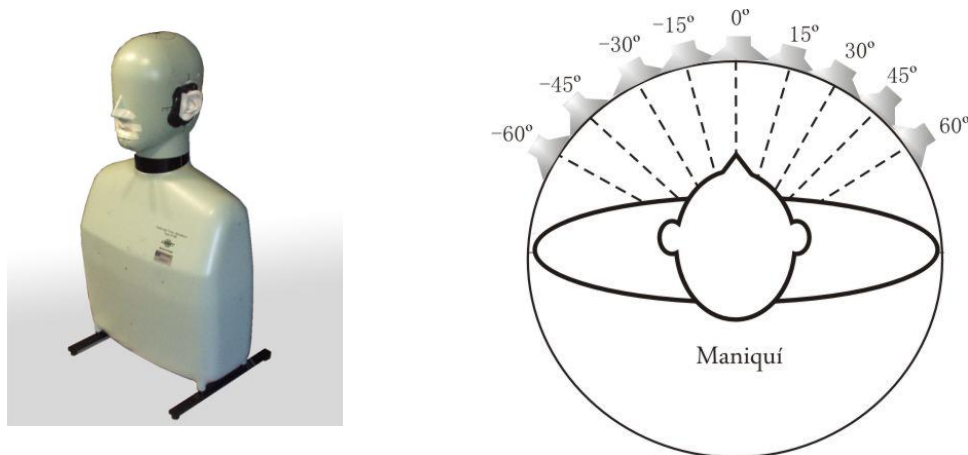


Figura 4-10: Maniquí acústico B&K HTS 4128 utilizado en las grabaciones binaurales realizadas en la cámara anecoica del CCADET (izquierda). Posición de los altavoces para las grabaciones binaurales (derecha).

El diagrama experimental para llevar a cabo las grabaciones binaurales a diferentes ángulos se muestra en la Figura 4-11.

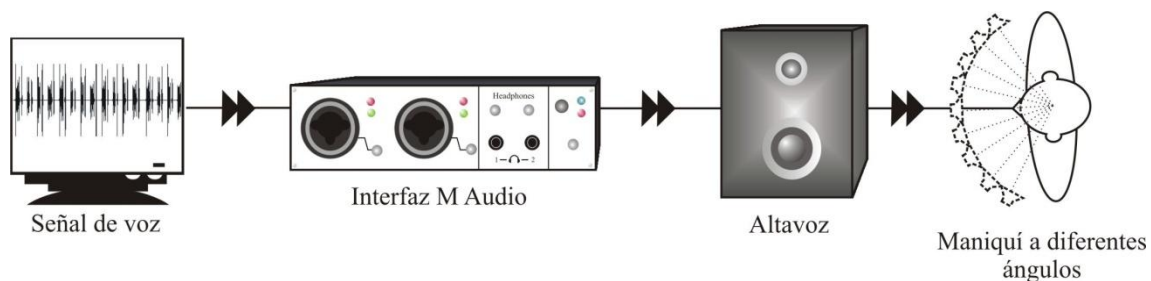


Figura 4-11: Diagrama experimental para las grabaciones binaurales a diferentes ángulos en el plano horizontal desde -60° hasta $+60^\circ$ en intervalos de 15° .

4.3.4 Procesamiento de las señales

Una vez grabadas las pistas de voz a diferentes ángulos, se contaminaron con ruido y reverberación como se detalla en la sección 4.1.5. En estas pruebas se utilizó ruido no correlacionado y reverberación correlacionada, debido a que en las pruebas anteriores se concluyó que bajo esas condiciones se obtenía menor inteligibilidad.

La Tabla 4-8 muestra el orden en que se presentaron las pistas de voz considerando el ángulo, y las diferentes condiciones acústicas (limpia, ruido y reverberación) utilizadas en cada una de las tres sesiones de la prueba, además se muestran los valores de la correlación interaural (IACC) para cada una de las listas utilizadas.

El orden de las pistas de voz se debe a una selección aleatoria considerando los diferentes ambientes acústicos estudiados y los ángulos.

Tabla 4-8: Pistas de voz utilizadas en las tres sesiones, indicando la pista, el ángulo la condición acústica (limpia, ruido o reverberación), y los valores de la correlación interaural (IACC).

Sesión 1			Sesión 2			Sesión 3		
Ángulo (°)	Condición acústica	IACC	Ángulo (°)	Condición acústica	IACC	Ángulo (°)	Condición acústica	IACC
15	L4/Ruido	0.2345	30	L3/Rever	0.9085	-30	L1/Ruido	0.1975
-30	L3/Rever	0.9123	-15	L4/Ruido	0.2279	45	L4/Rever	0.8718
-45	L1/Limpia	0.8301	45	L1/Limpia	0.8237	30	L2/Limpia	0.9052
45	L2/Ruido	0.1646	60	L2/Rever	0.7805	60	L3/Ruido	0.1260
60	L4/Limpia	0.7448	-60	L3/Ruido	0.1228	-15	L1/Rever	0.9380
-15	L3/Limpia	0.9155	-45	L4/Ruido	0.1571	-60	L4/Limpia	0.7436
-60	L2/Rever	0.7850	-30	L2/Limpia	0.9050	15	L3/Limpia	0.9263
30	L1/Ruido	0.2034	15	L1/Rever	0.8621	-45	L2/Rever	0.8550

La Figura 4-12 muestra los valores IACC en función del ángulo acimutal bajo condiciones de ruido no correlacionado y reverberación correlacionada.

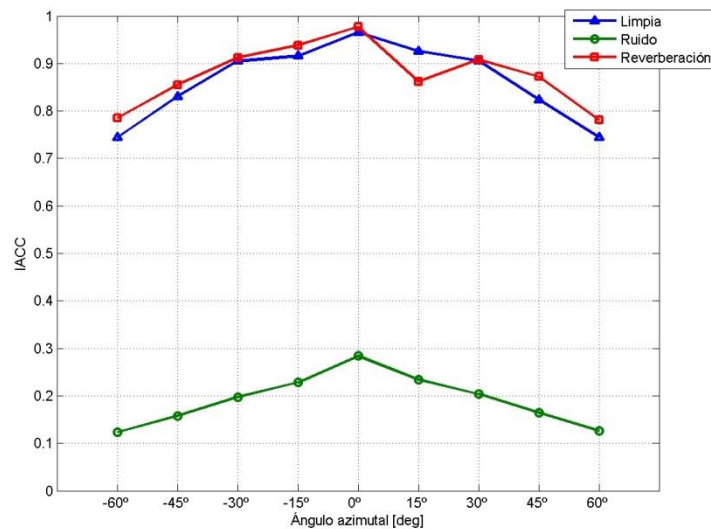


Figura 4-12: Valores de la correlación interaural (IACC) de las pistas de voz bajo condiciones de ruido no correlacionado y reverberación correlacionada en función del ángulo acimutal.

4.3.5 Desarrollo de la prueba

La prueba se realizó en el interior de la cámara anecoica del CCADET y se dividió en tres sesiones de aproximadamente 45 minutos cada una. En estas pruebas también se les solicitó a los participantes que señalaran la ubicación donde ellos creían que provenía la señal de voz. Este experimento también se realizó con audífonos. Para esta tarea, se les mostró un diagrama en el que se les presentaron 13 posibles posiciones de la fuente, que iba desde -90° hasta $+90^\circ$ en intervalos de 15° (Figura 4-13).

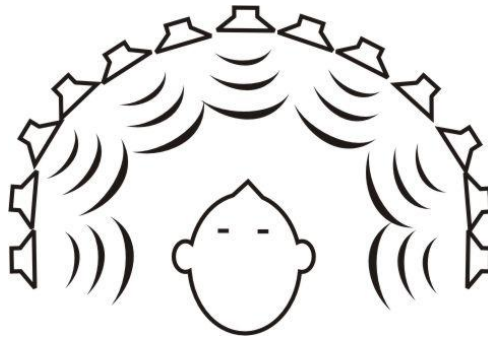


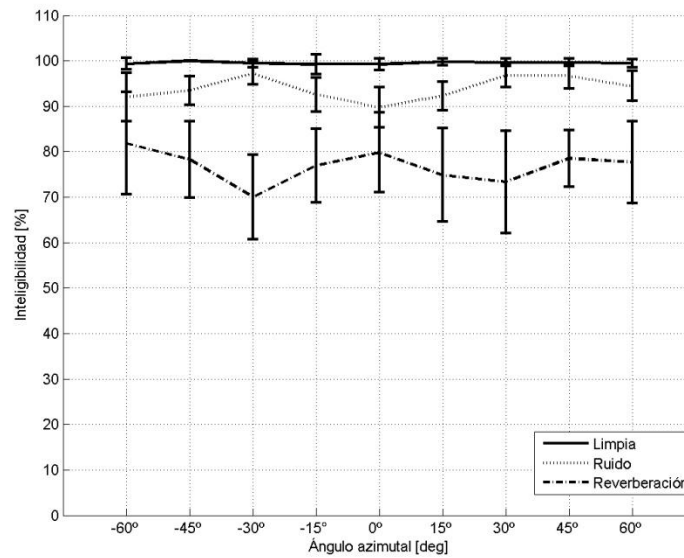
Figura 4-13: Diagrama mostrado a los participantes durante la prueba de inteligibilidad con el objetivo de que señalaran la posición de la cual creían que provenía la señal de voz.

4.3.6 Resultados

La Figura 4-14 muestra los porcentajes de inteligibilidad obtenidos a distintos ángulos, los resultados muestran una clara influencia del ángulo acimutal, especialmente cuando la señal de voz se presenta bajo condiciones acústicas adversas. Como se esperaba, cuando la señal de voz se presenta sin ningún disturbio, el porcentaje de inteligibilidad es alto para todos los ángulos estudiados.

Bajo condiciones de ruido no correlacionado, la inteligibilidad es mayor cuando el orador tiene una orientación de $+30^\circ$ y -30° con 96.6% y 97.2%, respectivamente; mientras que el porcentaje más bajo se obtiene a 0° (89.7%). Se observa un comportamiento simétrico de los porcentajes de inteligibilidad para los ángulos positivos y negativos.

Bajo condiciones de reverberación correlacionada, la inteligibilidad es mayor cuando el orador está orientado a -60° y -45° , con 81.8% y 78.2%, respectivamente, mientras que el porcentaje más bajo se obtiene para -30° y $+30^\circ$, con 70% y 73.3%, respectivamente.



	-60°	-45°	-30°	-15°	0°	15°	30°	45°	60°
Limpia	100 (1.2)	99 (0.0)	100 (0.9)	100 (2.1)	99.2 (1.2)	99 (0.7)	100 (0.8)	99 (0.8)	100 (0.9)
Ruido	92 (5.3)	93.4 (3.1)	97.2 (2.4)	92.5 (3.7)	89.7 (4.4)	92.2 (3.1)	96.6 (2.4)	96.6 (2.8)	94.4 (3.3)
Reverberación	81.8 (11.1)	78.2 (8.4)	70 (9.2)	76.9 (8.1)	79.8 (8.8)	74.8 (10.2)	73.3 (11.2)	78.5 (6.2)	77.6 (8.9)

Figura 4-14: Porcentaje de inteligibilidad a diferentes ángulos. La barra de error muestra la desviación estándar. En la tabla se muestra el promedio y la desviación estándar se muestra entre paréntesis.

La Tabla 4-9 muestra el número de palabras que fueron entendidas por al menos el 33% de los sujetos, y las cuales fueron utilizadas para el análisis de los resultados.

Tabla 4-9: Presentación de las señales bajo condiciones de ruido y reverberación para diferentes ángulos y número de palabras utilizadas en el análisis de los resultados.

Ángulo (°)	Ruido	Reverberación
	Número de palabras	Número de palabras
0	44	40
15	48	44
30	50	35
45	48	45
60	49	40
-15	50	46
-30	50	30
-45	50	46
-60	50	43

4.3.6.1 Localización de fuentes

La Figura 4-15 muestra los resultados obtenidos para la prueba de localización con las pistas de voz limpia, es decir, la señal de voz no estaba bajo ninguna condición acústica adversa (ruido y/o reverberación).

Los resultados muestran que hay una clara y correcta distinción entre el lado derecho e izquierdo. Sin embargo, en cada lado, las respuestas son dispersas en gran medida, además hay una tendencia a lateralizar demasiado la percepción del ángulo. Los ángulos que están ligeramente fuera del centro, ya sea a la derecha o izquierda ($\pm 15^\circ$), tienden a ser muy lateralizados, llegando incluso a reportarse en los ángulos extremos de $\pm 90^\circ$. El ángulo con mayor número de aciertos es $+60^\circ$; mientras que los ángulos con menor número de aciertos es a $\pm 15^\circ$. La línea diagonal corresponde a una indicación correcta del ángulo percibido.

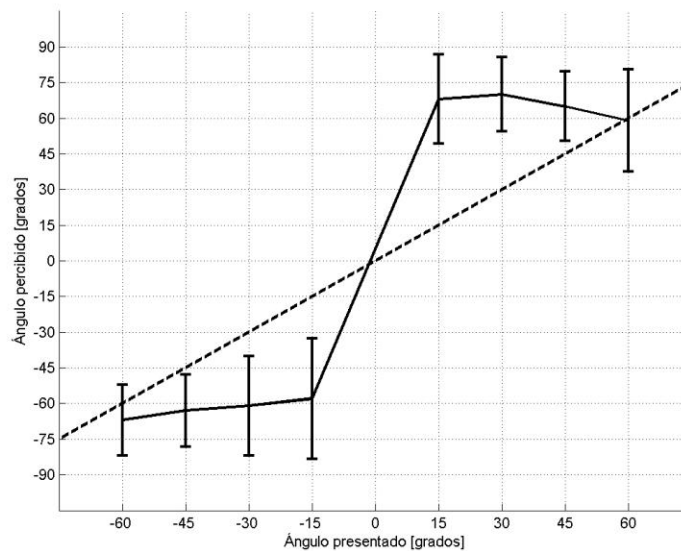


Figura 4-15: *Ángulos percibidos con la señal de voz limpia. En el eje de las abscisas se muestran los ángulos presentados al oyente, y en el eje de las ordenadas los ángulos percibidos por el oyente. Se muestra la desviación estándar para cada ángulo.*

4.3.7 Análisis de los resultados

Los resultados obtenidos en esta prueba son consistentes y es posible explicarlos en términos del modelo binaural de Ecuación – Cancelación (EC). Bajo condiciones de ruido no correlacionado, la etapa de cancelación del ruido es casi imposible de lograr, por lo que el sistema auditivo obtiene el máximo beneficio de la señal de voz recibida por ambos oídos para ángulos laterales intermedios

($\pm 30^\circ$), que es donde el espectro de voz es diferente para cada uno de los oídos, dada la cercanía de uno de ellos a la fuente sonora (a diferencia del ángulo a 0° donde las señales de voz son iguales para ambos oídos). Sin embargo, estas diferencias no son tan extremas como ocurre para el ángulo de $\pm 60^\circ$.

Bajo condiciones de reverberación correlacionada, ocurre el efecto contrario, esto es, el sistema auditivo reduce la inteligibilidad de la voz para ángulos laterales intermedios ($\pm 30^\circ$), donde el nivel de la señal de voz reverberada para cada uno de los oídos es diferente; sin embargo, la relación interaural de fase es constante para ambos oídos (debido a la forma en que se generó la respuesta a impulso para reverberar la señal de voz), de este modo es posible que en la etapa de cancelación se eliminara parte de la señal de voz. También se observa que los porcentajes más altos de inteligibilidad se obtienen cuando no hay ventaja binaural, es decir, para ángulos alrededor de 0° y de $\pm 60^\circ$, para el cual la diferencia interaural de nivel es más evidente, y está dominada por el nivel recibido por el oído que está ubicado más cerca del orador, por lo que el beneficio binaural se reduce casi a un caso monoaural.

Los resultados anteriores muestran que la lateralización del orador reduce la inteligibilidad binaural bajo condiciones de reverberación correlacionada. Sin embargo, es importante mencionar que esta condición es poco común encontrarla en situaciones prácticas. Lo anterior se explica con base en los principios de la acústica estadística de recintos [Pierce, 1981; Kuttruff, 2009]: una mayor reverberación implica un campo sonoro difuso en el cual la correlación espacial de dicho campo, en particular la correlación interaural (IACC), es pequeña en el intervalo de medias y altas frecuencias (alrededor y por arriba de los 1000 Hz). Sin embargo, bajo otras condiciones, las cuales incluyen: reverberación no correlacionada, ruido correlacionado y no correlacionado, se ha demostrado que la lateralización del orador generalmente ayuda a mejorar la inteligibilidad de la voz. Es posible encontrar estas condiciones acústicas en muchas situaciones prácticas, de modo que la lateralización del orador generalmente mejora la inteligibilidad de la voz bajo condiciones acústicas adversas.

4.3.8 Conclusiones

Se llevaron a cabo pruebas subjetivas de inteligibilidad a diferentes ángulos para determinar la posición en la cual el oyente recibe de manera más efectiva el mensaje bajo condiciones acústicas adversas (ruido y reverberación), con disturbios interauralmente correlacionados y no correlacionados. Bajo condiciones limpias de audición se obtuvo un porcentaje alto de

inteligibilidad (casi 100% de palabras entendidas correctamente) para todos los ángulos; sin embargo bajo condiciones de ruido o reverberación, la posición del orador respecto al oyente tiene una influencia significativa en la inteligibilidad. Bajo condiciones de ruido no correlacionado interauralmente la inteligibilidad mejora cuando el oyente no está directamente de frente al orador. Se obtiene un porcentaje máximo para ángulos alrededor de los $\pm 30^\circ$; mientras que bajo condiciones de reverberación interauralmente correlacionada la inteligibilidad disminuye alrededor de estos mismo ángulos ($\pm 30^\circ$), y mejora para ángulos alrededor de los 0° y $\pm 60^\circ$.

4.4 Prueba 4: Prueba de inteligibilidad con HRTF personalizadas y no personalizadas

4.4.1 Objetivo

- Evaluar el posible beneficio que se obtiene mediante el uso de HRTF individualizadas contra HRTF no individualizadas en la inteligibilidad de la palabra, así como la localización de fuentes.

4.4.2 HRTF individualizadas

4.4.2.1 Medición de las HRTF

Se llevaron a cabo mediciones de las HRTF en 7 sujetos a diferentes ángulos desde -60° a $+60^\circ$ en intervalos de 15° (Figura 4-10 derecha). Las mediciones se llevaron a cabo en la Cámara Semi-Anecoica del SRCM-Intel.

Se generó una señal de barrido lineal en frecuencia (*chirp*), la cual se reprodujo a través de un altavoz amplificado marca Mackie modelo HR624, para la grabación de la señal se utilizaron audífonos biaurales marca Head Acoustics modelo BHM III.2 (Figura 4-16). La reproducción y adquisición de los datos se realizó a través del software Analyzer ACQUA. La frecuencia de muestreo utilizada fue $f_s = 44100 \text{ Hz}$.

El micrófono utilizado para las mediciones se coloca sobre la cabeza (similar a un audífono de diadema), viene equipado con un par de conos de plástico, los cuales se colocan en la entrada del conducto auditivo del sujeto. Por cuestiones de higiene los conos fueron reemplazados por unos conos limpios para cada medición.



Figura 4-16: Micrófono binaural BHM III.2 empleado para las mediciones de las HRTF.

Para realizar las mediciones de las HRTF, se ubicaron los sujetos en el centro de la cámara sentados en una silla giratoria, se les pidió estar con la espalda recta y recargada en el respaldo de la silla, se colocaron los micrófonos y se les pidió que giraran. Con ayuda de un transportador y un hilo adherido a la diadema del micrófono se verificó la distancia y el ángulo.

Los conos del micrófono se posicionan en la entrada del conducto auditivo, por lo que las mediciones se realizan con el conducto auditivo parcialmente bloqueado, esto tiene ciertas ventajas, dado que el conducto auditivo se puede considerar como un tubo cerrado por un extremo con unas dimensiones promedio de 25 mm longitud y 8 mm de diámetro [Hammersøi y Møller, 1996], lo que genera una frecuencia de resonancia alrededor de los 3 kHz. Provocando la coloración de los sonidos que se escuchan en esa frecuencia. Las ventajas que presenta realizar las mediciones de HRTF bajo estas condiciones se detallan en [Møller *et al.*, 1995].

4.4.2.2 Participantes

Participaron 7 sujetos con características similares a las pruebas anteriores. La edad de los participantes oscila entre los 26 y 33 años de edad, con una edad promedio de 29.7 años.

4.4.2.3 Procesamiento de las señales de voz

Una vez medidas las HRTF para cada uno de los ángulos y para todos los sujetos, se *convolucionaron* con la señal de voz, como se indica en la ecuaciones (4-10) y (4-14). Después, de forma independiente se contaminó la señal resultante con ruido y reverberación, tal como se muestra en la Figura 4-17.

$$S_L(\theta) = S_L(t) * HRTF_L(\theta) \quad (4-13)$$

$$S_R(\theta) = S_R(t) * HRTF_R(\theta) \quad (4-14)$$

Donde $S_L(t), S_R(t)$ son las señales biaurales grabadas en el oído izquierdo y derecho, respectivamente, $HRTF_L(\theta), HRTF_R(\theta)$ son las funciones de transferencia referidas a la cabeza medidas en el oído izquierdo y derecho, respectivamente.

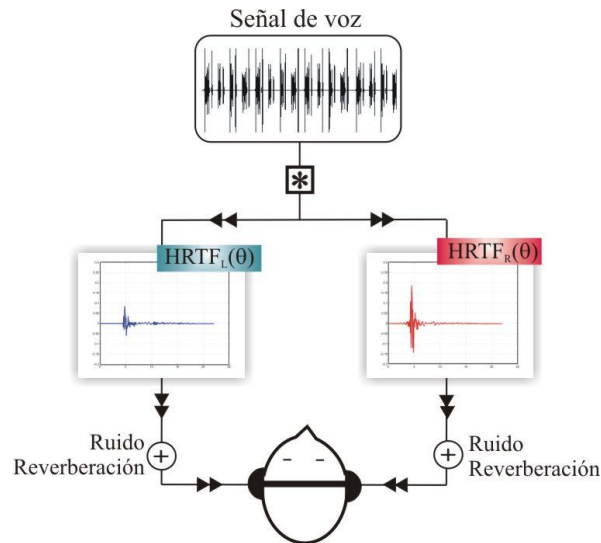


Figura 4-17: La señal de voz se convoluciona con cada una de las HRTF medidas para cada sujeto y para diferentes ángulos.

4.4.2.4 Desarrollo de la prueba

Las pruebas se dividieron en 4 sesiones. En la primera sesión se realizaron las mediciones de las HRTF, con una duración aproximada de 30 minutos. En las tres últimas sesiones se desarrollaron las pruebas de inteligibilidad y cada una tuvo una duración de 45 minutos aproximadamente. Todas las sesiones se realizaron en el interior de la cámara semi-anecoica. Los sujetos escuchaban la señal de voz convolucionada con sus propias HRTF medidas en cada uno de los ángulos.

Los sujetos escucharon las pistas de voz a través de unos audífonos de tipo intraural, marca Panasonic modelo RP-HJE100; por razones de higiene, y debido a que se trata de audífonos que se insertan en el conducto auditivo, a cada voluntario se le entregaron unos audífonos para su uso personal.

4.4.3 HRTF no individualizadas

Para esta prueba se utilizaron las respuestas impulso contenidas en una base de datos pública de mediciones de HRTF, realizadas por CIPIC Interface Laboratory [Algazi *et al.*, 2001]. Esta base de datos contiene mediciones realizadas en 45 sujetos en 1250 posiciones distintas y para ambos oídos. Los ángulos acimutales θ medidos fueron: -80° , -65° , -55° , desde -45° a $+45^\circ$ en intervalos de 5° , $+55^\circ$, $+65^\circ$ y $+80^\circ$, 25 puntos en total; el ángulo de elevación ϕ varía de -45° a $+230.625^\circ$ en intervalos de 5.625° , 50 puntos en total.

La ubicación de la fuente sonora se especifica por el ángulo acimutal θ y el ángulo de elevación ϕ , es decir (θ, ϕ) . La coordenada $(0^\circ, 0^\circ)$ corresponde a la posición justo frente al sujeto.

Para realizar las mediciones, el sujeto se ubicó sentado en el centro de un aro de un metro de radio, se utilizaron micrófonos de prueba marca Etymotic Research ER-7C. Las mediciones se hicieron con el conducto auditivo bloqueado y se obtuvieron 200 muestras de la respuesta impulso. La señal de excitación fue una señal de ruido pseudo aleatoria. La frecuencia de muestreo utilizada fue de $f_s = 44.1 \text{ kHz}$.

Esta base de datos cuenta además con información de mediciones antropométricas de 35 de los 45 sujetos. Diecisiete mediciones de la cabeza y torso y diez realizadas en las orejas. Incluye también información sobre género, edad, estatura y peso de los participantes.

Para este trabajo de tesis no se consideraron las mediciones antropométricas (cabeza, torso y orejas), como ya se explicó, estas medidas son particularmente de interés para la tesis doctoral mencionada en el inicio de este trabajo, la cual está actualmente en desarrollo y que complementa este proyecto de investigación. En este trabajo nos enfocamos principalmente en las diferencias de peso, estatura, edad y sexo, para asegurar diferencias significativas entre los sujetos donde fueron medidas las HRTF y los sujetos que participaron en las pruebas de inteligibilidad.

Se eligieron 7 funciones de transferencia medidas en personas con ciertas características anatómicas, para después aplicar estas funciones a personas con una anatomía completamente diferente. Los datos del lado izquierdo de la Tabla 4-10 corresponden a las características de las personas en la cuales se midieron las funciones de transferencia elegidas de la base de datos; mientras que las medidas del lado derecho corresponden a los sujetos que participaron en las pruebas de inteligibilidad.

Tabla 4-10: Características de las personas elegidas de la base de datos (lado izquierdo) y de las personas que participaron en la prueba (lado derecho).

	Sexo	Edad	Peso (Kg)	Estatura (m)		Sexo	Edad	Peso (Kg)	Estatura (m)
1	F	25	39.5	1.47	➔	M	29	60	1.76
2	F	35	66	1.67		M	30	80	1.76
3	M	20	118	2.00		M	32	73	1.70
4	F	21	50	1.54		M	26	78	1.77
5	M	63	77	1.83		M	33	63	1.75
6	F	30	93	1.78		M	27	72	1.74
7	F	21	48.5	1.52		M	31	69	1.74

De la tabla anterior se observa que hay diferencias significativas entre los sujetos que fueron medidos y a los que se aplicó la prueba.

4.4.3.1 Participantes

Participaron 7 sujetos con características similares a las pruebas anteriores. La edad de los participantes oscila entre los 22 y 27 años de edad, con una edad promedio de 25.5 años.

4.4.3.2 Desarrollo de la prueba

Esta prueba se dividió en dos sesiones de, aproximadamente, 45 minutos cada una. A cada sujeto se le asignaron las mediciones de HRTF de un sujeto en particular de la base de datos (ver Tabla 4-10). Al igual que en la prueba anterior, se utilizaron audífonos intraurales, marca Panasonic modelo RP-HJE100.

4.4.3.3 Resultados

En la Figura 4-18 se muestran las HRTF medidas para este trabajo de tesis, para los 7 sujetos; mientras que en la Figura 4-19 y Figura 4-20 se muestran las HRTF tomadas de la base de datos CIPIC.

4Prueba 4: HRTF individualizadas vs. HRTF no-individualizadas

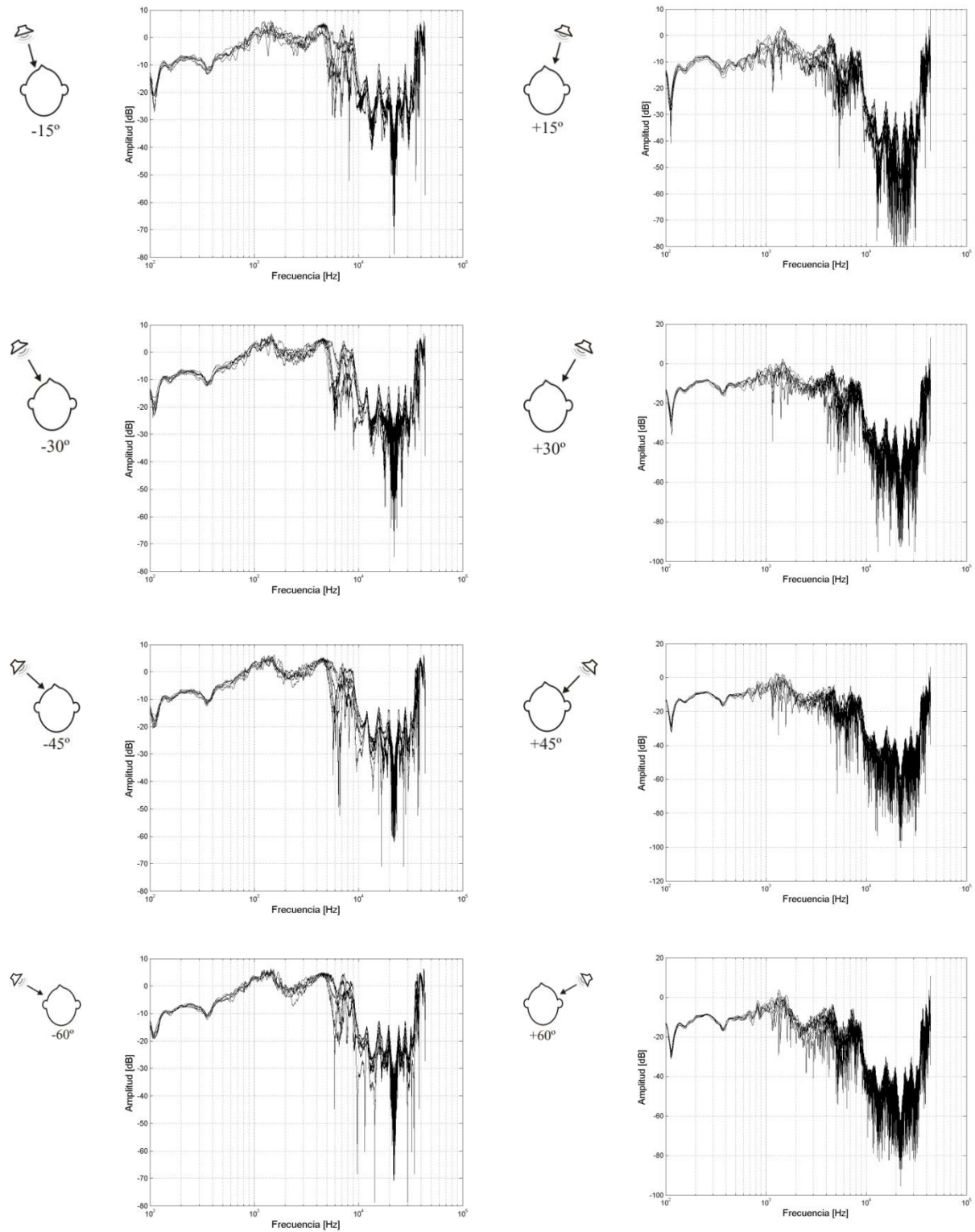


Figura 4-18: HRTF medidas en 7 sujetos en una cámara semi-anechoica, para los ángulos acimutales de -15° a -60° y de 15° a 60° en intervalos de 15° , en todos los casos el ángulo de elevación fue de 0° .

4Prueba 4: HRTF individualizadas vs. HRTF no-individualizadas

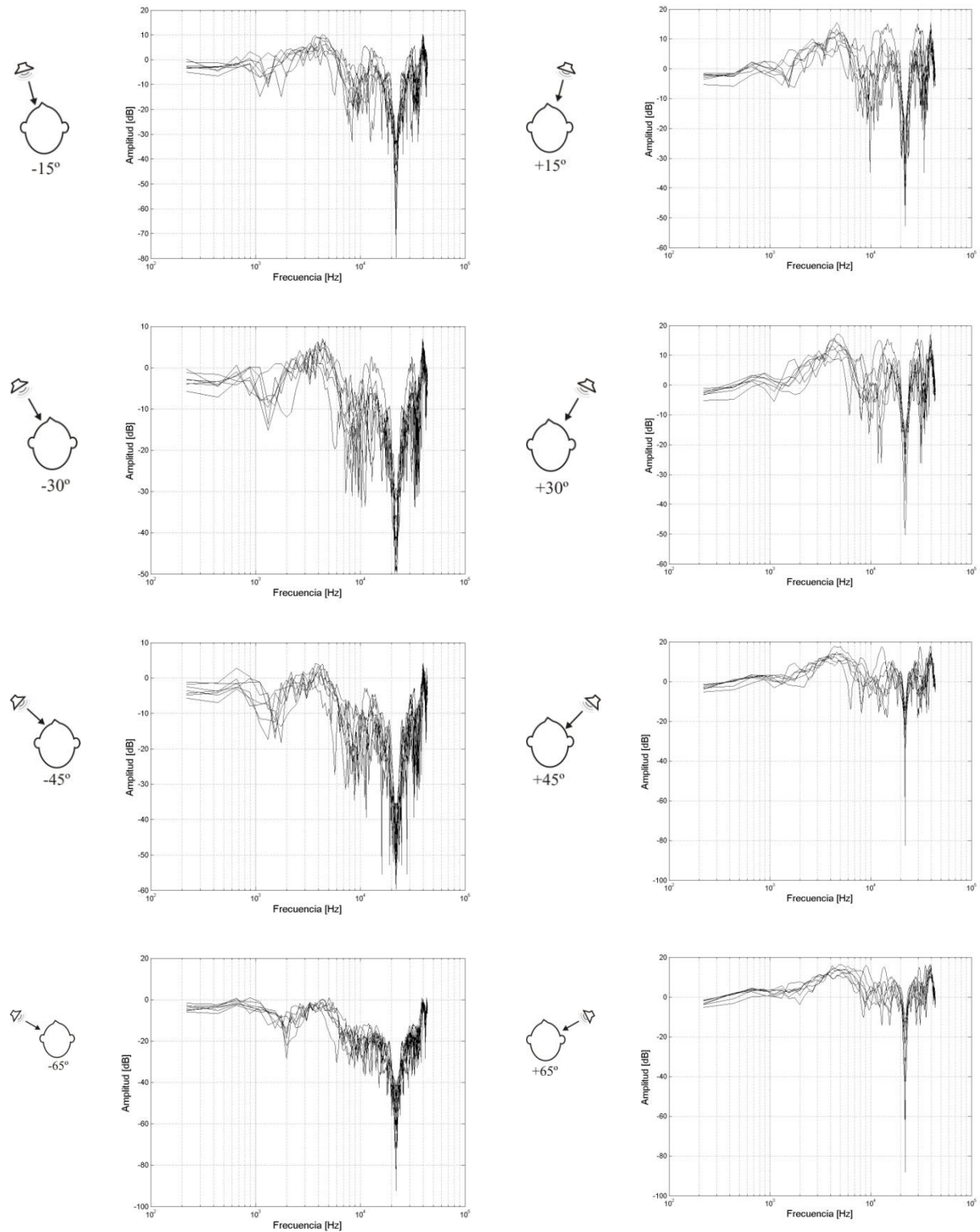


Figura 4-19: HRTF medidas por CIPIC Lab. Para los ángulos de -15° a -65° y de 15° a 65° en intervalos de 15° para el oído derecho.

4Prueba 4: HRTF individualizadas vs. HRTF no-individualizadas

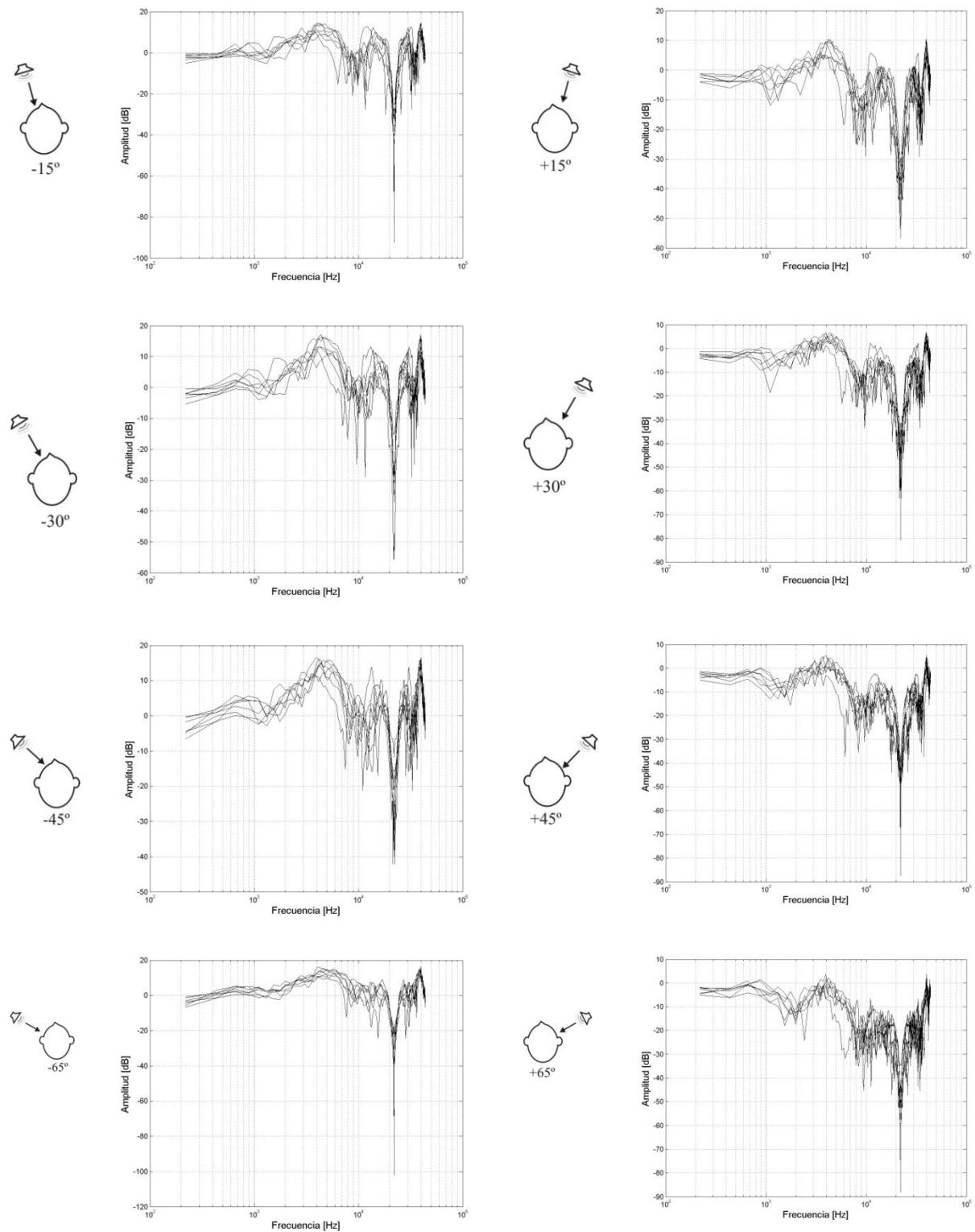


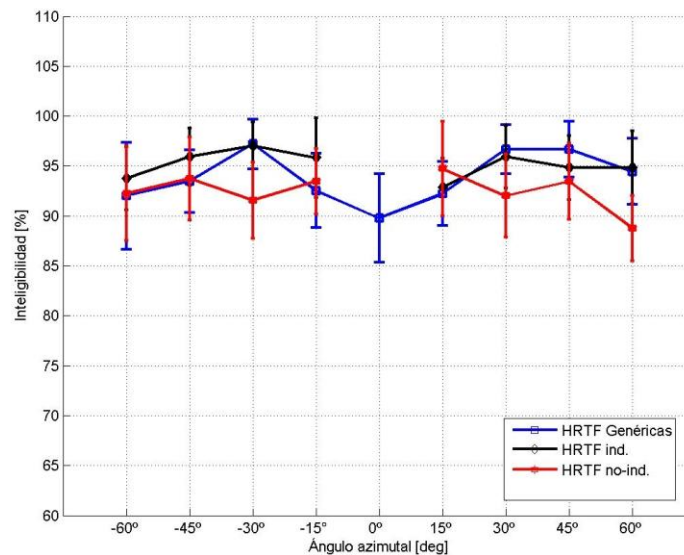
Figura 4-20: HRTF medidas por CIPIC Lab. Para los ángulos de -15° a -65° y de 15° a 65° en intervalos de 15° para el oído izquierdo.

4.4.3.3.1 HRTF individualizadas vs. HRTF no individualizadas bajo condiciones de ruido

Para facilitar la lectura de las gráficas, los resultados se muestran de forma independiente. Primero se muestran los resultados obtenidos bajo condiciones de ruido (Figura 4-21) y los obtenidos bajo condiciones de reverberación (Figura 4-22).

Se observa que si se utilizan HRTF individualizadas comparadas con las HRTF no individualizadas se obtiene una discreta mejoría, esto para la mayoría de los ángulos estudiados, a excepción del ángulo -15° , donde se obtuvo un decremento de 1.8% al utilizar HRTF personalizadas. Los ángulos que presentan una ventaja mayor son -30° y 60° , con 5.4% y 6%, respectivamente.

Los resultados de estas pruebas coinciden con lo publicado por [Drullman y Bronkhorst, 2000], quienes encontraron que no existe una diferencia significativa en la inteligibilidad, al utilizar HRTF individualizadas y no individualizadas. Estos resultados fueron encontrados para señales de voz con banda limitada a 4 kHz. Por otro lado, hay investigaciones que demuestran la importancia del uso de HRTF personalizadas principalmente en la localización de fuentes [Kistler *et al.*, 1992; Wenzel *et al.*, 1993; Møller *et al.*, 1995; Xie, 2002].



	-60°	-45°	-30°	-15°	0°	15°	30°	45°	60°
HRTF Genéricas	92 (5.3)	93.4 (3.15)	97.2 (2.4)	92.5 (3.7)	89.7 (4.4)	92.2 (3.1)	96.6 (2.4)	96.6 (2.8)	94.4 (3.3)
HRTF ind	93.7 (3.6)	95.9 (3.2)	97 (3.1)	95.8 (2.9)		92.8 (2.9)	95.9 (3.1)	94.8 (3.2)	94.8 (3.6)
HRTF no-ind	92.2 (3.6)	93.7 (4.4)	91.5 (4.6)	93.4 (4.1)		94.7 (4.1)	92 (4.6)	93.4 (4.4)	88.7 (3.6)

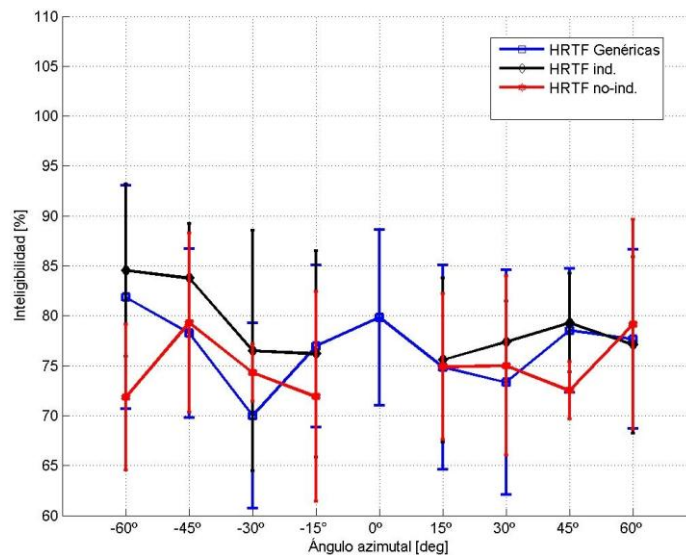
Figura 4-21: Porcentajes de inteligibilidad obtenidos con HRTF individualizadas (línea en color negro), HRTF no individualizadas (línea continua en color rojo) y HRTF genéricas (línea en color azul) bajo condiciones de ruido. La barra de error indica la desviación estándar. La tabla muestra el promedio y la desviación estándar entre paréntesis.

4.4.3.3.2 HRTF individualizadas vs. HRTF no individualizadas bajo condiciones de reverberación

En la Figura 4-22 se muestran los porcentajes de inteligibilidad obtenidos con HRTF individualizadas y no individualizadas bajo condiciones de reverberación. En la tabla de la figura se muestra el promedio y la desviación estándar entre paréntesis.

En este caso, la inteligibilidad presenta una ventaja muy variable utilizando HRTF individualizadas, incluso para algunos ángulos la ventaja se obtiene utilizando HRTF no individualizadas, como por ejemplo, cuando el orador está ubicado a 60°.

Utilizando HRTF individualizadas se observa una ventaja de 12.7%, 4.48% y 6.7%, comparando con las HRTF no individualizadas, para los ángulos -60°, -45° y +45°, respectivamente.



	-60°	-45°	-30°	-15°	0°	15°	30°	45°	60°
HRTF Genéricas	81.8 (11.1)	78.2 (8.4)	70 (9.2)	76.9 (8.1)	79.8 (8.8)	74.8 (10.2)	73.3 (11.2)	78.5 (6.2)	77.6 (8.9)
HRTF ind	84.5 (8.6)	83.7 (5.4)	76.4 (12)	76.1 (10.3)		75.5 (8.2)	77.3 (4.0)	79.2 (4.9)	77 (8.8)
HRTF no-ind	71.8 (7.2)	79.2 (8.9)	74.2 (2.8)	71.8 (10.4)		74.9 (10)	75 (4.5)	72.5 (10.7)	79.1 (8.0)

Figura 4-22: Porcentajes de inteligibilidad obtenidos con HRTF individualizadas (línea en color negro), HRTF no individualizadas (línea continua en color rojo) y HRTF genéricas (línea en color azul) en presencia de reverberación. La barra de error indica la desviación estándar. La tabla muestra el promedio y la desviación estándar entre paréntesis.

La Tabla 4-11 muestra el número de palabras que fueron entendidas por al menos el 33% de los sujetos, y las cuales fueron utilizadas para el análisis de los resultados.

Tabla 4-11: Número de palabras utilizadas en el análisis de los resultados para HRTF ind. y HRTF no-ind., a diferentes ángulos y bajo condiciones de ruido y reverberación

Ángulo (°)	Ruido	Reverberación
	Número de palabras	Número de palabras
	HRTF ind.	
15	48	45
30	49	36
45	47	40
60	47	38
-15	48	48
-30	48	31
-45	49	44
-60	48	37
	HRTF no-ind.	
15	46	37
30	50	36
45	48	40
60	47	37
-15	48	31
-30	49	35
-45	48	40
-60	48	37

4.4.3.3 Intervalos de confianza

La Tabla 4-12 muestra los intervalos de confianza calculados para las pruebas de inteligibilidad, utilizando HRTF individualizadas y no individualizadas bajo condiciones de ruido y reverberación.

Tabla 4-12: Intervalos de confianza del porcentaje de inteligibilidad utilizando HRTF individualizadas y no individualizadas bajo condiciones de ruido y reverberación. En negritas se muestra el promedio y entre paréntesis los intervalos de confianza con un nivel de confianza del 95%.

RUIDO							
HRTF individualizada							
-60°	-45°	-30°	-15°	15°	30°	45°	60°
93.7	95.9	97	95.8	92.8	95.9	94.8	94.8
(96.1, 91.3)	(98, 93.7)	(98.7, 95.2)	(98.7, 92.8)	(95, 90.7)	(98.2, 93.6)	(97.2, 92.4)	(97.5, 92.1)
HRTF no individualizada							
92.2	93.7	91.5	93.4	94.7	92	93.4	88.7
(95.7, 88.7)	(96.8, 90.6)	(94.3, 88.7)	(95.8, 91)	(97.7, 91.6)	(95.4, 88.5)	(96.7, 90.1)	(91.4, 86)

REVERBERACIÓN							
HRTF individualizada							
-60°	-45°	-30°	-15°	15°	30°	45°	60°
84.5	83.7	76.4	76.1	75.5	77.3	79.2	77
(90.9, 78.1)	(87.8, 79.7)	(85.4, 67.5)	(83.8, 68.5)	(81.6, 69.4)	(80.3, 74.3)	(82.9, 75.6)	(83.6, 70.5)
HRTF no individualizada							
71.8	79.2	74.2	71.8	74.9	75	72.5	79.1
(77.2, 66.4)	(85.9, 72.6)	(76.4, 72.1)	(79.6, 64.1)	(82.3, 67.4)	(78.3, 71.6)	(80.4, 64.5)	(85.1, 73.1)

4.4.4 Análisis de los resultados

El hecho de utilizar HRTF individualizadas ayuda a mejorar la inteligibilidad bajo condiciones acústicas adversas. Por otro lado, al utilizar HRTF no individualizadas hay un decremento en la inteligibilidad, esto se debe a que las características anatómicas de las personas en las que se midieron las funciones de transferencia son completamente diferentes a las de las personas que participaron en las pruebas de inteligibilidad.

4.4.5 Conclusiones

Se llevaron a cabo pruebas subjetivas de inteligibilidad para evaluar el beneficio que se obtiene mediante el uso de HRTF individualizadas comparado con HRTF no individualizadas. Se midieron HRTF en 7 sujetos a diferentes ángulos en el plano horizontal; mientras que las HRTF no individualizadas se tomaron de una base de datos pública disponible en internet. Las señales de voz fueron filtradas con las HRTF individualizadas y no individualizadas. Los resultados demuestran que el uso de las HRTF individualizadas ayuda a mejorar la inteligibilidad de la voz bajo condiciones acústicas adversas de audición.

5 CONCLUSIONES

Esta investigación doctoral forma parte de un proyecto de investigación de mayor alcance, desarrollado en el Grupo de Acústica y Vibraciones del CCADET-UNAM en colaboración con Intel Tecnología de México S.A. de C.V., cuyo objetivo general es aplicar técnicas de reproducción de sonido binaural en dispositivos móviles para mejorar la inteligibilidad de la voz, a través del uso de HRTF individualizadas para cada usuario mediante procesamiento de imágenes antropométricas. Los resultados obtenidos específicamente en la presente investigación doctoral aportan información útil sobre el efecto que tienen las técnicas de reproducción de sonido binaural en la inteligibilidad de la voz, considerando diferentes aspectos, tales como: condiciones adversas de ruido y reverberación, la correlación interaural del disturbio acústico, así como la influencia del ángulo de incidencia de la señal de voz y el uso de HRTF individualizadas.

En esta investigación se llevaron a cabo cuatro pruebas de inteligibilidad de la voz reproducida a través de audífonos, cada una con objetivos específicos. El objetivo de la primera prueba fue cuantificar el beneficio que se obtiene al utilizar técnicas de grabación y reproducción binaural, contra técnicas estándares de reproducción monofónicas o estereofónicas. El objetivo de la segunda prueba fue investigar el efecto de la correlación interaural (IACC) en la inteligibilidad binaural de la voz. En la tercera prueba se evaluó la inteligibilidad a diferentes ángulos acimutales (plano horizontal). El objetivo de la última prueba fue evaluar el posible beneficio que se obtiene mediante el uso de funciones de transferencia referidas a la cabeza (HRTF) individualizadas, contra HRTF no individualizadas, en un sistema de reproducción de sonido binaural a través de audífonos.

Con el objetivo de presentar condiciones de audición excesivamente nocivas, que garantizaran una pérdida significativa de la inteligibilidad y con el objetivo de simular ambientes acústicos adversos, las pruebas se realizaron bajo condiciones extremas de ruido y reverberación. Para el ruido, se consideró una relación de señal a ruido de $SNR = -10\text{dB}$; mientras que el tiempo de reverberación fue de $T_{60} = 10$ s. Estos disturbios acústicos fueron agregados de manera artificial a las señales de voz grabadas previamente.

Las conclusiones de este trabajo se mencionan a continuación:

La inteligibilidad de la voz es afectada por diversos factores, como el ruido y la reverberación; sin embargo, incluso dentro de estos ambientes poco favorables para la comunicación oral, es posible

mejorar la inteligibilidad a través de la aplicación de técnicas de reproducción binaural, como lo demuestran los resultados de esta investigación, algunos de los cuales confirman resultados ya publicados y otros son originales, según se indica más adelante.

Algunos otros factores que logran mejorar la inteligibilidad (considerando la reproducción del sonido de manera binaural a través de audífonos), son:

- Modificar el ángulo de incidencia de la señal de voz, es decir, que el oyente no reciba la señal de voz desde una dirección frontal, sino que la ubicación de la fuente de voz sea desde un ángulo lateral específico en el plano horizontal. En este trabajo quedó demostrado que para los ángulos $\pm 30^\circ$ la inteligibilidad tiene una mejoría significativa (comparando con el porcentaje de inteligibilidad obtenido a 0°), de 7.2%, considerando condiciones de ruido (Ver Sección 4.3.6).
- Si el disturbio acústico que reciben ambos oídos es igual será posible cancelar la señal no deseada. Lo anterior quedó demostrada con los resultados obtenidos para la presentación monofónica bajo condiciones de ruido correlacionado y no correlacionado.
- Utilizar funciones de transferencia referidas a la cabeza (HRTF) personalizadas en la reproducción binaural de las señales de voz.

Las pruebas subjetivas de inteligibilidad son una herramienta útil para evaluar la calidad acústica de la voz, ya sea en un recinto o bien en este caso, dirigido a dispositivos móviles. Un aspecto conveniente de estas pruebas es que permiten conocer la experiencia de la persona involucrada en la prueba. Por lo que, para la evaluación de las tecnologías binaurales propuestas en este trabajo, las pruebas subjetivas son de mucha utilidad ya que proporcionan información directamente del usuario, respecto a la aplicación de estas tecnologías. Sin embargo, se puede también considerar la implementación de pruebas objetivas, basadas según los investigadores [Schroeder, 1981; Houtgast y Steeneken, 1985a; Houtgast y Steeneken, 1985b] en la medición física de niveles de ruido, tiempos de reverberación, correlación interaural, para realizar comparaciones.

Este trabajo aporta evidencias de que las técnicas binaurales de reproducción de sonido a través de audífonos son una buena estrategia para mejorar la calidad acústica en dispositivos móviles de comunicación.

Es posible aplicar los resultados de este trabajo a sistemas de audio ó videoconferencias, con la condición de que utilicen audífonos como medio de reproducción, para que las condiciones de espacialización se mantengan y por otro lado, evitar el efecto de cruce de canales (*crosstalk*) que es muy común en la reproducción a través de altavoces, el cual consiste en que la señal que es emitida por el altavoz derecho llegue también al oído izquierdo y viceversa.

Las aportaciones originales de este trabajo se mencionan a continuación:

- Se utilizó material de voz en español en pruebas de inteligibilidad biaural a través de audífonos, este material está conformado por palabras bisílabas con significado, que en la literatura especializada existe de forma muy limitada [Salinas, 2007]. La ventaja de utilizar palabras de este tipo, es que se presenta un panorama realista en las pruebas realizadas, ya que la comunicación en cualquier sistema de comunicación se realiza a través de palabras con significado. Algunos estudios realizados en la Universidad Austral de Chile proponen material para la medición de inteligibilidad de la voz en español; sin embargo este material está basado principalmente en logátomos, y además por ser diseñado en otro país muchos de los términos son desconocidos en nuestro lenguaje [Rosas y Sommerhoff, 2008].
- El presente trabajo aporta propuestas metodológicas y resultados experimentales para investigaciones futuras de la influencia del ángulo de incidencia de la señal de voz en la inteligibilidad biaural; un aspecto que ha sido muy poco estudiado. Ya que por otra parte, en la literatura se encuentra mucha información sobre el estudio de la localización biaural de fuentes sonoras [Wenzel, 1993; Møller, 1996; Xu *et al.*, 2007; Hu *et al.*, 2008].
- Resultados originales adicionales de este trabajo son los que se obtuvieron al comparar la correlación interaural del disturbio acústico (ya sea ruido o reverberación), en la inteligibilidad biaural de la voz. Se observa que la correlación interaural tiene una influencia importante en la inteligibilidad para los dos diferentes ambientes acústicos analizados (como se muestra en las Secciones 4.2.6). Los estudios revisados [Licklider, 1948; Pollack y Trittipoe, 1959] presentan información acerca de la correlación interaural; sin embargo, a diferencia del presente trabajo, estos no realizan comparaciones entre las señales de voz y el disturbio acústico, en los casos en los que están fuertemente correlacionados ($IACC > 0.8$) y no correlacionados ($IACC < 0.2$).

- Otro aporte original de este trabajo, es el estudio de la influencia que tiene el ángulo de incidencia de la señal de voz en la inteligibilidad biaural, cuando la señal de voz se presenta bajo condiciones acústicas adversas. Existe un trabajo previo [Nordlund y Fritzell, 1963] que detalla la forma en que un mensaje de voz se modifica en relación con el ángulo de incidencia; sin embargo dicho trabajo no investiga qué ocurre cuando la señal de voz está inmersa en un ambiente reverberante.

Para el trabajo futuro en esta línea de investigación, se recomienda lo siguiente:

- Implementar modelos computacionales, como los propuestos en la literatura [Beutelmann y Brand, 2006; van Wijngaarden y Drullman, 2008; Lavandier y Culling, 2010; Beutelmann *et al.*, 2010; Jelfs *et al.*, 2011] para medir de forma objetiva la inteligibilidad biaural de la voz y comparar con los resultados obtenidos en este trabajo a través de pruebas subjetivas.
- Considerar ambientes cotidianos específicos, tanto de ruido como de reverberación; por ejemplo: ruido de cafetería, ruido urbano, entre otros; mientras que en el caso de la reverberación, tiempos de reverberación típicos de lugares públicos como cafeterías, estaciones de transporte, salones de clases, etc.
- Hacer un estudio más amplio de la inteligibilidad a diferentes ángulos, con una subdivisión más fina del ángulo acimutal y considerando también ángulos de elevación.
- Evaluar los modelos objetivos de inteligibilidad biaural publicados en la literatura [Beutelmann y Brand, 2006; van Wijngaarden y Drullman, 2008; Lavandier y Culling, 2010; Beutelmann *et al.*, 2010; Jelfs *et al.*, 2011] para validarlos y establecer sus límites de aplicación, así como para establecer en qué casos son todavía necesarias las pruebas subjetivas.
- Evaluar el umbral de recepción de la voz (SRT, Speech Reception Threshold) bajo condiciones adversas de audición, además de evaluar el porcentaje de palabras correctamente entendidas. Esto, debido al uso relativamente frecuente del índice SRT en la literatura especializada sobre el tema.

REFERENCIAS

AES Staff, (2006). "Binaural Technology for Mobile Applications", *J. Audio Eng. Soc.*, 54 (10), pp. 990 – 995.

Algazi, V. R., Duda, R., Thompson, D., Avendano, C. (2001). "The CIPIC HRTF Database", *In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pp. 99 – 102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24.

Disponible en: <http://interface.cipic.ucdavis.edu/sound/hrtf.html>

Arsenault, M. D., Punch, J. L. (1999). "Nonsense-syllable recognition in noise using monaural and binaural listening strategies", *J. Acoust. Soc. Am.*, 105 (3), pp. 1821 - 1830.

Bai Mingsian, Shih Geng-Yu, Lee Chih-Chung (2007), "Comparative study of audio spatializers for dual-loudspeakers mobile phones", *J. Acous. Soc. Am.*, 121 (1), pp 298-308.

Barry A. (1992), "A review of the cocktail party effect", *J. of the American Voice I/O Society.*, 12, pp. 35 – 50.

Begault D., Wenzel E. (1992), "Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces", *J. Aviation Psychology*, pp. 1-22.

Begault D., Wenzel E. (1993), "Headphone localization of speech", *Hum. Factors*, 35, pp. 361-376.

Benitez, L., Speaks, C., (1968), "A test of speech intelligibility in the Spanish language", *Int. Audiol.*, 7 (1), pp. 16-22.

Berkow R., Talbott J. (1978), *El manual Merck*. Sexta edición. Ed. Merck Sharp y Dohme International.

Berruecos P., Rodriguez J. L. (1967), "Determination of the phonetic percent in the Spanish language spoken in México City, and formation of P.B. lists of trochaic words", *Int. Audiol.* 6(2), pp. 211-216.

Beutelmann, R., Brand, T. (2006). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners", *J. Acoust. Soc. Am.*, 120 (1), pp. 331 – 342.

Beutelmann, R., Brand, T., Kollmeier, B. (2010). "Revision, extension, and evaluation of a binaural speech intelligibility model", *J. Acoust. Soc. Am.*, 127 (4), pp. 2479 – 2497.

Blauert, J. (1971). "Localization and the Law of the first wavefront in the median plane", *J. Acoust. Soc. Am.*, 50 (2), pp. 466 – 470.

Blauert, J. (1997). "Spatial hearing. The psychophysics of human sound localization", The MIT Press. Cambridge, MA.

Blauert, J. (1998). "The AUDIS catalog of human hrtfs". In *Proc. 16th ICA, Seattle, WA*, pp. 2901–2902.

Bronkhorst, A. W. (2000), "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions", *Acustica.*, 86, pp. 117 – 128.

Bronkhorst, A. W., Plomp R. (1988), "The effect of head-induced interaural time and level differences on speech intelligibility in noise", *J. Acoust. Soc. Am.*, 83 (4), pp. 1508 – 1516.

Bronkhorst, A. W., Plomp R. (1989), "Binaural speech intelligibility in noise for hearing-impaired listeners", *J. Acoust. Soc. Am.*, 86 (4), pp. 1374 – 1383.

Brungart, D., Rabinowitz, W. (1999), "Auditory localization of nearby sources. Head-related transfer functions", *J. Acous. Soc. Am.*, 106 (3), pp. 1465-1479.

Burkhard, M.D., Sachs, R.M. (1975), "Anthropometric manikin for acoustic research", *J. Acoust. Soc. Am.*, 58 (1), pp. 214 - 222.

Cancel, C. A., (1965), "Multiple-choice intelligibility lists for Spanish speech audiometry", *Int. Audiol.*, 4 (2), pp. 91-93.

Castañeda, R., Pérez, S. (1991), “Análisis fonético de las listas de palabras de uso más extendido en logaudiometría”, *Anales de la Sociedad Mexicana de Otorrinolaringología*. Vol. XXXVI, pp. 23 – 30.

Cheng, C., Wakefield, G. (2001), “Introduction to head-related Transfer Functions (HRTF’s): Representations of HRTF’s in Time, Frequency, and Space”, *J. Audio Eng. Soc.*, 49 (4), pp. 231 – 249.

Cherry, Colin E. (1953), ”Some experiments on the recognition of speech, with one and two ears”, *J. Acoust. Soc. Am.*, 25 (5), pp. 975 – 979.

Colburn, H. Steven, C., Suzzane, Wan, Rui (2007), “Spatial effects on speech intelligibility”, Paper presented at the 19th International Congress on Acoustics. Madrid, España.

Committee on Hearing, Bioacoustics and Biomechanics (CHABA), (1988). “Speech understanding and aging”. *J. Acoust. Soc. Am.*, 83, pp. 859 – 895.

Dirks, D. D., Wilson, R. H. (1969a), ”The effect of spatially separated sound sources on speech intelligibility”, *J. Speech Hear. Res.*, 12, pp. 5 – 38.

Dirks, D. D., Wilson, R. H. (1969b), ”Binaural hearing of speech for aided and unaided conditions”, *J. Speech Hear. Res.*, 12, pp. 650 – 664.

Duda, R., (1993). “Modeling Head Related Transfer Functions”, in *Proc. 27th Ann. Asilomar Conf. Signals, Systems and Computers*, pp. 996 – 1000, Asilomar, CA, Nov. 1–3.

Duquesnoy, A. J. (1983), ”Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons”, *J. Acoust. Soc. Am.*, 74 (3), pp. 739 – 543.

Duquesnoy, A. J., Plomp R. (1980), ”Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis”, *J. Acoust. Soc. Am.*, 68 (2), pp. 537 – 544.

Durlach, N.I., (1963), “Equalization and cancellation theory of binaural masking-level differences”, *J. Acoust. Soc. Am.*, 35 (8), pp. 1206-1218.

Drullman, R., Bronkhorst, A. W. (2000), "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation", *J. Acoust. Soc. Am.*, 107 (4), pp. 2224 – 2235.

Egan, J.P. (1948), "Articulation testing methods", *Laryngoscope*. 58 (9), pp. 955-931.

Ferrer, O. (1960), "Speech audiometry: a discrimination test for Spanish language", *Laryngoscope*. 70 (11), pp. 1541-1551.

Flanagan, J. L., (1960), "Analog measurements of sound radiation from the mouth", *J. Acoust. Soc. Am.*, 32, pp. 1613 – 1620.

Fletcher, N.H. (2007), *Springer Handbook of Acoustics*. 1st Ed. Springer, pp. 1182.

French, N. R., Steinberg, J. C. (1947), "Factors governing the intelligibility of speech sounds", *J. Acoust. Soc. Am.*, 19, pp. 90 – 119.

Gardner, B., Martin, K., (1994), "HRTF measurements of a KEMAR dummy-head microphone," *MIT Lab Perceptual Computing Tech. Rep. #280*.

Disponible en: <http://sound.media.mit.edu/pub/Data/KEMAR>

Gelfand, S., Hochberg, I. (1976), "Binaural and monaural speech discrimination under reverberation", *Audiology*., 15 (1), pp. 72 – 84.

Goldstein, B. (2005). "Sensación y percepción", Thomson, Sexta Edición.

Grassi E., Tulsi J., Shamma, S. (2003). "Measurement of Head-Related Transfer Functions based on the empirical transfer function estimate", *In Proc. International Conference on Auditory Display*. Pp. 119-122. Boston, MA, USA, July 6-9.

Disponible en: <http://www.isr.umd.edu/Labs/NSL/Software.htm>

Hamidi, F., Kapralos, B. (2009), "A Review of Spatial Sound for Virtual Environments and Games with Graphics Processing Units", *The Open Virtual Reality Journal 1*, pp. 8-17

HammersøI, D., Møller, H. (1996), “Sound transmission to and within the human ear canal”, *J. Acoust. Soc. Am.*, 100 (1), pp 408-427.

Harikrishnan, N., Sengupta, D., Velmurugan, R., Rao, P. (2011), “Azimuth-dependent Spatialization for a Teleconference Audio Display”, *Proc. Of the National Conference on Communications (NCC)*, January, Bangalore, India.

Hartmann, W. M. (1999), “How we localize sound”, *Physics Today*, pp. 24-29.

Hartmann, W. M., Constan, Z. (2002), “Interaural level differences and the level-meter model”, *J. Acous. Soc. Am.*, 112 (3), pp 1037-11045.

Hawley, M., Litovsky, R., Colburn, J. (1999). “Speech intelligibility and localization in a multiple-source environment”, *J. Acoust. Soc. Am.*, 105, pp. 3436-3448.

Hawley, M., Litovsky, R., Culling, J. (2004). “The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer”, *J. Acoust. Soc. Am.*, 115 (2), pp 833-843.

Hirsh, I. (1950). “The relation between localization and intelligibility”, *J. Acoust. Soc. Am.*, 22 (2), pp 196-200.

Houtgast, T., Steeneken, H.J.M. (1973). “The Modulation Transfer Function in room acoustics as a predictor of speech intelligibility”. *Acustica* 28, pp. 63 – 73.

Houtgast, T., Steeneken, H.J.M. (1984). “A multi-language evaluation of the RASTI-Method for estimating speech intelligibility in auditoria”, *Acustica* 54, pp 185 - 199.

Houtgast, T., Steeneken, H.J.M. (1985a). “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria”. *J. Acoust. Soc. of Am.*, 77, pp. 1069 – 1077.

Houtgast, T., Steeneken, H.J.M.(1985b). Technical Review No. 3, The Modulation Transfer Function in room acoustics. Marlborough, MA: Brüel & Kjaer Instruments, pp. 1-44.

Hu H., Zhou L., Ma H., Wu Z., (2008). “HRTF personalization based on artificial neural network in individual virtual auditory space”, *Applied Acoustics*. 69, pp. 163-172.

Hughes, D. (1989). “Glass cockpit study reveals human factors problems” *Aviation Week and Space Technology*. 131 (6), pp. 32-36

IRCAM Database (2003)

Disponible en: <http://recherche.ircam.fr/equipes/salles/listen/context.html>

Jelfs, S., Culling, J. F., Lavandier, M. (2011). “Revision and validation of a binaural model for speech intelligibility in noise”, *Hearing Research*. 275, pp. 96 – 104.

Kanada, Y. (2004). “Multi-Context Voice Communication Controlled By Using An Auditory Virtual Space”, *2nd IASTED International Conference on Communication and Computer Networks (CCN 2004)*, pp. 1 – 6. Nov. 8 – 10.

Kanada, Y. (2005). “Simulated virtual market place using voiscap communication medium”, *In Proc. 13th ACM Int. Conf. Multimedia*, pp. 794 – 795.

Kazuhiro, I., Motokuni, I., Atsue, I., Masayuki, M. (2007). “Median plane localization using a parametric model of the head-related transfer function based on spectral cues”, *Applied Acoustics*. 68, pp 835-850.

Kim, S., Frisina, R., Mapes, F., Hickman, E. (2006), “Effect of age on binaural speech intelligibility in normal hearing adults”, *Speech Communication*, 48, pp. 591 – 597.

Kistler, D.J., Wightman, F.L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91(3), pp.1637-1647.

Kock, W. E. (1950). “Binaural localization and masking”, *J. Acoust. Soc. Am.*, 22 (6), pp. 801 – 804.

Koenig, W. (1950). “Subjective effects in binaural hearing”, *J. Acoust. Soc. Am.*, 22 (1), pp. 61– 62.

Kuttruff, Heinrich. (2009). Room Acoustics, fifth ed., Spon Press.

Kryter, K. D. (1962). "Methods for the calculation and use of the articulation index", *J. Acoust. Soc. Am.*, 34, pp. 1689 – 1697.

Lavandier, M., Culling, J. F. (2008). "Speech segregation in rooms: Monaural, binaural and interacting effects of reverberation on target and interferer", *J. Acoust. Soc. Am.*, 123 (4), pp. 2237-2248.

Libbey, B., Rogers, P. (2004). "The effect of overlap-masking on binaural reverberant word intelligibility", *J. Acoust. Soc. Am.*, 116 (5), pp. 3141 - 3151.

Licklider, J. (1948), "The influence of interaural phase relations upon masking of speech by white noise," *J. Acoust. Soc. Am.*, 20, pp. 150–159.

Loomis, J., Golledge, R., Klatzky, R., Speigle, J., Tietz, J., (1994), "Personal guidance system for the visually impaired", *Proceedings of the First Annual International ACM/SIGCAPH Conference on Assistive Technologies*, pp. 85-90, Marina Del Rey. California, October 31-November 1.

Loomis, J., Golledge, R., Klatzky, R. (1998). "Navigation system for the blind: Auditory display modes and guidance", *Presence: Teleoperators and Virtual Environments* 7, pp. 193-203.

Loomis, J., Hebert, C., Cicinelli, J. (1990), "Active localization of virtual sounds," *J. Acoust. Soc. Am.*, 88 (4), pp. 1757–1764.

MacKeith, N. W., Coles, R. R. A. (1971). "Binaural advantages in hearing of speech". *J. Laryngol. Otol.*, 85, pp. 213 – 232.

Marshall, A. H., Barron, M. (1981). "Spatial impression due to early lateral reflections in concert halls", *J. Sound. Vib.*, 77, pp. 211-232.

Mazelova, J., Popelar, J., Syka, J., (2003). "Auditory function in presbycusis: peripheral vs. central changes", *Exp. Gerontol.*, 38, pp. 87 – 94.

Miller, G. A., Heise, G. A., Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials", *J. Exp. Psychol.*, 41, pp. 329 – 335.

Møller, H. (1992). "Fundamentals of Binaural Technology", *Applied Acoustics.*, 36, pp. 171-218.

Møller, H., Sørensen, M., Hammersøi, D., Jensen, C. (1995). "Head-Related Transfer Functions of Human Subjects", *J. Audio Eng. Soc.*, 43 (5), pp. 300-321.

Moncur, J. P., Dirks, D. D. (1967). "Binaural and monaural speech intelligibility in reverberation", *J. Speech Hear. Res.*, 10, pp. 186 -195.

Moore, B. (2004). "An introduction to the psychology of hearing". Quinta edición. United Kingdom. Elsevier Academic Press, pp. 248 – 253.

Morfey, C. (2001). "Dictionary of Acoustics". United Kingdom. Academic Press, pp. 291.

Nábelek, A. K., Mason, D. (1981). "Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms", *J. Speech Hear. Res.*, 24, pp. 375 -383.

Nábelek, A. K., Pickett, J.M. (1974). "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids", *J. Acoust. Soc. of Am.*, 56 (2), pp.628 – 639.

Nábelek, A. K., Robinson, P. K. (1982). "Monaural and binaural speech perception in reverberation for listeners of various ages". *J. Acoust. Soc. Am.*, 71(5), pp. 1242 – 1248.

Nakado, T., Nishino, T., Takeda, K. (2008), "Head-Related Transfer Function measurement in sagittal and frontal coordinates", *Acoust. Sci. & Tech.*, 29 (5), pp. 335-337.

Disponible en: <http://www.sp.m.is.nagoya-u.ac.jp/HRTF/database.html>

Neovius, L., Raghavendra, P. (1993). Evaluation of comprehension of KTH Text-to-Speech with "Listening Speed" Program. *Proceedings of Eurospeech 93* (3): 1687-1690.

Nishino, T., Inoue, N., Takeda, K., Itakura, F. (2007), "Estimation of HRTFs on the horizontal plane using physical features", *Applied Acoustics.*, 68, pp. 897-908.

Nordlund, B., (1962), "Physical factors in angular localization", *Acta Otolaryng.* 54, pp. 75-93.

Nordlund, B., Fritzell, B., (1963), "The influence of azimuth on speech signals", *Acta Otolaryng.* 56, pp.633 -642.

Nordlund, B., Lidén, G., (1963), "An artificial head", *Acta Otolaryng.* 56, pp. 493-499.

Olsen, W., Carhart, R., (1967). "Development of test procedures for evaluation of binaural hearing aids", *Bull. Prosthetics Res.* 10, pp. 22 – 49.

Pierce, Allan D., (1981). *Acoustics: An Introduction to Its Physical Principles and Applications*, McGraw-Hill Book Company, New York.

Pollack, I., Pickett, J. M. (1958). "Stereophonic listening and speech intelligibility against voice babbles," *J. Acoust. Soc. Am.* 30, pp. 131–133.

Ramos, O., Calvo, G., Tommasini, F. (2007). "Modelo Acústico de Cabeza y Torso Mediante Análisis de Componentes Principales". *In Mecánica Computacional*. Vol. XXVI, pp. 46 – 58.

Rosas, C., Sommerhoff J., (2008). "Inteligibilidad acústica en español: una propuesta para su medición". *Estudios Filológicos* 43, pp. 179 – 190.

Salinas, G. (2007). "Validación mediante índices acústicos de material de voz para medir la inteligibilidad". Tesis de Licenciatura. Universidad Nacional Autónoma de México. México, D.F.

Schroeder, M. R. (1981). "Modulation transfer functions: Definition and measurement". *Acustica* 49, pp. 179 – 182.

Steeneken, H.J.M., Houtgast T. (1980). "A physical method for measuring speech-transmission quality". *J. Acoust. Soc. Am.*, 67, pp. 318 – 326.

Spiegel, M., Schiller, J., Srinivasan, A. (2009). *Probability and statistics*, Serie Schaum. 3a. edición. Mc. Graw Hill. N.Y.

Spieth, W., Curtis J. F., Webster J. C. (1954). “Responding to one of two simultaneous messages”. *J. Acoust. Soc. Am.*, 26(1), pp. 391–396.

Strutt, J. W. (Lord Rayleigh). (1907). “On our perception of sound direction”. *Philos. Mag.*, 13, pp. 214 – 232.

Takane, S., Arai, D., Miyajima, T., Watanabe, K., Suzuki, Y., Sone, T. (2002), “A database of Head-Related Transfer Functions in whole directions on upper hemisphere”, *Acoust. Sci. & Tech.*, 23 (3), pp. 160-162.

Disponible en: <http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf/>

H. Sutou (2005), Shimada Laboratory (Nagaoka University of Technology, Japan), HRTF Database.

Disponible en: <http://audio.nagaokaut.ac.jp/hrtf/>

Suzuki Y., Murase A., Itoh M, Sakamoto S., (2008) “Determination of filtering parameters for dichotic-listening binaural hearing aids”, Paper presented at the 2nd annual meeting of the European Acoustics Association.

Takeuchi, T., Nelson, P. A., Hamada, H. (2001), “Robustness to head misalignment of virtual sound imaging systems”, *J. Acous. Soc. Am.*, 109 (3), pp. 958-971.

Tato, J. (1949), “Lecciones de Audiometria”, Ed. El Ateneo, Buenos Aires.

van Wijngaarden, S., Drullman, R. (2009), “Binaural intelligibility prediction based in the speech transmission index”, *J. Acous. Soc. Am.*, 123 (6), pp. 4514-4523.

Wan, R., Durlach, N. I., Colburn, H. S., (2010), “Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers” , *J. Acous. Soc. Am.*, 128 (6), pp. 3678-3690.

Wenzel, E., Arruda M., Kistler, D., Wightman F., (1993), “Localization using nonindividualized head-related transfer functions”, *J. Acous. Soc. Am.*, 94 (1), pp. 111-123.

Webster, J. C., Thompson, P. O. (1954), “Responding both of two overlapping messages”, *J. Acous. Soc. Am.*, 26(1), pp. 396–402.

Xie, B.S., (2002), “Effect of Head Size on Virtual Sound Image Localization”. *Applied Acoustics* 21(5), pp. 1–7.

Xu S., Li Z., Salvendy G., (2007), “Individualization of Head-Related Transfer Function for Three-Dimensional Virtual Auditory Display: A Review”, ICVR’07 Proceedings of the 2nd. International conference on virtual reality, pp. 397 – 407.

Zubick, H., Irizarry, L. M., Rosen, L., Feudo, P. Kelly, J. H., Strome, M., (1983), “Development of speech audiometry materials for native Spanish-speaking adults”, *Audiology* 22, pp. 88 – 102.

Zwicker, E., Fastl, H. (1999), *Pshychoacoustics, Facts and models*, segunda edición, Springer

APÉNDICE A

En este apéndice se muestran las palabras empleadas en las pruebas de inteligibilidad desarrolladas en este trabajo. Como se mencionó anteriormente, se muestra que son 4 listas cada una con 50 palabras. Se observa que todas son palabras graves conformadas por dos sílabas [Castañeda y Pérez, 1991; Salinas, 2007].

Lista 1

nube	clavo	grasa	nina	ellos	suela	selva	gestor	simple	dime
dejo	surco	neta	botes	sigo	duque	caro	listo	persa	lenta
críos	pleno	medios	perros	piano	mimo	cierta	pera	toro	celda
cuales	flaca	miden	calor	choca	cita	crean	cifra	deme	tiendas
beca	torno	reno	duna	llenos	diosa	une	prima	veinte	nada

Lista 2

cedros	pluma	nombre	manto	perla	jalan	miope	laca	talco	dique
gose	fierro	cumbres	cebra	cena	nuca	puse	niña	seda	lista
meta	lina	disco	timbre	celo	noche	Viena	himno	conde	seso
sella	pista	ciega	dije	tira	codo	cero	cera	tiro	sigo
duda	duelo	trenza	norte	lince	nená	pacto	alla	saco	cura

Lista 3

Corea	pura	canción	busto	hacia	damas	ciega	tina	fino	padre
dulce	vino	reto	turco	cama	pajes	libre	vienen	cielo	pardo
siglo	níquel	tecleo	viernes	lloro	cabe	deme	regla	necio	onda
yeso	monte	tarde	quepa	luces	Carmen	mismo	suela	dota	nadie
taches	Nilo	flanes	celtas	premios	corta	surco	dardo	trance	pica

Lista 4

radio	algún	unos	Diego	pili	brazo	senda	control	dante	buque
mide	cano	reto	seda	calle	sebo	clame	tape	dones	queso
sello	dieta	freno	jaque	lila	lacre	liso	Carmen	seta	esos
mini	tigre	cerca	pisen	cardo	plato	curas	lista	nave	siete
Irma	hombro	sudo	prensa	hambre	tapia	sones	feria	nulo	asno

APÉNDICE B

En este apéndice se muestran las palabras que no se consideraron en el análisis de los resultados, se muestran en diferentes tablas según las pruebas realizadas.

Tabla B-1: Lista de palabras que no fueron consideradas en el análisis de los resultados en las pruebas realizadas para distintas presentaciones y bajo condiciones no correlacionadas de ruido y reverberación (Prueba 1).

Presentación	Disturbio	Lista	Palabras no consideradas para el análisis
Monoaural	Ruido	4	mini, cano, jaque, pisen, pili, clame.
	Reverberación	2	cedros, goce, sella, fierro, lina, ciega, manto, cebra, perla, celo, jalan, nuca, codo, puse, pacto, laca, himno, talco, seso
Monofónica	Ruido	3	fino
	Reverberación	4	mini, cano, pili, cardo, lacre, tapia, clame, tape, Carmen
Biaural a 0°	Ruido	1	críos, flaca, nina, mimo, une, deme
	Reverberación	3	pura, reto, hacia, cama, pajes, cabe, corta, deme, padre, pardo.
Biaural a 30°	Ruido	2	pusé
	Reverberación	1	flaca, nina, mimo, une

Tabla B-2: Lista de palabras que no fueron consideradas en el análisis de los resultados en las pruebas realizadas para distintas presentaciones y bajo condiciones correlacionadas de ruido y reverberación (Prueba 2).

Presentación	Disturbio	Lista	Palabras no consideradas para el análisis
Monofónica	Ruido	3	Corea, pura, flanes, pajes, cabe, Carmen, corta, fino
	Reverberación	4	radio, mide, sello, mini, algún, cano, freno, sudo, pisen, prensa, pili, cardo, sebo, lacre, tapia, senda, clame, curas, tape, Carmen, feria, dante, seta, nave, esos, asno.
Biaural a 0°	Reverberación	3	corea, taches, pura, flanes, celtas, cama, pajes, cabe, Carmen, corta, vienen, dardo, pardo
Biaural a 30°	Ruido	2	fierro
	Reverberación	1	flaca, miden, reno, suela, mimo, caro, crean, simple, toro, veinte.

Tabla B-3: Lista de palabras que no fueron consideradas en el análisis de los resultados en las pruebas realizadas a distintos ángulos (Prueba 3).

Ángulo (°)	Disturbio	Lista	Palabras no consideradas para el análisis
+15	Ruido	4	mide, curas
+45		2	codo
+60		3	pura
+15	Reverberación	1	beca, flaca, selva, cierta, simple, toro.
+30		3	pura, tecleo, tarde, flanes, celtas, hacia, cama, pajes, cabe, corta, vienen, dardo, dota, padre, pardo
+45		4	cano, pili, cardo, sones, seta
+60		2	fierro, manto, cebra, celo, jalan, cero, laca, cera, talco, seso
-15	Reverberación	1	flaca, une, simple, toro
-30		3	Corea, taches, pura, reto, tecleo quepa, celtas, hacia, cama, damas, pajes, cabe, Carmen, corta, deme, vienen, dardo, padre, pardo, onda
-45		2	jalan, cero, seso
-60		2	fierro, manto, celo, codo, cero, seso

Tabla B-4: Lista de palabras que no fueron consideradas en el análisis de los resultados en las pruebas con HRTF individualizadas (Prueba 4).

Ángulo (°)	Disturbio	Lista	Palabras no consideradas para el análisis
+15	Ruido	4	pili, curas
+30		1	flaca
+45		2	puse, laca
+60		3	pura, flanes, cabe
-15	Ruido	4	mide, cano
-30		1	críos, flaca
-45		4	cano
-60		3	Pura, pajes
+15	Reverberación	1	flaca, miden, caro, une, toro
+30		3	corea, pura, níquel, flanes, celtas, hacia, cama, pajes, cabe, corta, vienen, padre, pardo, onda
+45		4	mide, sello, algún, cano, freno, pili, lacre, clame, tape, seta
+60		2	goce, fierro, lina, norte, perla, jalan, codo, cero, laca, saco, seso
-15	Reverberación	1	flaca, veinte
-30		3	Corea, taches, pura, níquel, tecleo, tarde, celtas, hacia, cama, pajes, cabe, Carmen, corta, deme, vienen, regla, trance, pardo, onda.
-45		2	manto, jalan, cero, laca, cera
-60		2	sella, fierro, lina, manto, cebra, norte, perla, celo, jalan, cero, laca, seso

Tabla B-4: Lista de palabras que no fueron consideradas en el análisis de los resultados en las pruebas con HRTF no individualizadas (Prueba 4).

Ángulo (°)	Disturbio	Lista	Palabras no consideradas para el análisis
+15	Ruido	4	mide, cano, pili, Carmen
+45		2	laca
+60		3	pura, corta, fino
-15	Ruido	4	mide, cano
-30		1	une
-45		4	mide, cano
-60		3	pura, cabe
+15	Reverberación	1	dejo, críos, flaca, miden, reno, mimo, selva, caro, crean, une, simple, toro, veinte
+30		3	pura, reto, tecleo, celtas, hacia, pajes, cabe, Carmen, corta, vienen, regla, padre, pardo, onda
+45		4	mide, cano, pili, cardo, lacre, tapia, tape, Carmen, seta
+60		2	fierro, lina, cebra, perla, celo, jalan, cero, laca, cera, talco, saco, seso
-15	Reverberación	1	dejo, crios, pleno, flaca, torno, medios, miden, reno, nina, mimo, selva, caro, crean, une, simple, toro, deme, veinte, lenta
-30		3	pura, monte, celtas, hacia, cama, damas, pajes, cabe, Carmen, corta, vienen, regla, dota, padre, pardo
-45		2	lina, manto, perla, celo, jalan, cero, laca, dique, seso
-60		2	sella, fierro, norte, perla, celo, jalan, cero, laca, cera, saco, seso