

COMPARACIÓN ENTRE LAS TÉCNICAS BINAURAL Y ESTEREO PARA LA LOCALIZACIÓN DE FUENTES EN EL ESPACIO 3D

PACS: 43.66.Pn

Robert Barti. Fabiola Ordoyo
Departamento de Acústica. Ingeniería y Arquitectura La Salle
Pg. Bonanova
8 08022 Barcelona
Tel: 932 902 427
Fax: 932 902 416
E-mail: robert@salleURL.edu

ABSTRACT

In the most advanced acoustic applications, sounds are recorded in order to have a good sensation of realism when they are played through headphones or loudspeakers. Two technologies are available at the moment: binaural technique seems better than stereo technique for a first time. In this paper the target is to obtain the best technique of recording signal showing natural sound. Different sound sources were recorded with different microphone techniques, in a well acoustic conditioned room. Distance to microphone and horizontal angle is modified many times. All the sound sources were placed over the horizontal ear's plane. These signals were played through headphones, and listeners must indicate the origin of the sound (distance and angle). With these results interesting conclusions were obtained.

RESUMEN

En las aplicaciones más avanzadas de acústica, se graban las señales de sonido, para obtener el máximo realismo a la hora de ser reproducidas mediante auriculares o altavoces. La tecnología binaural vs la estereo parece en principio más ventajosa. Con éste trabajo, se trata de establecer que técnica de grabación resulta más natural. Para ello se utiliza una sala acondicionada acústicamente para grabar, distintos sonidos procedentes de distintas fuentes, utilizando diferentes técnicas microfónicas. Se varía tanto la distancia a los micrófonos, como su ángulo de posición. Todos los sonidos se sitúan en el plano horizontal del oído. Estas señales se reproducen mediante auriculares, y los sujetos deben indicar la procedencia del sonido y su distancia. A partir de los resultados se sacan una serie de conclusiones interesantes.

OBJETIVOS

Se pretende obtener un sistema de grabación de dos canales que permita reproducir con extrema fidelidad el sonido. Se parte de un sistema digital de grabación con tecnología de 24 bits que asegura un margen dinámico y una respuesta en frecuencia más que suficientes. La aplicación principal radica en la grabación de sonido en el interior del vehículo automóvil, para ser posteriormente escuchada y valorada. La sensación de realismo es pues nuestra máxima aspiración.

TECNOLOGÍA ESTEREO

Esta tecnología tiene una larga tradición y experiencia y se puede encontrar muy buena documentación. Se basa en la utilización de dos micrófonos con diferentes posiciones de montaje y de respuesta direccional. Los parámetros que hay que tener en cuenta son: la fusión, la localización, y la profundidad. Estos parámetros nos dan información sobre diferencia de nivel y de tiempo entre las dos señales microfónicas recibidas. Hay tres tipos de configuración básica: el par espaciado, el par coincidente y el par casi-coincidente.

Micrófono par espaciado: Se basa en diferencias temporales. Proporciona una buena imagen estéreo. Poca compatibilidad en sistemas monofónicos. Precisa de micrófonos omnidireccionales. Necesita separaciones de más de 90 cm.

Micrófono par coincidente: Basada en diferencias de nivel. Ofrece una menor imagen estéreo. Compatible con tomas monofónicas. Precisa de micrófonos directivos. Técnicas X_Y, Blumlein.

Micrófono par casi-coincidente: Combina las diferencias de tiempo y de nivel. Compromiso entre imagen estéreo y compatibilidad con monofonía. Precisa de micrófonos directivos. Técnicas ORTF, OOS, etc.

TÉCNICA BINAURAL

Esta tecnología se basa en el uso de un maniquí que permite crear el mismo campo acústico interferente, que el cuerpo humano. Uno de los principales inconvenientes, radica en que la función HRTF, es cambiante con cada individuo, de manera que las sensaciones que se pueden percibir, con una grabación que introduce una HRTF distinta a la nuestra, pueden ser confusas. La imagen sonora, la realiza nuestro cerebro a partir de las informaciones binaurales percibidas. Durante la fase de aprendizaje (infancia) sabemos localizar fuentes sonoras, así como analizar sus características, familiaridad de un sonido, mensaje acústico, etc. Es en ésta fase del aprendizaje, donde nuestra fisiología del oído externo, y nuestro cuerpo, nos determinan una función HRTF propia. Las diferencias de niveles y de tiempos entre los mensajes recibidos por nuestros oídos (los ITD y los ILD) nos permiten identificar la procedencia del sonido en el espacio 3D. La técnica binaural se basa en el uso de un maniquí más o menos realista en formas, que incorpora dos micrófonos situados en los puntos donde estaría la entrada del pabellón auditivo. También han aparecido recientemente sistemas basados en cascos auriculares que incorporan los micrófonos. Estos sistemas también se han analizado.

PRIMERAS PRUEBAS CON VEHÍCULOS. RUIDO INTERIOR

Mediante éstas técnicas se han grabado señales en el interior de vehículos. Para ello se ha montado un sistema muy ligero pero a la vez robusto para fijar los micrófonos. En la fotografía siguiente se puede ver un ejemplo de montaje. Aunque en la imagen no se aprecia, la varilla inclinada permite situar a los dos micrófonos a la altura de los oídos de una persona. Este sistema debe montarse sin la influencia de la persona. Por ello sólo puede colocarse en el asiento del acompañante del vehículo. Tanto la distancia de separación como el ángulo de inclinación de los micrófonos, puede ajustarse libremente.



Este sistema debe montarse sin la influencia de la persona. Por ello sólo puede colocarse en el asiento del acompañante del vehículo. Tanto la distancia de separación como el ángulo de inclinación de los micrófonos, puede ajustarse libremente.

Cuando se escucha el ruido interior de vehículos, el usuario pone su atención en las sensaciones que los diferentes ruidos le producen. La forma de grabar las señales y de reproducirlas será determinante. En éste trabajo se han

comparado diversas grabaciones realizadas con técnicas estéreo y técnicas binaurales. Los

mejores resultados se han obtenido curiosamente con técnicas estéreo "clásicas", aunque "a priori", la binaural parecía superior. Con éste modelo de microfónica, como el de la fotografía, se han conseguido los resultados más realistas. La separación está realmente lejos de la que ofrece la cabeza humana.

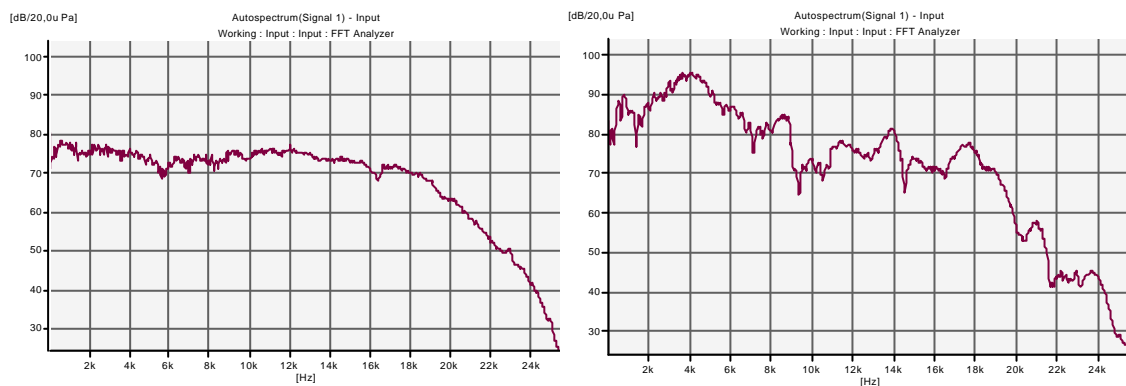
Uno de los problemas detectados en las grabaciones es la oscilación de los micrófonos con el traqueteo del coche. Estas oscilaciones perfectamente visibles, son de muy baja frecuencia. Esta señal disminuye el margen dinámico de la medición, ya que el previo que lleva el micrófono situado inmediatamente después del micro, no puede ser atenuado o ajustado. Esta baja frecuencia se traduce en la reproducción, en una modulación en amplitud, bastante perceptible en algunos casos, y en un exceso de baja frecuencia, ya que el sistema tiende a reproducir todo el margen frecuencial. Nuestro oído no tiene éste grado de sensibilidad a las bajas frecuencias, y por consiguiente ese sonido suena muy poco real.

ALTAVOCES vs AURICULARES

A la hora de reproducir el sonido, se puede escoger entre auriculares o altavoces. Los primeros permiten una localización intra-craneal, aunque tienen una mayor fidelidad, ya que disponen de una fuente única, mientras que los segundos permiten una reproducción más cercana al 3D pero con los inconvenientes de los rayos cruzados, respuesta frecuencial más pobre, y disponer de fuentes múltiples de sonido. Valorando los pros y los contras se opta por la reproducción con auriculares, ya que a las ventajas comentadas anteriormente, hay que añadir la mayor inmunidad a las condiciones de la sala, así como al ruido ambiente.

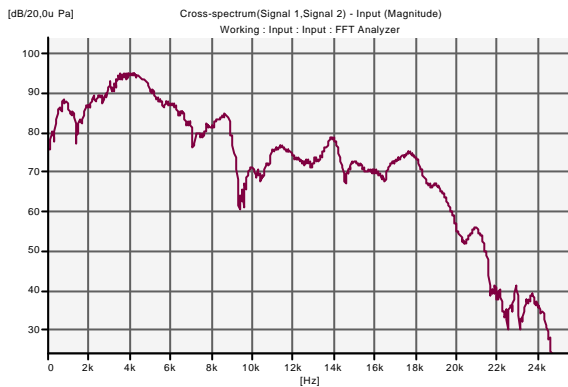
ANÁLISIS FRECUENCIAL

Es evidente que la respuesta frecuencial, aun teniendo los mismos micrófonos va a ser distinta, ya que en el caso del maniquí, la influencia de la cabeza y torso, modificará su sensibilidad frecuencial, la cual a su vez será función del ángulo de incidencia preferente del sonido. A continuación se muestra la diferencia entre una técnica estéreo y una binaural.



A la izquierda tenemos el espectro obtenido de la señal estéreo, y a la derecha el espectro de la señal binaural equivalente. En ambos casos se parte de una fuente de ruido blanco radiando de forma uniforme en cámara anecoica. Notamos no sólo lo diferente de las curvas, sino también las diferencias de los niveles. A la izquierda tal como recibimos el sonido en el oído externo, a la derecha el filtrado introducido por el conducto auditivo, y las difracciones del pabellón, torso y hombros. En ambos casos se trata de la información correspondiente al sonido recibido frontalmente.

La dificultad de las mediciones binaurales, radica en la ausencia, por el momento, de parámetros de medida binaurales. Todos los existentes actualmente se basan en mediciones monoaurales. Por tanto a la hora de evaluar o cuantificar ciertos aspectos del sonido, nos encontraremos siempre con la dificultad de establecer una escala que permita comparar los resultados. El espectro cruzado entre ambos canales nos puede aportar alguna información adicional. En la figura siguiente, se muestra el espectro cruzado para el ejemplo anterior de la



respuesta binaural. Se observan ciertas diferencias que podrían aportar información única del registro binaural. Falta desarrollar una técnica de análisis que aproveche éstas diferencias y que permita cuantificar los resultados medidos.

Como se ha comentado anteriormente, se observan diferencias de la calidad del sonido cuando se utilizan diferentes técnicas de grabación del sonido. En general se puede decir que no existe una técnica que sea sistemáticamente mejor que otra. En

algunos casos la técnica estéreo se muestra más natural para detectar algunos aspectos. En otros casos, es la técnica binaural la que ofrece mejores resultados, siempre desde el punto de vista subjetivo. Nos interesa poder establecer comparaciones respecto de una referencia, antes de abordar el problema más complejo de escuchar el ruido interior de un coche.

EXPERIMENTACIÓN SUBJETIVA EN EL LABORATORIO

Para llevar a cabo las pruebas subjetivas, se recurre a la preparación de un experimento controlado en el laboratorio. Concretamente se trata de grabar los sonidos procedentes de diversas fuentes sonoras de uso familiar, un despertador, un móvil, un temporizador, etc. Dichos elementos se ubican en diferentes puntos del plano horizontal del oído de una persona sentada. Las grabaciones se realizan en la sala de grabación que dispone nuestro departamento, que se caracteriza por tener una respuesta muy plana tanto a baja como a alta frecuencia con un TR60 entorno a los 0,5 seg. Las señales de los diferentes equipos se graban directamente en disco duro, y se usan diferentes técnicas microfónicas, y dos maniqués disponibles en ese momento. Para las pruebas se utiliza la técnica de grabación en campo próximo, ya que los objetos no estaban más lejos de 2,5m.

El objetivo principal es encontrar la técnica que ofrece un mayor realismo en la localización de las fuentes. El escenario acústico se ha mantenido exactamente igual en las pruebas subjetivas con los diferentes individuos. El experimento ha consistido en localizar, para cada sonido y cada ubicación, la procedencia del sonido escuchado, tanto el ángulo de procedencia, como la distancia de separación. Las personas que realizan los tests, se sientan en el mismo punto ocupado por los micrófonos en el proceso de grabación. En el suelo de la sala existen unas marcas separadas cada 20 cm i con 10° de resolución, siguiendo las indicaciones de J. Blauert.

Los sujetos escuchaban los sonidos a través de los auriculares. Para evitar desviaciones y falseamiento de resultados, las pruebas se hacían siguiendo un orden totalmente aleatorio, de manera que la secuencia de sonidos recibidos era distinta para cada individuo. Inicialmente los sujetos identificaban siempre las fuentes en la parte posterior. Se observó que si cerraban los ojos, se mejoraban los resultados. Con los ojos abiertos, se condiciona mucho a la identificación de la fuente, ya que ésta no existe para no dar pistas. También se tuvo cuidado de que los sujetos no conocieran nada de la prueba antes de realizarla, para no influir en sus decisiones.

Una de las cuestiones que más impacto ha tenido entre las personas que realizan el test, es que un mismo objeto varía de posición según la técnica de grabación utilizada, aunque realmente el objeto permanece siempre en la misma posición. Los sonidos se escuchaban una sola vez, aunque se podía repetir la escucha si la localización no era clara hasta un máximo de tres veces.

Es importante destacar que los sonidos se terminan de forma brusca y sin reverberación asociada de la sala en las colas. La duración de cada sonido oscilaba sobre los 2 segundos. Los sonidos utilizados son muy diversos, para evitar la influencia de la variable "sonido". Dado que se trabaja con sujetos es importante que éstos se sienten a gusto, para no condicionar una respuesta errática o equivocada. En éste sentido, las pruebas realizadas, fueron muy cortas,

evitando el cansancio. Las fuentes de ruido escogidas han sido: Motor de lavadora, una lijadora, un despertador, un teléfono móvil, un temporizador mecánico, y un radio-CD portátil sencillo, reproduciendo música pop.

En la figura siguiente se muestra el escenario de las pruebas subjetivas. Destacamos las marcas realizadas en el suelo. El sujeto escucha el sonido a través de los auriculares. La orientación es la misma que la de los micrófonos o maniquíes en el proceso de grabación.

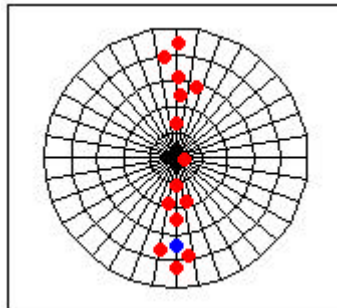


La sala reproduce una sala "normal", con superficies más o menos absorbentes, y con paredes que ofrecen reflexiones.

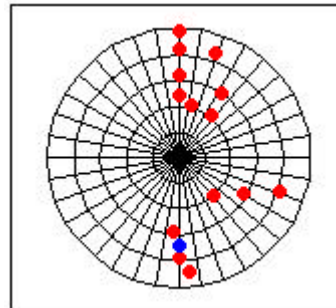
A continuación se muestran algunos de los resultados obtenidos.

RESULTADOS OBTENIDOS

Destacamos el conocido "front-back error", cuando se identifica a una fuente en la posición simétrica y opuesta a la real. En la figura siguiente se muestran dos casos, para el sonido del teléfono móvil, a la izquierda para la grabación con maniquí, y a la derecha para grabación con tecnología estéreo y 30 cm.



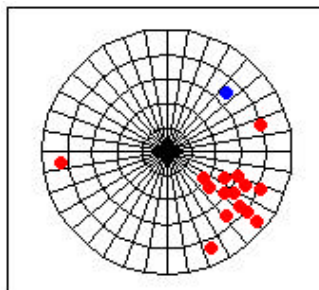
Maniquí comercial



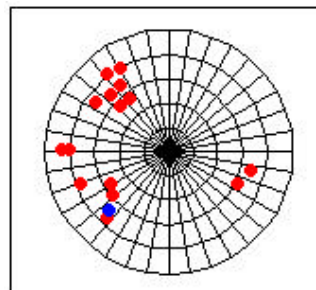
Estéreo 30 cm.

El punto azul indica la posición exacta de la fuente. Los puntos rojos son las distintas respuestas dadas por los individuos durante la prueba. Nótese que aproximadamente la mitad de las personas que realizan el test están dentro del error. En el caso de la técnica estéreo, resulta curioso observar un cierto desplazamiento hacia la derecha en buena parte de los encuestados. Esta desviación puede ser producida por la superficie acristalada que comunica la sala de grabación con el control, i que afecta a la grabación estéreo. La grabación binaural no ofrece éste defecto. El error F-B es bien conocido y no puede imputarse al tipo o modelo de maniquí utilizado.

Si tomamos muestras sobre los conos de confusión, obtenemos los resultados que se muestran en las siguientes gráficas. En éstos puntos se confunde de forma simétrica la posición de la fuente.



Estéreo a 90 cm.

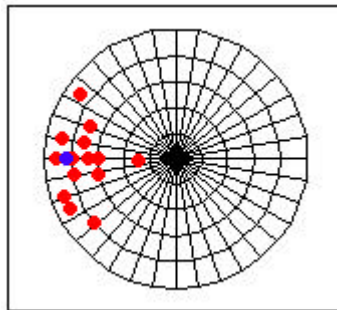


Estéreo a 30 cm.

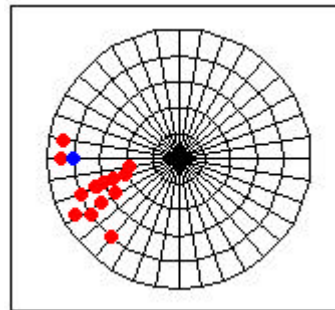
Se comparan dos técnicas estéreo, donde se muestra como en ambos casos la respuesta más votada es la simétrica a al real respecto el eje D-I. Siguiendo el mismo criterio anterior, el punto azul indica la ubicación exacta de la fuente. Los ejemplos son para las configuraciones estéreo a 30 cm y 90 cm. En ambos casos las desviaciones son similares. Se muestra un caso de cada uno de ellos. Entre ambas, la técnica estéreo a 90 cm ofrece una mayor fiabilidad, aunque mantiene el error.

Finalmente se comparan dos maniqués, uno comercial y el que disponemos en el departamento, hecho sobre la base de un maniquí de unos grandes almacenes. Este maniquí lleva dos micrófonos de precisión pre-polarizados en su interior. Los diafragmas están justo al principio del que sería el conducto auditivo. Mediante filtrado se puede añadir el efecto del conducto auditivo para visualizar la señal percibida. La superficie exterior del maniquí se ha recubierto de látex, material muy similar a la textura de la piel. Las formas y dimensiones también se ajustan a los patrones estándar. El interior está relleno de material que evita resonancias mecánicas.

En las figuras siguientes se muestran los resultados.



Maniquí comercial



Maniquí departamento

El punto azul indica la procedencia de la fuente de sonido, en éste caso el temporizador mecánico. La persona está sentada mirando hacia la parte superior de la figura. Se observa como, en promedio, el maniquí comercial ofrece una mayor precisión en la localización de la fuente, aunque los resultados ofrecen una dispersión notable alrededor del punto. Respecto a la distancia de la fuente, también se observa que el maniquí comercial consigue mayor precisión.

Nuestro maniquí aunque de promedio se aleja más de la correcta localización de la fuente, se observa un cierto "offset" para un 75% de los sujetos. Si se corrige éste offset, se obtendrá un maniquí con unas prestaciones notablemente superiores a las del maniquí comercial. Por ahora hemos caracterizado el sistema, más adelante se indicaran las mejoras añadidas.

CONCLUSIONES

1. Con auriculares se evidencia una cierta dificultad en determinar la distancia de las fuentes.
2. Para fuentes con una cierta directividad, y con radiación importante de alta frecuencia la técnica estéreo a 90 cm se muestra como la más real de todas.
3. Para fuentes más omnidireccionales, con mayor contenido de baja frecuencia, la técnica binaural es la más eficiente.
4. Aun con las ventajas de las grabaciones binaurales, el hecho de llevar asociadas una función HRTF, complica notablemente el proceso de análisis. En éste sentido, las técnicas estéreo compatibles con las grabaciones monoaurales, resultan ventajosas.
5. Es necesario el desarrollo de parámetros que puedan medir aspectos de las señales binaurales. En un futuro, éste será el patrón a seguir para realizar las medidas acústicas.
6. Los cascos que incorporan micrófonos para hacer supuestas grabaciones binaurales ofrecen los peores resultados del grupo. Es preferible una configuración simple X-Y en lugar de éstos elementos.

7. Las grabaciones en interior de vehículo carecen de realismo debido a la sensibilidad del micrófono para frecuencias muy bajas. Nuestro oído no es tan sensible. La falta de realismo se acentúa en la reproducción mediante altavoces.