EVALUACIÓN Y COMPARACIÓN DE CINCO MODELOS DE PERCEPCIÓN BINAURAL

*Jairo Alberto Hurtado Londoño**

*Pedro Raúl Vizcaya Guarín***

Resumen: tomando como referencia la base de datos de HRTF (Head-Related Transfer Function – Función de Transferencia de la Cabeza) [Begault, 2000] [Brown y Duda, 1998] realizada en el MIT [Martin y Gardner, 1994], se desarrollaron cuatro modelos adicionales (LPC [Kay y Marple, 1981], retardo de modelo esférico de la cabeza, retardo de la HRTF y retardo de la HRTF con escalización en magnitud) para generar archivos de espacialización a partir de una entrada monofónica, para ser evaluados en naturalidad del sonido producido, percepción y seguimiento de trayectorias y complejidad de implementación. Los resultados obtenidos demostraron que es posible alcanzar percepción lateral con modelos simples de espacialización, pero ningún modelo presentó confiabilidad con respecto a la simulación de elevación.

Palabras clave: binaural, sonido 3D, percepción de elevación y azimut.

Abstract: In this paper, four models for binaural perception based on the Head Related Transfer Function (HTRF) developed in MIT [Martin and Gardner, 1994] are presented and compared: linear prediction code (LPC) model of the HRTF (Kay and Marple, 1981) delay of the spherical head model, delay of the HRTF and delay of the HRTF plus magnitude equalization. Spatial hearing files from a monophonic source are generated and evaluated from the point of view of naturalness of the final sound, perception of path and follow up as well as complexity for implementation. Results showed that it is possible to get lateral perception with simple spatial hearing models but no one is reliable enough for simulation of elevation.

Key words: Binaural, 3-D sound, azimuth and elevation perception.

* *Ingeniero electrónico, magíster en ingeniería electrónica, Pontificia Universidad Javeriana. Profesor instructor, Departamento de Electrónica, Pontificia Universidad Javeriana. Correo electrónico: jhurtado@javeriana.edu.co*

** *Ingeniero electrónico, Pontificia Universidad Javeriana; Master of Science y PhD., Rensselaer Polytechnic Institute. Profesor titular, Departamento de Electrónica, Pontificia Universidad Javeriana. Correo electrónico: pvizcaya@javeriana.edu.co*

1. Introducción

El sonido tridimensional (3D), como usualmente se le conoce en la literatura técnica, se ha ido convirtiendo en los últimos años en parte importante de los sistemas científicos, comerciales y de entretenimiento [Begault, 2000] [Brown y Duda, 1998].

Desde un punto de vista estrictamente perceptivo, estamos constantemente expuestos a diversos sonidos que llegan a nuestros oídos, de los cuales somos capaces de interpretar la posición en el espacio de la fuente de dichos sonidos. Dicha interpretación tridimensional es un complejo sistema en el cual están involucradas características físicas como el tamaño y forma del torso, hombros, cabeza y orejas, que producen procesos de reflexión, refracción y difracción de onda. Posteriormente estas ondas llegan al oído y son procesadas y analizadas por el cerebro, que finalmente es quien realiza la interpretación espacial [Sibbald, ear, s.f.].

Existen dos fundamentos en los que se basa la percepción del espacio sonoro:

1. Fundamento fisiológico: la información es recogida por los dos oídos y es procesada posteriormente en el cerebro, dándonos una percepción unificada (denominada técnicamente "fusión binaural[1]") [García de la Torre, s.f.].
2. Fundamento psicológico: el aspecto psicológico de la audición se considera como el más importante, no existiendo una geometría en la percepción auditiva [García de la Torre, s.f.].

Estos dos fundamentos son los que marcan la pauta en la evaluación de los modelos desarrollados en este trabajo, ya que además de generar sonido binaural se quiso evaluar las características que dicho sonido generado presentaba de acuerdo con el modelo de espacialización que le fuese aplicado.

2. Desarrollo de los modelos

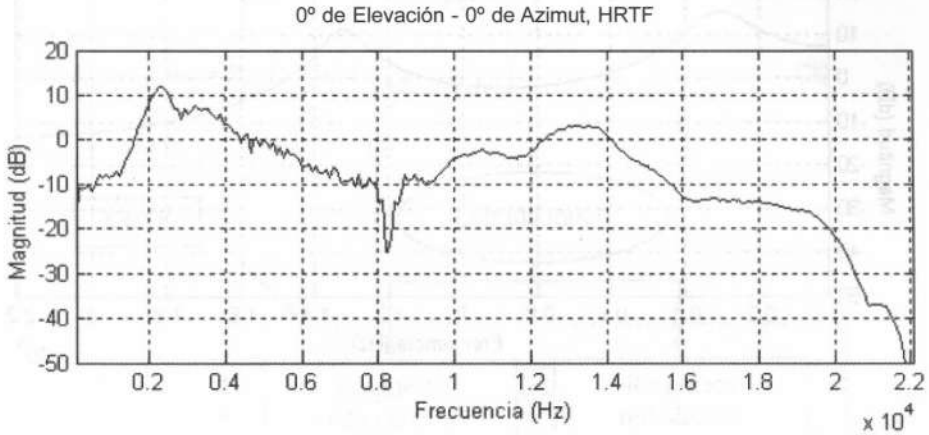
Los modelos a los cuales se les hizo la evaluación fueron HRTF, LPC, retardo con modelos de cabeza esférica (delay), retardo según la respuesta de la HRTF (delay DB) y retardo con escalización en amplitud según respuesta del HRTF (delay and mag).

A partir de la respuesta en frecuencia de la HRTF (Figura 1a) y en la respuesta en tiempo de la HRIR (Head-Related Impulse Response - Respuesta impulso de la cabeza) (Figura 1b), se determinó realizar un modelo LPC (Codificación Predictiva Lineal) de orden 4 [Gold y Morgan, 2000], [Martens, s.f.] [Stoica y Moses, 1997], de tal forma que se obtuviera una respuesta similar a la de la HRTF, pero con eliminación de redundancia, lo que se traduce en un modelo más simple (Figuras 2a y 2b).

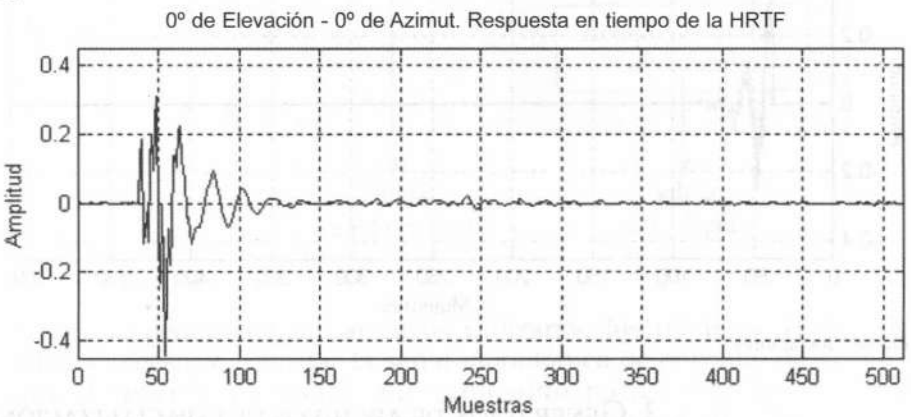
[1] *Binaural: capacidad de escuchar mediante la utilización de dos receptores u oídos.*

Figura 1. Respuesta en frecuencia (a) y en tiempo (b) de la HRTF de elevación 0° y Azimut 0°

a)



b)



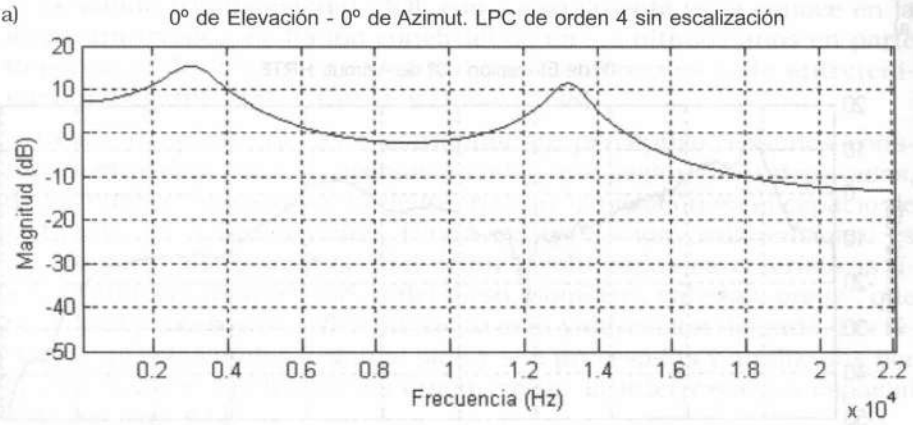
Fuente: los autores.

Con el ánimo de realizar modelos bastante simples a fin de evaluar su posible implementación en sistemas con procesamiento en tiempo real, se tomó el modelo de cabeza esférica para realizar los cálculos respectivos de retardo que habrían entre el tiempo de llegada de la señal a ambos oídos.

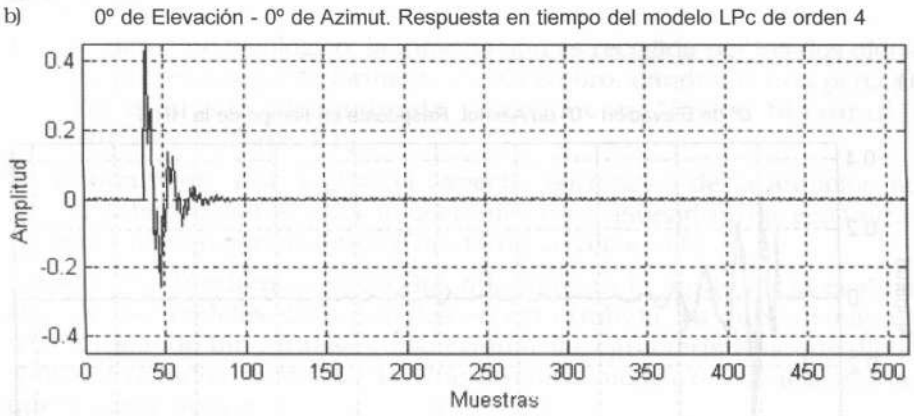
De igual forma, con los tiempos de retardo de la base de datos de la HRTF se realizó el otro modelo y, finalmente, agregándole la escalización en magnitud a este último modelo, se definió el quinto modelo.

Figura 2. Respuesta en frecuencia (a) y en tiempo (b) del Modelo LPC de la HRTF de 0° de elevación y 0° de Azimut

a)



b)



Fuente: los autores.

3. Generación de archivos de espacialización por medio de la HRTF y demás modelos

Una vez definidos los modelos, se procede a la realización de los archivos de espacialización. La forma de realizar dichos archivos se basa en la HRIR, que es la HRTF expresada en el dominio del tiempo, que al convolucionarla con la entrada monofónica que se quiere espacializar y posteriormente alimentar dicha señal de salida a un par de audífonos o parlantes debidamente compensados [Sibbald, virtual, s.f.] [Sibbald, hearing, s.f.] de tal forma que se minimice o evite la diafonía[2] [Brown y Duda, 1998] se logra el efecto de sonido tridimensional, como se puede ver en la Figura 3.

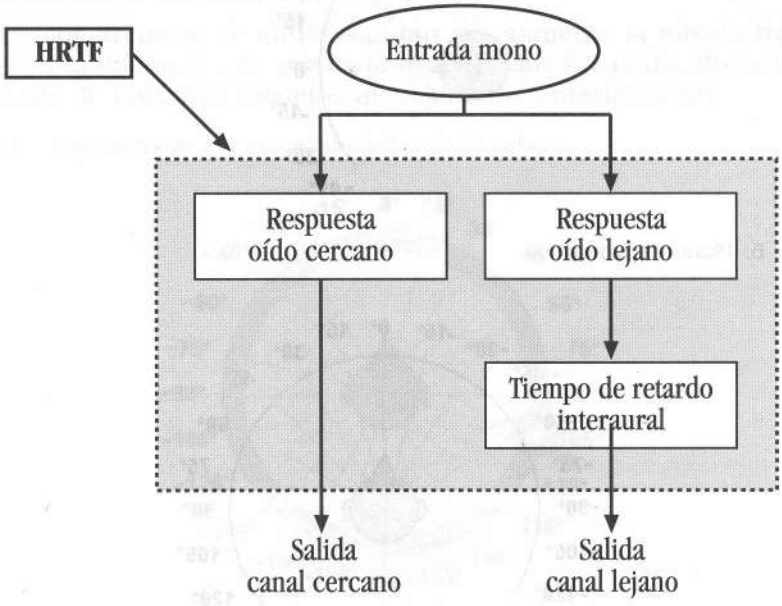
[2] *Cross talk o diafonía: acción por la que la señal de un canal aparece en canales adyacentes, como por ejemplo del canal izquierdo al canal derecho [Martens, s.f.].*

Las expresiones matemáticas que establecen las relaciones de lo anteriormente expuesto son:

$$\text{Salida oído izq} = X_{in}\left(t * \text{HRTF}_{\text{Oído izquierdo}}\right) \quad (1)$$

$$\text{Salida oído der} = X_{in}\left(t * \text{HRTF}_{\text{Oído derecho}}\right) \quad (2)$$

Figura 3. Modelo de síntesis binaural



Fuente: los autores.

Para la generación de archivos utilizando los modelos de retardo, sólo es necesario retardar la señal monofónica de entrada, de acuerdo con el retardo necesario en cada oído para crear el efecto de espacialización, como lo determinan las ecuaciones (3) y (4).

$$\text{Salida oído izq} = X_{in}\left(t\text{-}t_{\text{Arribo oído izquierdo}}\right) \quad (3)$$

$$\text{Salida oído der} = X_{in}\left(t\text{-}t_{\text{Arribo oído derecho}}\right) \quad (4)$$

Los modelos de HRTF y LPC sirven para generar tanto sonido con variación azimutal como con variación de elevación, mientras que los modelos a partir de retardos sólo se aplican a espacialización con 0° de elevación y entre 0° y 360° de azimut.

Elevación 0° y azimut 0° se toman como el punto frente a la nariz; hacia arriba de dicho punto la elevación se considera positiva y por debajo, elevación negativa. De igual forma, cualquier incremento de án-

gulo en el mismo giro de las manecillas del reloj se considerará ángulo de azimut positivo, tal y como se muestra en la Figura 4 y en la Figura 5, respectivamente.

Figura 4. Plano de elevación

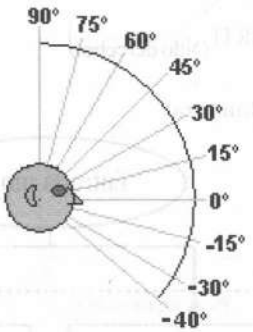
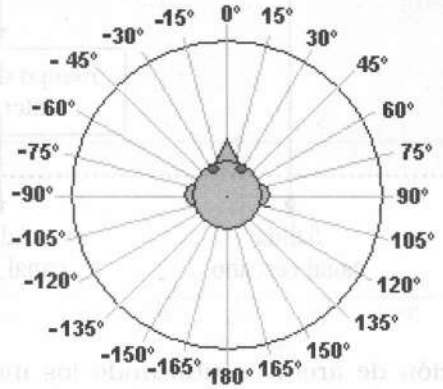


Figura 5. Plano de azimut



4. Evaluación de los modelos y resultados

La etapa de evaluación se efectuó mediante la aplicación de una encuesta en la cual las personas, de acuerdo con los archivos de audio que se les indicaban, debían responder acerca de la naturalidad y trayectoria aparente del movimiento (posición azimutal relativa y elevación) de los sonidos que se reproducían según la lectura del archivo.

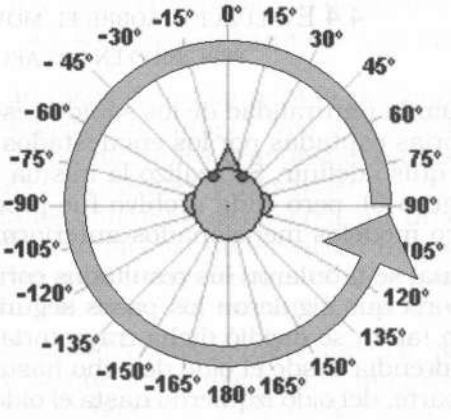
La encuesta fue realizada a un grupo de 35 personas voluntarias, 18 hombres y 17 mujeres. Para la realización de la prueba era indispensable que las personas usaran audífonos, los cuales eran suministrados en el sitio de la prueba, donde se tenía un ruido ambiente de oficina. Todas las pruebas fueron hechas en el mismo equipo, garantizando la misma tarjeta de sonido y el mismo programa de reproducción para los archivos de *audio wave*.

4.1 Evaluación de los modelos

El sonido que se les pedía comparar a las personas del grupo de prueba debía simular unos pasos alrededor de la persona, a la altura de las orejas (0° de elevación). La trayectoria de los pasos se iniciaba en el oído derecho y empezaban a desplazarse en el sentido contrario a las manecillas del reloj, pasando primero por el frente de la persona, luego por el oído izquierdo, posteriormente por detrás, para finalizar nuevamente junto al oído derecho. Cada paso está 15° después del paso anterior, tal como se muestra en la Figura 6.

Todos los archivos de audio seguían exactamente la misma trayectoria, con la diferencia de que cada uno de ellos fue realizado utilizando alguno de los cinco modelos mencionados anteriormente.

Figura 6. Trayectoria de los pasos alrededor de la cabeza



4.2 Resultados de la encuesta acerca de la naturalidad del sonido

Como se quería evaluar la calidad de los modelos realizados en comparación con la HRTF, se pidió a las personas encuestadas que escogieran el sonido que percibieran más natural, siempre comparando el sonido generado por medio del modelo de HRTF contra el sonido generado por la utilización de algún otro modelo.

Luego de hacer el análisis de los resultados de la encuesta, se obtiene una estimación del intervalo de confianza de la distribución muestral del 95% para cada uno de los modelos, en lo que se refiere a naturalidad, con lo cual se aprecia un bajo sobrelapamiento de los intervalos de confianza del modelo HRTF con respecto al modelo de delay y de delay HRTF. Con respecto a los modelos de LPC y delay and mag, los intervalos de confianza poseen un gran intervalo de sobrelapamiento, lo que significa que estos modelos, que son más simples, no difieren demasiado del modelo con HRTF en lo que a percepción de naturalidad se refiere. Esto podría facilitar la implementación de un sistema de espacialización en tiempo real, ya que el tiempo de procesamiento sería inferior al utilizar un modelo más simple que el HRTF.

4.3 Evaluación acerca de la naturalidad del sonido discriminado por sexo

Durante el análisis de resultados se notaron algunas diferencias entre apreciaciones que tuvieron los hombres y las mujeres, lo cual indica algunas tendencias diferentes de acuerdo con el género; por tal razón se procedió a realizar un análisis comparativo entre hombres y mujeres en lo que respecta a percepción de la naturalidad de los sonidos generados.

Luego del análisis se llegó a la conclusión de que existía una mayor tendencia de confianza de las mujeres por la HRTF, mientras que en el caso los hombres esta tendencia parece no existir o no estar plenamente definida, indicando diferentes inclinaciones de naturalidad según el modelo aplicado.

4.4 Evaluación sobre el movimiento azimutal percibido en los archivos escuchados

Además de evaluar la naturalidad de los sonidos escuchados, se evaluaron las trayectorias captadas por los encuestados con respecto a la trayectoria que se quiso definir. Se realizó la misma trayectoria en los cinco archivos (Figura 6), pero cada archivo fue procesado por medio de uno de los cinco modelos mencionados anteriormente.

La forma en la cual se tabularon los resultados correspondió directamente a la trayectoria que siguieron los pasos según la percepción de cada persona. Para tal fin se dividió dicha trayectoria en dos partes: la primera parte comprendía desde el oído derecho hasta el oído izquierdo (ida) y la segunda parte, del oído izquierdo hasta el oído derecho (vuelta).

Los resultados obtenidos al discriminar la trayectoria de ida y de vuelta llevan a afirmar que la mayoría de las personas no alcanza a discriminar adecuadamente los sonidos de adelante y de atrás con tan sólo el uso de modelos utilizados para la simulación de especialización; sin embargo, con estos mismos modelos sí se logran identificar los sonidos laterales.

Debido a que los resultados anteriores son contradictorios en lo que se refiere a la intención de movimiento con respecto a la percepción de movimiento, se decidió entonces realizar el análisis de resultados no solamente por trayecto de ida y vuelta, sino tomando en cuenta todo el trayecto realizado, de tal forma que las respuestas dadas correspondan a la composición de ambos trayectos.

Al realizar el análisis de porcentaje de confianza de la trayectoria verdadera, con respecto a la trayectoria percibida, el resultado indica que la mayoría de las personas no discrimina adecuadamente los trayectos propuestos; la percepción de la trayectoria por detrás es la que posee mayor cantidad de votos de la muestra; sin embargo, de acuerdo con el análisis estadístico, la proporción no llega a ser lo suficiente-

mente alta, es decir, tampoco se puede afirmar que las personas escuchan los sonidos por detrás cuando sólo son estimuladas por sonidos binaurales, es decir, ninguno de los modelos aplicados resultó ser completo para utilizarlos en la proyección de espacialización del sonido.

De igual forma, los resultados de la encuesta permiten establecer que aunque la intención del archivo de audio era dar la percepción de movimiento por delante y luego por detrás de la cabeza, una buena proporción de personas percibieron el efecto contrario (por detrás y luego por delante). Aunque esta afirmación puede sonar un poco pesimista en principio, proporciona nuevas áreas de investigación, debido a que realmente sí se logra crear la sensación de movimiento, algo que en definitiva se quería alcanzar con los modelos; pero lamentablemente dicha percepción de movimiento no es localizada por la mayoría de las personas encuestadas en los lugares que los archivos de audio quieren indicar.

4.5 Evaluación sobre el movimiento de elevación en los archivos que sólo tenían simulación de movimiento azimutal

Aunque en los archivos de audio para la percepción de movimiento de azimut no se implementó ningún tipo de elevación –siempre estuvo en 0°–, sí existieron múltiples respuestas en las cuales las personas afirmaban diversos niveles y movimientos de elevación. Así, al analizar los resultados obtenidos se observa que una buena proporción de personas percibieron efectos de elevación independientemente del modelo utilizado. De igual forma, se realizó el análisis tomando ambos trayectos, el de ida y el de vuelta, como un sólo trayecto completo.

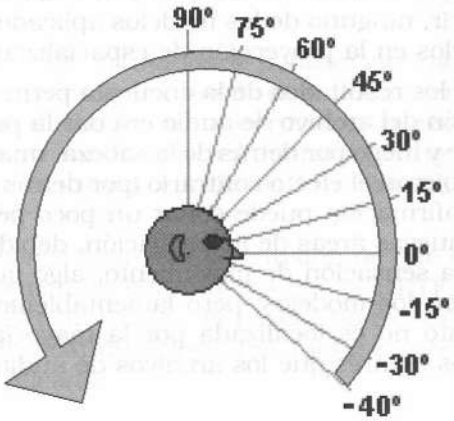
Con los dos análisis anteriores se pudo apreciar que aunque el archivo de audio no haya querido implementar ninguna elevación, varias personas sí sintieron dicho efecto y, al realizar el análisis estadístico de proporción, el resultado indica que la mayoría de las personas sí percibieron efectos de elevación en el sonido, en contraposición de lo se quiso realizar en el archivo de audio.

4.6 Evaluación sobre el movimiento azimutal en los archivos que sólo tenían simulación de movimiento de elevación

Al igual que se realizaron archivos para identificar la percepción azimutal, se realizaron archivos para la percepción de la elevación. La trayectoria que se pretendía simular era de unos pasos que empezaban en -40° frente a la persona (cada paso estaba 10° más arriba del anterior), pasaban por encima de la cabeza y terminaban nuevamente en -40° detrás de la persona (cada paso estaba 10° más abajo del anterior), tal como se muestra en la Figura 7.

Los archivos fueron desarrollados por medio de los modelos de HRTF y LPC. Aunque el archivo tenía como intención simular el efecto de elevación, este efecto fue percibido por una buena cantidad de las personas, pero no de forma coherente, es decir, la mayoría percibió diferentes tipos de trayectorias de elevación.

Figura 7. Trayectoria de los pasos por encima de la cabeza



Al analizar si se percibía trayectoria azimutal en un archivo que no tenía este tipo de variación, se observaron diferentes percepciones de las personas encuestadas con respecto al azimut, donde la mayoría de ellas indica alguna trayectoria significativamente diferente a la diseñada en el archivo de audio.

4.7 Aplicación de menús auditivos y espacialización estática

Con el fin de demostrar la posibilidad de uso comercial de los modelos se realizó la espacialización de personajes en la narración de un texto con varios de ellos. En dicha espacialización cada personaje es ubicado en un lugar diferente para facilitar a quien escucha su identificación.

En el ámbito musical se realizaron grabaciones de estudio de tres instrumentos (guitarra, voz masculina y voz femenina) y posteriormente se les hizo una espacialización de tal forma que se pretendía crear un ambiente de espacio y localización de los instrumentos.

5. Conclusiones

El modelo de la HRTF es incompleto debido a que no proporciona adecuadamente los resultados de espacialización requeridos. Los otros modelos desarrollados, aunque son más simples que el de la HRTF, no difieren demasiado de este modelo en lo que a aceptación se refiere. Los costos en cuanto a especificaciones debido a una posible implementación en tiempo real de los modelos de HRTF en comparación con los demás modelos serían menores para los segundos en lo relacionado con tamaño de memoria y velocidad de procesamiento.

Aunque los modelos de retardo sólo se apliquen a localización azimutal con elevación 0°, su simplicidad los hace efectivos en aplicaciones de espacialización musical o menús auditivos virtuales.

La tendencia de las personas encuestadas a inferir que el sonido estaba detrás de ellas, puede ser debida al procesamiento que hace el cerebro, debido a que los demás sensores (vista, tacto) no advierten sobre la presencia de dicha fuente sonora. Para obtener mejores resultados en cuanto espacialización se refiere, sería necesario involucrar más sentidos para aumentar la percepción espacial, como por ejemplo la visión.

Los modelos utilizados para la espacialización con elevación son de rendimiento bastante bajo. Esto puede ser debido a que las personas no poseen ningún punto de referencia que pueda ubicarlos directamente a cierta altura de la fuente sonora, como ocurriría si escucháramos algún sonido a cierta altura sin tener ninguna referencia adicional para poder localizarlo.

En definitiva, es posible simular espacialización lateral por medio de dos fuentes, mediante el uso de un par de audífonos, pero con este mismo par de audífonos, utilizando los modelos mencionados, no fue posible simular adecuadamente los efectos de elevación. En la vida diaria nosotros no logramos percibir una fuente sonora fácilmente sin un adecuado direccionamiento de nuestros sensores auditivos, que en nuestro caso se logra con el movimiento de cabeza para ubicar las orejas en la dirección en la cual el sonido tiene más energía. Así mismo, para lograr mayor realidad, sería necesario la implementación de un sistema en tiempo real para que las personas logren ubicar más fácilmente la fuente sonora.

El uso de procesamiento binaural en menús auditivos y espacialización musical podría aumentar el nivel de aceptación del sonido sin incrementar los costos de producción. Es posible la creación de archivos de audio en procesamiento de *playback* para la generación de espacialización, lo que puede hacerse sin tener la obligación de realizar dichas grabaciones con espacialización, sino simplemente grabando cada señal por un canal diferente.

Referencias

Begault, D. R. (2000), *3-D Sound For Virtual Reality and Multimedia*, NASA.

Brown, Ph., R. O. Duda (1998), A Structural Model for Binaural Sound Synthesis, en: *IEEE Transaction on Speech and Audio Processing*, Vol. 6, No. 5, September.

Diccionario bilingüe de Audio Profesional (s.f.), en: http://www.doctorproaudio.com/doctor/diccionario.htm.

García de la Torre, A. (s.f.), Música y espacio, en: http://www.espacioluke.com/Marzo2001/alfonso.html.

Gold, B., N. Morgan (2000), *Speech and Audio Signal Processing*, John Wiley and Sons.

HeadRoom Corporation (s.f.), *The Psychoacoustics of Headphone*.

KAY, S., S. MARPLE (1981), Spectrum Analysis - A Modern Perspective, en: *Proceedings of the IEEE*, Vol. 69, No. 11, November.

MARTIN, K., B. GARDNER (1994), HRTF Measurements of a KEMAR Dummy-Head Microphone. Media Lab Perceptual Computing, MIT, May.

MARTENS, W. L.(s.f.), *Spatial Audio Terminology*, University of Aizu.

STOICA, P., R. MOSES (1997), *Introduction to Spectral Analysis*, Prentice Hall.

SIBBALD, A.(s.f.), Virtual Ear Technology. Sensaura. *3D Positional Audio.*

SIBBALD, A.(s.f.), Virtual Audio for Headphones. Sensaura. *3D Positional Audio.*

SIBBALD, A. (s.f.), Hearing in Three Dimensions. Sensaura. *3D Positional Audio.*

SIBBALD, A. (s.f.), An Introduction to Sound and Hearing. Sensaura. *3D Positional Audio.*