TECNIACUSTICA AVILA · 1999

ESTUDIO SUBJETIVO DE DISTINTOS MODELOS DE LOCALIZACIÓN DE FUENTES

REFERENCIA PACS: 43.66

Jesús Alba, Joaquín Martínez, José Javier López, Jaime Ramis.
Departamento de Comunicaciones
Departamento de Física Aplicada
Escuela Politécnica Superior de Gandia
Ctra. Nazaret-Oliva s/n
46730 GRAO DE GANDIA, Valencia
Tel: 34 962 849 300
Fax: 34 962 849 309
E-mail: jjlopez@dcom.upv.es, jramis@fis.upv.es, jesalba@fis.upv.es

SUMMARY

This paper presents results of a subjective study carried out with a set of people in order to compare different source localization models in spatial hearing for 3-D sound reproduction. Only sources in the horizontal plane have been studied. ITD, IID, combination of both and HRTF based models have been compared. Results obtained are coherent with theory giving in many cases valuable information that could be directly used in the design of virtual realty systems.

1. INTRODUCCIÓN

A partir de los estudios fisiológicos sobre localización espacial de fuentes [1], se pueden crear modelos de localización de distinta complejidad que den lugar a sistemas de síntesis de fuentes virtuales en distintos puntos del espacio. Existen gran cantidad de referencias, algunas incluso de gran antigüedad, [2] [3], donde se tratan estos temas. Sin embargo, no se disponía de ninguna experiencia cercana previa en estos aspectos, tan intrínsecamente relacionados con la fisiología del sistema auditivo humano y por tanto tan subjetivos y difíciles de interpretar desde un punto de vista teórico, por lo que se decidió realizar una experimentación *insitu* en nuestros laboratorios que contribuyera al conocimientos sobre dichos aspectos y su aplicación al diseño de sistemas de sonido 3-D.

Para ello, se realizó un estudio subjetivo con distintas personas para comparar distintos modelos de localización de fuentes, limitándonos en principio a la localización en el plano horizontal (la sencillez y la naturaleza de los modelos ITD e IID no permiten realizar estudios en elevación). La reproducción se realizó colocando unos auriculares sobre los oídos del oyente, con el objeto de independizar el estudio de otros factores que pueden aparecer en la reproducción mediante altavoces (transaural). Los distintos modelos que se han comparado son:

- Diferencia de Tiempo Interaural (ITD)
- Diferencia de Intensidad Interaural (IID)
- Combinación de ITD y IID
- Base de datos de medidas de la HRTF de un maniquí (KEMAR)

Los tres primeros modelos son sencillos de implementar a nivel software. La ITD se implementa retardando convenientemente la señal entre los dos oídos. La IID se implementa aplicando factores de ganancia diferente a cada una de las dos señales binaurales. Para el último se utilizaron las medidas de la respuesta al impulso de la cabeza (HRTF) realizadas sobre un

maniquí acústico comercial (KEMAR) que se encuentran disponibles al dominio público en el MIT, [4], medidas por el método de las MLS, [6].

2. CONFIGURACIÓN DE LOS EXPERIMENTOS

Hay que señalar que para llevar a cabo esta experiencia se ha contado con la colaboración de treinta oyentes. Cada oyente se expone ante una secuencia de 64 eventos sonoros. En principio no hay dos eventos sonoros iguales, pues cada uno de ellos se corresponde con una posición del espacio para uno de los cuatro modelos propuestos.

Con cada modelo se han procesado 16 eventos diferentes desde 0º hasta 360º, es decir, las posiciones tienen una separación de 22,5º. Recordemos que no existe ningún tipo de elevación, sólo se ha variado el acimut dentro del plano horizontal. Si por cada uno de los cuatro modelos utilizados se procesan 16 eventos diferentes obtenemos un total de 64 sucesos sonoros diferentes tal como indicábamos anteriormente.

La reproducción de los sucesos sonoros se ha realizado mediante auriculares. Las 64 secuencias de sonido han sido barajadas para exponerlas de forma aleatoria. De este modo, los oyentes tendrían más dificultades para reconocer la procedencia de los sonidos y el método es más fiable. Más adelante veremos que estas dificultades han influido indirectamente en los resultados, y que evitarlas hubiera supuesto el desarrollo de un estudio en profundidad sobre asuntos que en principio resultaban irrelevantes.

La señal de prueba utilizada fue un ruido de banda ancha similar al ruido blanco. Como se puede comprobar en [1], de este modo, la evaluación del suceso sonoro es más crítica. La señal era la misma para las 64 secuencias procesadas. La duración aproximada de la misma era de unos dos segundos. Mediante un bucle, cada una de las secuencias se presentaban indefinidamente, hasta que el individuo bajo prueba estuviera seguro de su decisión.

Antes de comenzar la exposición de los sucesos sonoros, a los sujetos se les enviaba cuatro sonidos de referencia correspondientes a 0, 90, 180 y 270 grados. Los sonidos de referencia estaban procesados empleando la base de datos (HRTF del maniquí). Indudablemente estas señales son las que mayor calidad presentan tal como veremos más adelante.

En principio, podríamos pensar que los oyentes, tras una sesión tan extensa, se verían afectados por la fatiga. En consecuencia, éstos podrían perder precisión en la evaluación subjetiva de los sucesos. Pero no fue el caso. Conforme la experiencia iba llegando a su fin, los sujetos iban afinando su oído.

Como consecuencia de esta adaptación, se decidió realizar dos grupos de experimentos. En el primer grupo los oyentes no se han sometido a ningún entrenamiento previo, mientras que en el segundo grupo si se han sometido a entrenamiento. Por tanto las primeras representaciones obtenidas, poseen alguna dispersión no deseada para ciertos ángulos. No obstante, de las gráficas correspondientes a esta primera experiencia se obtienen conclusiones que confirman las predicciones teóricas.

3. RESULTADOS SIN ENTRENAMIENTO PREVIO DE LOS SUJETOS

Se sometió a los treinta oyentes a los 64 eventos acústicos comentados anteriormente. Los resultados para los cuatro modelos de localización se muestran en la figura 1.

La interpretación de las gráficas que se muestran en las siguientes figuras es muy sencilla: el eje de abscisas representa el ángulo real del suceso sonoro, mientras que en el eje de ordenadas encontramos el ángulo percibido por el oyente. Idealmente debería aparecer sobre la

bisectriz de los dos ejes un cuadrado por cada ángulo real. Sin embargo esto no es así. En función de la cantidad de sujetos que coincidan en señalar una determinada dirección de la fuente, los cuadrados serán más grandes o más pequeños. Por lo tanto, se está representando el histograma bidimensional, correspondiente a las posiciones presentadas y percibidas de la fuente virtual.

Observando las gráficas de la figura 1r, se puede pensar que la distinción delantera/trasera para el ángulo de 0º no es muy acertada. Esto es debido a que, a lo largo de la experiencia llevada a cabo con cada uno de los voluntarios, el primer sonido que se le presentaba era el correspondiente a 0º. Como hemos indicado anteriormente, el aprendizaje ha jugado un papel muy importante en la precisión de los sujetos. Esto nos induce a pensar que los resultados obtenidos son absolutamente coherentes.

a) modelo ITD.

b) modelo IID.

c) modelo combinado ITD + IID.

d) modelo con la HRTF de un maniquí.

Figura 1. Resultados de la localización para los distintos modelos.

El modelo ITD mostrado en la figura 1a, sin ninguna duda es el que peores resultados reporta. Tal y como sabemos, el nivel de presión sonora sobre los dos oídos siempre es el mismo para cualquier posición de la fuente sonora. Sólo la fase de la señal varía en función de la posición de la fuente en el espacio. Este hecho dificultaba gravemente la localización con éxito de la fuente sonora virtual.

Salta a la vista que la distinción delantera trasera es nefasta. Para un valor cualquiera del ángulo de la fuente comprendido entre 0º y 180º, los sujetos (en general) eran incapaces de ubicar a la fuente en el cuadrante adecuado (es decir, de 0º a 90º ó de 90º a 18 0º). Por ello se observa una elevada dispersión entre 0º y 180º cuando la fuente se encontraba a la derecha del sujeto y algo similar entre 180º y 360º cuando la fuente se encontraba a su izquierda.

El modelo IID mostrado en la figura 1b proporciona resultados mejores, aun así, el resultado se haya bastante lejos de parecerse al que proporciona la HRTF. Si comparamos la figura 1b con la figura 1d, veremos que en la gráfica correspondiente al modelo IID los cuadrados parecen estar un poco más organizados, da la impresión de que intentan acercarse un poco a la bisectriz. Por otra parte se observa una gran dispersión en los datos obtenidos para el ángulo de cero grados. Del mismo modo que ocurría anteriormente, este suceso sonoro fue expuesto entre los primeros (concretamente el segundo). Esto sumado al gran parecido con el anterior (HRTF a cero grados), redunda en la desorientación del sujeto.

En la figura 1c, se muestra la combinación de los dos modelos matemáticos ITD más IID. De la distribución de cuadrados no podemos llegar a obtener conclusiones distintas de las comentadas hasta el momento. Tal vez estos elementos se encuentren más próximos a la bisectriz que los otros dos modelos matemáticos pero prácticamente es inapreciable.

La posición correspondiente a cero grados procesada mediante este modelo, se expuso también entre las tres primeras. Esto ha influido en la precisión de los sujetos por dos motivos ya conocidos. Porque no podían haberse adaptado en tan corto espacio de tiempo, y porque los tres sonidos eran prácticamente iguales.

Ya hemos comprobado que el modelo ITD es el menos afortunado de los tres modelos matemáticos. Sin embargo, aunque parezca un tanto paradójico, para el ITD se ha obtenido una dispersión menor de resultados para el ángulo de cero grados que con el resto de sistemas. Esto encuentra su justificación en el hecho de que el evento sonoro correspondiente a este ángulo se emitiera el noveno en la secuencia de los sesenta y cuatro sucesos. Un hecho que refuerza aún más la idea relativa al aprendizaje y la adaptación de las personas en experiencias similares a las descritas en este capítulo.

4. RESULTADOS CON ENTRENAMIENTO PREVIO DE LOS SUJETOS

Para solucionar los problemas de aprendizaje de los sujetos y el reflejo de éstos sobre los resultados de las medidas, se creyó oportuno repetir la experiencia siendo un poco más cautos.

Para asegurar la adaptación de cada uno de los sujetos bajo prueba al modelo de localización de fuentes, se emitieron hasta treinta y dos eventos sonoros en una fase previa de aprendizaje. De este modo se asegura una adaptación previa de los oyentes mejorando la localización de eventos posteriores. Hemos contado con la colaboración de sólo diez sujetos en el desarrollo de esta segunda fase de la experiencia.

En las figuras 2 se representan los resultados con entrenamiento previo, correspondientes a los diez sujetos que se presentaron a las pruebas. De estos diez oyentes, la mitad utilizaron unos auriculares de un modelo, mientras que la otra mitad utilizó un modelo distinto. Esto se

utilizará en apartados posteriores para verificar la influencia de dicha variable sobre la localización. En la presentación de los resultados en este punto, no se realiza distinción alguna entre los que llevaban unos auriculares u otros.

Queda claro que existe una migración de los cuadrados hacia la bisectriz de los ejes. Se concluye entonces, que la precisión en la localización de las fuentes virtuales por parte de los oyentes, mejora de forma sensible si son sometidos a un entrenamiento previo.

5. INFLUENCIA DEL TIPO Y CALIDAD DE LOS AURICULARES

Con el objeto de contrastar la influencia de los auriculares, de los diez sujetos estudiados en el apartado anterior, cinco de ellos utilizaron unos auriculares de mayor calidad que presentaban un diseño más ergonómico y una mayor fidelidad.

Los auriculares de mayor calidad se corresponden con el modelo HD-545 de la marca Sennheiser. Son auriculares del tipo *abierto*, pero dotados de un borde de material blanco y esponjoso que rodea todo el pabellón auricular sin aplastarlo contra la cabeza y que proporciona un buen contacto del auricular con ésta, mejorando la reproducción de las bajas frecuencias.

Los auriculares considerados como de inferior calidad (a tenor de los resultados obtenidos), se corresponden con el clásico modelo que produce una presión del pabellón auricular con la cabeza tras su colocación.

Podemos afirmar que la respuesta de los auriculares influye notablemente en la valoración subjetiva de los voluntarios. Si nos fijamos con detalle en las gráficas a y b de la figura 3, apreciamos que la distinción delantera/trasera empleando los auriculares de mejor calidad es mucho mayor que haciendo uso de los de calidad inferior.

A pesar de que todas las personas que se han sometido a este experimento han tenido dificultades para distinguir entre 0º y 180º en sus correspondientes sucesos sonoros, se puede apreciar una ligera mejora en la gráfica de la figura 3 respecto de cualquier otra representación para los ángulos mencionados de las presentadas anteriormente en este artículo.

a) modelo ITD. c) modelo combinado ITD + IID.

b) modelo IID. d) modelo con la HRTF de un maniquí.

Figura 2. Resultados de la localización con entrenamiento previo de los oyentes.

a) auriculares de buena calidad. b) auriculares de peor calidad.

Figura 3. Influencia de la calidad de los auriculares en la localización.

6. CONCLUSIONES

Para obtener resultados todavía mejores que con la HRTF medida en el maniquí, la experiencia tendría que desarrollarse contando con la HRTF individualizada, propia del sujeto bajo prueba como base para el filtrado.

Jens Blauert sugiere en [1] que, para una presentación perfecta de sonido 3-D mediante auriculares, es necesario medir cada una de las HRTF de los oyentes potenciales. Sin embargo Wenzel en [5] propone, para superar la imposibilidad (en la práctica) de medir la HRTF de cada

oyente, el entrenamiento previo de los oyentes que van a usar el sistema 3-D, usando unas HRTF no individualizadas estándar.

En la actualizad, las tendencias parece que se sitúan en la línea de disponer de un conjunto de HRTF's más o menos amplio y aplicar el más adecuado al oyente que va a utilizar el sistema en función de algún aspecto físico sencillo de su anatomía, como puede ser: la distancia interaural, el tamaño de la cabeza, el tamaño de la oreja, etc.

Un estudio más minucioso, con HRTF's individualizadas podría llegar a concluir cuál es la limitación real de la reproducción mediante auriculares, y obtendríamos de forma absoluta la desviación de cada uno de los modelos, por comparación con los resultados que genera la HRTF de cada persona.

Para concluir, cabe añadir el interés de estos estudios subjetivos, como se puede comprobar en otros trabajos similares de reciente aparición (Octubre de 1998) en el *Journal of the Audio Enginering Society* [7], el cual presenta gran similitud con el que se ha presentado en este artículo.

REFERENCIAS

[1] Blauert J., Spatial Hearing (Revised Edition), MIT Press, Cambridge, MA, 1997
[2] Rayleigh, On our Perception of Sound Direction,. Phil. Magazine 13 (6) pp. 214-232 (1907)
[3] Rayleigh, On the Acoustic Shadow of a Sphere, Phil. Transactions of the Royal Society 203 pp. 87-99 (1904)
[4] Gardner W.G. & Martin K., HRTF Measurements of a KEMAR Dummy-Head Microphone, MIT Media Lab Perceptual Computing Internal Report 280, MIT, Boston, MA, 1994
[5] Wenzel, E. M., Wightman F. L., Kistler D. and Foster S. H., The Convolvotron: Realtime synthesis of out-of-head localization, Proceeding of Second Joint Meeting Acoustic Societies of America and Japan, Honolulu, H, 1988
[6] Schoroeder, M.R., '*Integrated impulse method measuring sound decay without impulses'*. J.Acoust. Soc. Am., 66 (1979) 497-500
[7] Zhang M.,Tan K. and Er M.H., Three-Dimensional Sound Synthesis Based on Head-Related Transfer Functions, Journal of the Audio Engineering Society 46 (10) pp. 836-844 (1998)