

## COMPARACIÓN DE LA PERCEPCIÓN SONORA DE DISTANCIA CON LOS SISTEMAS WAVE FIELD SYNTHESIS Y VECTOR BASE AMPLITUDE PANNING

REFERENCIA PACS 43.66.Qp

Gutiérrez-Parera, Pablo; López Monfort, José Javier  
Universidad Politécnica de Valencia  
iTeAM, Instituto de Telecomunicaciones y Aplicaciones Multimedia  
Camino de Vera s/n, Edificio 8G esc D  
46022 Valencia  
Tel: +34 963 877 007 (Ext. 73008) Fax: +34 963 879 583  
pabgupa@iteam.upv.es; jjlopez@dcom.upv.es

### ABSTRACT

Creating a realistic distance perception with spatial audio reproduction systems is not an easy task. Wave Field Synthesis (WFS) is capable to synthesize the wavefront curvature produced by a virtual source. Previous studies suggest that this curvature can be an additional cue for the listener to extrapolate distance. A subjective perceptual test has been carried out to compare the capabilities of WFS and Vector Base Amplitude Panning (VBAP) to reproduce accurately sound distances. Different variables were studied; type of sound, listening angle and reverberation at different distances. The analysis of the collected data suggests that WFS is better at reproducing distances than panning systems.

### RESUMEN

Reproducir correctamente la percepción de distancia sonora no es una tarea fácil para los sistemas de sonido espacial. Wave Field Synthesis (WFS) sintetiza la curvatura del frente de onda producida por una fuente sonora virtual. Estudios previos sugieren que esta curvatura puede proporcionar al oyente información adicional para extrapolar la distancia. Se llevó a cabo un test de percepción subjetiva para comparar la capacidad de WFS y Vector Base Amplitude Panning (VBAP) para reproducir distancias sonoras con precisión. Se estudiaron distintas variables: tipo de sonido, ángulo de escucha y reverberación a varias distancias. El análisis de los datos obtenidos sugiere que WFS reproduce mejor las distancias que los sistemas de panoramización.

### 1. INTRODUCCIÓN

En cualquier campo sonoro existe una sensación de profundidad espacial. Esta sensación es la responsable de la percepción de la perspectiva en una escena acústica [1]. La profundidad se considera un atributo esencial para la percepción sonora espacial de la escena acústica, y en esta profundidad cumple un papel primordial la distancia sonora de la fuente.

Sistemas como WFS y VBAP, reproducidos sobre *arrays* de altavoces dispuestos envolviendo a los oyentes, tienen una gran capacidad para posicionar sonidos en diferentes ángulos del plano horizontal alrededor del oyente [2] [3]. Sin embargo su capacidad para situar sonidos en diferentes distancias no ha sido comparada, pudiendo resultar un criterio importante a la hora de conseguir un sonido espacial más completo.

Se han descrito varios factores responsables de la percepción de la distancia por medio del sonido (también llamados *indicios* acústicos), de los cuales los principales son:

- *Nivel de presión sonora*: En general, la presión sonora para una posición concreta decrece cuando aumenta la distancia entre el oyente y una fuente sonora de nivel constante. En condiciones ideales (una fuente puntual en campo libre) la disminución del nivel de presión sonora en función de la distancia sigue la ley del cuadrado inverso, por la que se produce una reducción de 6dB del nivel de presión cada vez que se dobla la distancia a la fuente. El nivel de presión sonora se considera el factor más claro y evidente para determinar la distancia mediante la escucha [4].
- *Relación sonido directo - sonido reverberante*: En entornos con superficies reflectantes, cuando la fuente sonora está cerca al oyente éste percibe más cantidad de sonido directo que reverberante, y a medida que la fuente se aleja la cantidad de sonido directo que percibe se reduce, aumentando el sonido reverberante. Este factor es por tanto determinante en espacios cerrados, donde siempre hay un cierto grado de reverberación. Sin embargo, en muchos espacios abiertos también se producen reverberaciones y por tanto la relación sonido directo-sonido reverberante también varía con la distancia.
- *Espectro*: Para distancias mayores de unos quince metros las propiedades absorbentes del aire atenúan las altas frecuencias. Por otro lado, cuando la fuente sonora está muy cerca de la cabeza del oyente se puede producir un aumento de las bajas frecuencias, debido a la curvatura del campo sonoro.
- *Diferencias interaurales*: Las diferencias relativas de la onda acústica al llegar a los oídos se suelen clasificar en diferencias interaurales de tiempo (ITD) y diferencias interaurales de nivel (ILD). Las ITD prevalecen a frecuencias bajas hasta aproximadamente 1–1,5kHz, mientras que las ILD tienen mayor influencia en la localización de fuentes a partir de la mencionada frecuencia. Además, las diferencias interaurales tienen una relación directa con la forma del frente de onda, dependiendo de la distancia a la fuente sonora la curvatura del frente de ondas se atenúa hasta quedar como onda plana. Esta variación es percibida por el oyente [5] y junto con ligeros movimientos de cabeza puede emplear esta información para realizar un paralaje acústico que le ayude a percibir mejor la distancia.

Existe otro factor adicional de suma importancia para la localización espacial sonora y en concreto para la determinación de la distancia, que es el *conocimiento de la fuente sonora y del entorno acústico*. Este factor se puede considerar como complementario a todos los anteriores y en muchos casos es determinante para la correcta localización. La familiaridad con el tipo de fuente sonora en cualquier caso siempre ayuda a calcular las distancias, dado que se puede hacer una estimación de la potencia acústica de la fuente, así como del contenido espectral y temporal de la misma.

## 2. MOTIVACIÓN Y OBJETIVOS

WFS permite sintetizar en toda la zona de escucha la curvatura del frente de ondas que llega al oyente, proveniente de una fuente localizada a una cierta distancia. Algunos estudios afirman que la sensación de distancia sonora es difícil de percibir mediante WFS [6], sin embargo, otros trabajos [7] sugieren lo contrario. Con la intención de ayudar a esclarecer este aspecto, se ha desarrollado un test de percepción subjetiva. Dicho test busca comparar la capacidad para reproducir una sensación de distancia realista tanto de WFS como de un sistema de panoramización por amplitud como VBAP.

En principio, la curvatura del frente de ondas tiene un efecto de más importancia para la *función de transferencia de la cabeza (HRTF)* en distancias cortas. Sin embargo, los pequeños movimientos que el oyente realiza en una escucha normal, pueden aportar información adicional de paralaje que ayude a extrapolar la distancia de la fuente sonora, incluso cuando ésta se encuentre en campo lejano. En base a este razonamiento, el objetivo de este trabajo es averiguar si WFS proporciona alguna ventaja para la percepción de la distancia o no, comparado con sistemas que emplean el efecto *phantom* mediante panoramización por amplitud. Se eligió VBAP como sistema de panoramización por amplitud porque proporciona ecuaciones analíticas para el caso de múltiples altavoces.

En estos experimentos, los oyentes permanecieron sentados, pero podían mover ligeramente la cabeza con la intención de que pudieran percibir hipotéticos efectos de paralaje que ayudaran a discernir la distancia. Se estudió la influencia de varios factores: el tipo de sonido, el ángulo de escucha y la reverberación, a distintas distancias.

### 3. TEST DE PERCEPCIÓN SUBJETIVA

Se diseñó un test de percepción según un modelo con respuestas de *escala directa por intervalos*. Este tipo de prueba permite realizar análisis mediante métodos cuantitativos, basados en la asunción de una distribución normal, permitiendo hacer pruebas-t y ANOVAs [8].

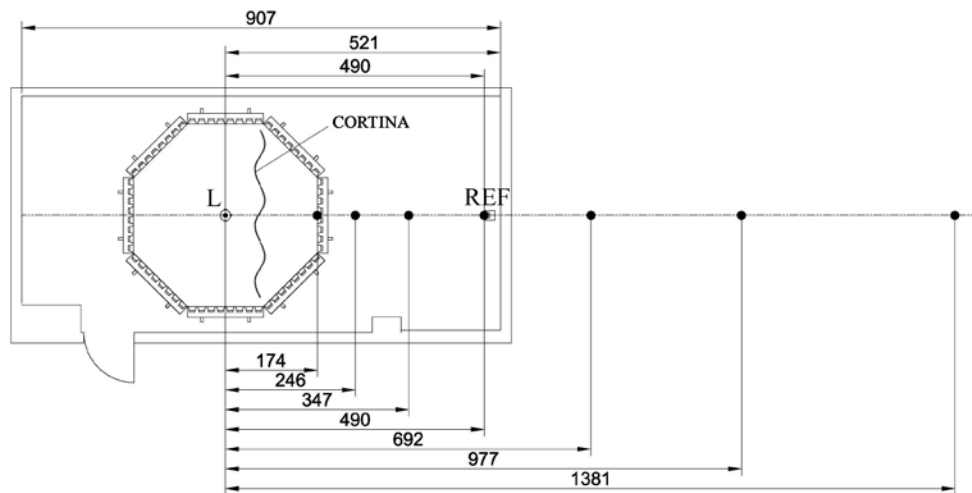
La mecánica básica del test consistió en comparar los sonidos reproducidos por los sistemas WFS y VBAP con una referencia externa. Los sistemas WFS y VBAP sintetizaban las distintas distancias por medio de unos *arrays* de altavoces situados en una posición fija con respecto al oyente, mientras que la referencia externa se reproducía por un altavoz independiente situado a una distancia fija. De esta manera, cada sonido sintetizado a una distancia concreta (ya fuese por WFS o VBAP) debía ser comparado y evaluado con otro sonido reproducido a una distancia fija por el altavoz externo, debiendo determinarse la distancia relativa del sonido sintetizado con respecto a la posición de referencia.

La figura 1 muestra un esquema del montaje que se realizó para el test. Para reproducir WFS y VBAP se empleó un *array* octogonal de 64 altavoces (con separación entre sus centros acústicos de 18 cm) dispuesto alrededor de la posición del oyente. En la figura 2 se encuentran marcados en gris los altavoces que intervenían en la reproducción de VBAP. Para evitar la influencia visual durante la realización del test se usó una cortina acústicamente transparente colocada delante del oyente. El test se realizó en una sala tratada acústicamente con un  $T_{60}(1\text{kHz}) < 0,25\text{s}$  y un volumen de  $96\text{m}^3$ , que pertenece a las instalaciones del ITEAM en la Universidad Politécnica de Valencia. Además se tomaron en cuenta varias normas y recomendaciones que tienen que ver con evaluación subjetiva de sonido [9] [10] [11].

Se estudió la influencia de varios factores:

- *Siete distancias* (posición de las fuentes sintetizadas). Estas distancias se comparaban con la posición de referencia, estando tres por delante y tres por detrás de la distancia de referencia: 1,74 – 2,46 – 3,47 – 4,9 – 6,92 – 9,77 – 13,81 metros, siendo 4,9 metros la posición de referencia.
- Cuatro *tipos de sonido* distintos: ruido rosa, voz, sonido de guitarra y de una puerta cerrándose. Según el sonido sea o no conocido por el oyente, según sea un sonido sintético o uno existente en la naturaleza, según el contenido espectral del propio sonido y su estructura temporal la percepción de la distancia se puede ver influida.
- *Reverberación*. Se presentaron los mismos estímulos sonoros con y sin reverberación. Para esta reverberación se sintetizaron las primeras reflexiones basándose en el método de las imágenes [12] y se añadió una pequeña cola al sonido para simular las últimas reflexiones. En total la reverberación añadida era de 20ms, diseñada así para que no hubiese una diferencia abrupta entre los sonidos sin y con reverberación.

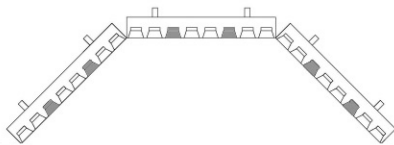
- Dos *ángulos de escucha*, a  $0^\circ$  y  $90^\circ$  de acimut. Según este ángulo la información que llega a cada oído es ligeramente distinta provocando diferencias interaurales de tiempo (ITD) y de nivel (ILD). La elección de los  $90^\circ$  se tomó para tener una representación de la máxima diferencia interaural.



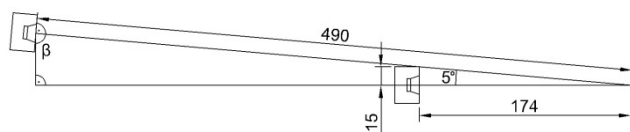
**Figura 1.** Montaje de *arrays* y altavoz de referencia con las distancias sintetizadas (m)

Teniendo en cuenta todas las características descritas de los distintos sonidos a emplear en el test, el *número de estímulos* a presentar a cada sujeto fue de 224. Además se incluyeron 8 referencias ocultas para cada ángulo de escucha, obteniendo finalmente la cifra de *240 estímulos totales* que se presentaron aleatoriamente a cada sujeto. Por razones prácticas el número de estímulos se dividió en dos partes, presentando en una primera parte los estímulos a  $0^\circ$  del plano horizontal con respecto al oyente, y en una segunda parte a  $90^\circ$  del plano horizontal. Así el número de estímulos presentado en cada parte era de  $112 + 8$  referencias ocultas = 120 estímulos para cada ángulo de escucha.

El altavoz de referencia, situado a la distancia de referencia de 4,9m detrás de los *arrays*, se elevó en altura para evitar una posible pérdida de altas frecuencias debido al efecto sombra que el *array* colocado delante pudiera producir en la posición del oyente. Se calculó el ángulo necesario ( $5^\circ$ ) para respetar la distancia de referencia y la inclinación necesaria para que el centro acústico del altavoz estuviese dirigido al oyente (ver figura 3) Este procedimiento utilizado en otros tests de percepción subjetiva sobre distancia sonora [13] [6] es posible gracias a que la capacidad humana para percibir sonidos en altura es muy limitada.



**Figura 2.** Detalle de los *arrays*. En gris los altavoces usados en VBAP

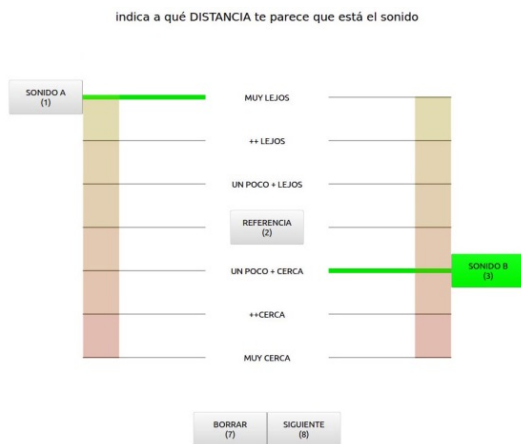


**Figura 3.** Perfil de posición del altavoz de referencia (a la izquierda) con respecto a los *arrays* y el oyente

El sistema fue además calibrado y ecualizado para igualar los niveles y las respuestas en frecuencia del altavoz de referencia con la reproducción de WFS y VBAP por los *arrays*. El nivel de presión sonora del ruido rosa en la posición de referencia era de 69dBA.

25 personas realizaron el test, 15 hombres y 10 mujeres, todos con audición normal y edades comprendidas entre los 24 y 35 años. Para la realización del test se desarrolló una interfaz gráfica que controlaba el propio participante mediante un mando de videojuegos (figura 4). Además éstos se sentaban en una banqueta regulable en altura que permitía ajustar la altura de sus oídos con la de los altavoces. La cortina acústicamente transparente completaba el montaje final, ocultando todos los altavoces frontales para evitar influencias visuales (figura 5).

Todos los participantes realizaron una fase de adiestramiento en la que pudieron escuchar los diferentes tipos de sonido así como las distancias límite, es decir la más lejana y la más cercana. De esta forma los participantes podían tener una referencia del rango concreto en que estarían comprendidos todos los estímulos del test.



**Figura 4.** Interfaz gráfica del test



**Figura 5.** Participante realizando el test

#### 4. RESULTADOS

Para medir la fiabilidad de los resultados se calculó el alfa de Cronbach de los datos obtenidos. La fiabilidad alfa de las respuestas de distancia es de  $\alpha=0,992$  (para  $N=25$  sujetos) lo que indica una muy buena fiabilidad. Además, el análisis de las respuestas sobre las referencias ocultas confirmó un alto grado de consistencia general de todas las respuestas.

- Resultados generales

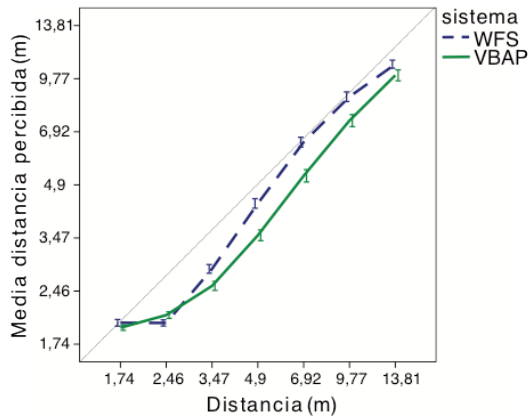
En la figura 6 se ha representado la media de las respuestas de distancia para cada sistema. Como se puede ver ambos sistemas consiguen resultados coherentes. Además se aprecia como los dos tienden a acercar las distancias sintetizadas. WFS parece acercarse más al comportamiento ideal (diagonal de la gráfica).

En la figura 7 podemos ver la divergencia de las respuestas de WFS y VBAP para la posición de referencia (posición 3, a 4,9m) y también para las referencias ocultas (ref). Se observa como las referencias ocultas tienen una divergencia prácticamente inexistente, indicando un comportamiento casi ideal, mientras que WFS y VBAP tienen una mayor divergencia, apreciándose una diferencia entre los dos sistemas, teniendo WFS un comportamiento mejor con diferencia significativa estadísticamente tal como muestran los intervalos de confianza (IC) al 95%.

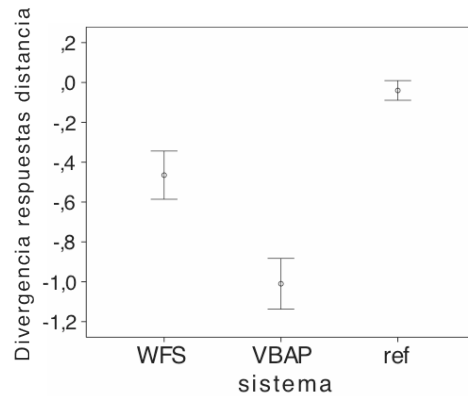
La comparación entre sistemas para todas las posiciones se puede ver en la figura 8, donde se ha representado la divergencia total en las respuestas de distancia de WFS y VBAP. En estas gráficas se puede ver el comportamiento general de un sistema con respecto al otro, observándose una menor divergencia para WFS.

Una prueba-T pareada contrastando los sistemas utilizados (WFS y VBAP) muestra que efectivamente el sistema tiene una influencia altamente significativa sobre las respuestas de distancia ( $t=16$ ,  $gl=2799$ ,  $p<0.001$ ).

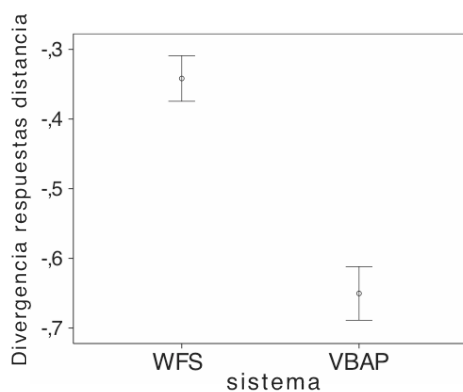
La divergencia en el comportamiento de los sistemas para cada una de las distancias está reflejada en la figura 9, donde se puede ver que aunque los dos sistemas tienen una distribución de la divergencia similar, WFS tiene un mejor comportamiento que VBAP, especialmente para las distancias situadas en torno a la referencia y por detrás de ésta.



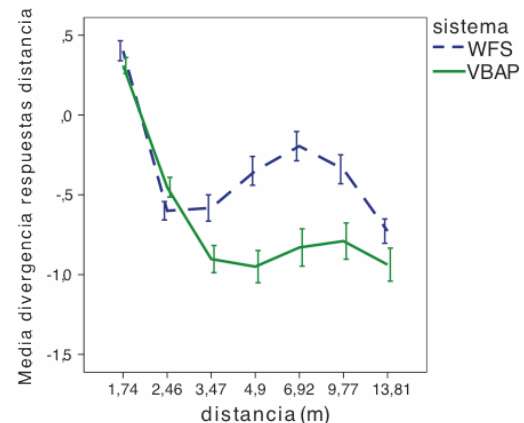
**Figura 6.** Media (N=25) de las respuestas de distancia para WFS y VBAP. IC 95%



**Figura 7.** Divergencia de las resp. de distancia en la posición de referencia (4,9m). IC 95%



**Figura 8.** Divergencia de todas las resp. de distancia para WFS y VBAP. IC 95%



**Figura 9.** Divergencia de respuesta de cada distancia para WFS y VBAP. IC 95%

- Influencia de la reverberación

La reverberación añadida no introduce ningún efecto destacable, a la vista de la figura 10. Para medir este efecto se realizó un análisis de varianza (ANOVA), obteniéndose que la influencia de la reverberación no es significativa ( $p=0,388$ ) en el experimento.

- Influencia del tipo de sonido

La representación de la divergencia de las respuestas de distancia para cada tipo de sonido ofrece resultados llamativos (ver figura 11). En esta gráfica se aprecia que el sonido de guitarra es el que mejores resultados tiene, seguido del sonido de la puerta. La voz sería el siguiente tipo de sonido por acierto de resultados, seguido muy de cerca por el ruido rosa. Destaca la diferencia en la divergencia de distancia del ruido rosa y la voz según el sistema, teniendo claramente mejores resultados WFS.

Con un ANOVA de un factor se comprueba que el tipo de sonido tiene una influencia muy significativa ( $F=131.62$ ,  $gl=3$ ,  $p<0.001$ ) sobre los aciertos de distancia. También se muestra una relación cruzada entre el tipo de sonido y el sistema con significatividad muy alta ( $F=8.6$ ,  $gl=3$ ,  $p<0.001$ ), que corresponde a los mejores resultados de WFS en general y especialmente para ruido rosa y voz.

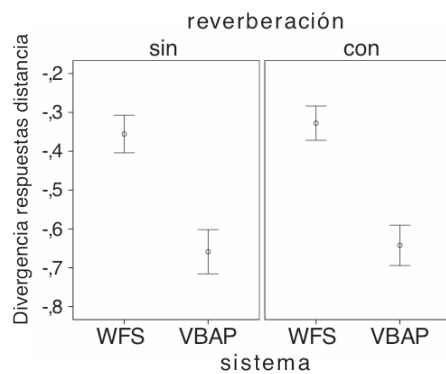
- Influencia del ángulo de escucha

El test se realizó a  $0^\circ$  y a  $90^\circ$  con el mismo número de estímulos en cada ángulo de escucha. En la figura 12 se puede ver cómo el número de aciertos a  $0^\circ$  es superior a los aciertos a  $90^\circ$ . Los resultados del ANOVA de un factor demuestran que el ángulo de escucha tiene una

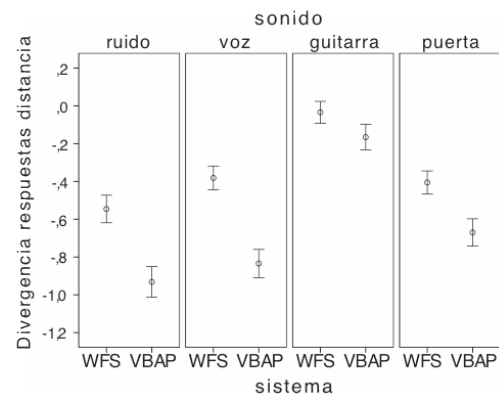
influencia muy significativa ( $F=56.08$ ,  $gl=1$ ,  $p<0.001$ ), pero que su relación cruzada con el sistema no es significativa ( $F=2.55$ ,  $gl=1$ ,  $p=0.110$ ).

- Influencia del sujeto

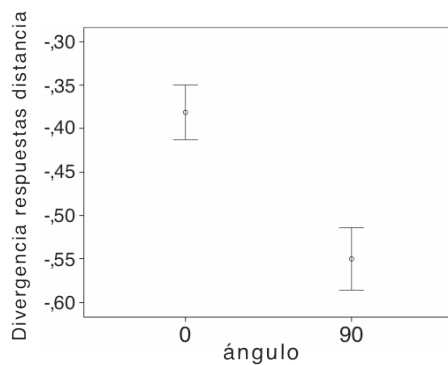
Con un ANOVA de un factor se puede ver una relación cruzada muy significativa entre el sujeto y el sistema ( $F=2.33$ ,  $gl=24$ ,  $p<0.001$ ). Al representar gráficamente la divergencia en las respuestas de distancia para cada sujeto, teniendo en cuenta WFS y VBAP por separado, destaca cómo algunos sujetos consiguen buenos resultados para WFS mientras que obtienen malos resultados para VBAP con mucha diferencia (ver figura 13).



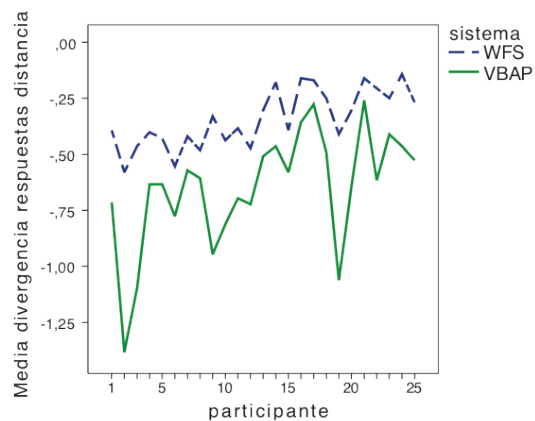
**Figura 10.** Divergencia de respuestas de distancia sin y con reverberación. IC 95%



**Figura 11.** Divergencia de respuestas de distancia según el tipo de sonido. IC 95%



**Figura 12.** Divergencia de respuestas de distancia para escucha a 0° y 90°. IC 95%



**Figura 13.** Divergencia de la media de las resp. de distancia de cada sujeto. IC 95%

## 5. CONCLUSIONES

En base a los resultados obtenidos con el test de percepción subjetiva, podemos enumerar las siguientes conclusiones:

- Ambos sistemas de sonido espacial (WFS y VBAP) se han demostrado capaces de simular una cierta sensación sonora de distancia, que creemos mayor a la simplemente basada en la disminución de nivel de la señal.
- El tipo de sonido es determinante en la percepción de la distancia, obteniendo mejores resultados los sonidos impulsivos. Además WFS es capaz de reproducir mejor que VBAP la distancia sonora para otro tipo de sonidos que no sean impulsivos.
- El ángulo de escucha influye mucho en la percepción de la distancia sonora, pero no es concluyente su relación con los sistemas WFS o VBAP.

- La introducción de reflexiones de primer orden no ha mostrado una mejoría apreciable en la percepción de la distancia. Queda pendiente verificar hasta qué punto la introducción de múltiples órdenes de reflexión mejoraría esta sensación.

- De los resultados de este test se puede concluir que WFS ha mostrado una capacidad global mejor que VBAP para reproducir distancia sonora, al menos para fuentes situadas delante del oyente.

Estos resultados nos llevan a pensar que el nivel de presión sonora es el principal factor responsable de la percepción de distancia. Además los resultados parecen mostrar que la capacidad del sistema de sonido WFS de sintetizar frentes de onda interviene especialmente en la localización de fuentes sonoras en distancia.

## 6. REFERENCIAS

[1] Rumsey, F. "Spatial Quality Evaluation for reproduced Sound: Terminology, Meaning and a Scened-based paradigm" *Journal of the Audio Eng. Soc.*, Vol. 50 (9), pp. 651-666, 2002.

[2] Berkhout, A; Vogel, P; & de Vries, D. "Use of wave field synthesis for natural reinforcement sound". *92th AES convention*. Vienna: AES, 1992.

[3] Pulkki, V. *Spatial sound generation and perception by amplitude panning techniques*. PhD Thesis, Helsinki University of Technology. Helsinki, 2001.

[4] Mershon, D. H; & King, E. "Intensity and reverberation as factors in the auditory perception of egocentric distance". *Perception & Psychophysics*, 18 (6), 409-415, 1975.

[5] Zahorik, P. "Assessing auditory distance perception using virtual acoustics" *Journal of the Acoustical Society of America*, Vol. 111 (4), pp. 1832-1846, 2002.

[6] Wittek, H; Kerber, S; Rumsey, F; & Theile, G. "Spatial perception in WFS rendered sound fields: Distance of real and virtual nearby sources", *AES 116th convention*, Berlin, 2004.

[7] Nogues, M; Corteel E. "Monitoring distance effect with Wave Field Synthesis" *Proc. of the 6<sup>th</sup> Int. Conf. Digital Audio Effects*, London, UK, 2003.

[8] Bech, S; & Zacharov, N. *Perceptual Audio Evaluation-Theory, Method and Application*. John Wiley & Sons Ltd, Sussex (England), 2006.

[9] Rec UIT-R BS.1116-1. Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio incluyendo los sistemas de sonido multicanal. Ginebra: Unión Internacional de Telecomunicaciones, 1997.

[10] Rec UIT-R BS.1534-1. Método para la evaluación subjetiva del nivel de calidad intermedia de los sistemas de codificación. Ginebra: Unión Internacional de Telecomunicaciones, 2003.

[11] EBU Tech 3276-2nd Edition. Listening conditions for the assessment of sound programme material. European Broadcasting Union, Geneva, 1998.

[12] Juan, L; & Yonghong, Y. "Distance perception synthesis in 3D audio rendering using loudspeaker array", *Proc. of the IEEE International Conference on Multimedia Technology (ICMT)*, pp. 290-293, 2011.

[13] Nielsen, S. "Depth perception - finding a design goal for sound reproduction systems". *90th AES convention*. Paris, 1991.