On the distance perception in spatial audio system: a comparison between Wave-Field Synthesis and Panning Systems.

Pablo Gutierrez-Parera, Jose J. Lopez, Emanuel Aguilera

iTEAM Institute, Universitat Politècnica de València, Spain

Abstract

Creating a realistic distance perception by means of spatial audio reproduction systems is not an easy task. Cues such as the ratio between the direct signal and the level of reverberation have been traditionally employed in stereo and surround systems. With the introduction of advanced spatial audio systems such as Wave Field Synthesis (WFS), it is possible to synthesize within the whole listening area the correct wavefront curvature produced by a virtual source located at a given distance. Some previous studies suggest that this curvature can be an additional cue for the listener to extrapolate distance. In this work, 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.

Index Terms: Wave Field Synthesis, Vector Base Amplitude Panning, sound distance, perceptual test.

1. Introduction

Every sound field produces a spatial depth sensation. This feeling is responsible for the perspective perception of an acoustic scene. The same pattern is valid for an artificial acoustic scene. Depth is considered to be an essential attribute for spatial sound perception and the sensation of depth is highly related to the perception of distance to the sound source [1]. The successful creation of depth in a sound scene is a challenge for a spatial audio reproduction system.

For conditions where both listener and sound source are stationary, at least four possible acoustic distance cues have already been suggested to play a relevant role [2-3]:

- *Intensity:* An acoustic point source in free field obeys an inverse-square law, losing 6dB on doubling distance to the listener. It is the most important cue, but implies a previous knowledge of the source power.

- *Direct-to-reverberant energy ratio:* In environments with sound reflecting surfaces, the ratio of energy reaching a listener directly (without contact with reflecting surfaces) to energy reaching the listener after reflecting surface contact (reverberant energy), decreases systematically with increases in source distance.

- *Spectrum:* For distances greater than 15 m the sound absorbing properties of air significantly modify the sound source spectrum, mainly high frequencies. Also, when the sound source is close to the listener's head, some low frequency increase may occur due to the curvature of the sound field.

- *Binaural differences:* Relative differences in acoustic waves when they reach the ears are usually classified as Interaural Level Differences (ILD) and Interaural Time Differences (ITD) forming the Head Related Transfer Function (HRTF). Although the HRTF in far field the does not strictly change with the distance [3], slight and natural movements of the listener modify this information allowing distance perception in the form of acoustic parallax.

The successful synthesis of distance in a sound scene is a challenge for a spatial audio reproduction system.

In the field of music production and cinema, direct-to-reverberant energy ratio as well as distance attenuation have been employed extensively as the main distance cues with acceptable results. These techniques have been commonly applied to stereo and surround (5.1) mixes. However, with the introduction of other advanced spatial audio systems such as WFS, new possibilities have emerged.

With WFS, it is possible to synthesize within the whole listening area the correct curvature of the wavefront arriving to the listener from a source located at a certain distance. Although distance perception has been said to be difficult to perceive with WFS [4], other studies [5] suggest the opposite. This works is intended to provide deeper insight into this point by means of different perceptual listening tests aimed at comparing the performance of WFS and panning systems such as VBAP with respect to their capabilities to recreate realistic distance perception cues.

This paper is structured as follows. Section 2 briefly presents the Wave Field Synthesis principles in order to introduce the reader in the topic. Section 3 focuses the objective of this work by means of a starting hypothesis. In section 4 the description of the test and the experiments procedure are explained in detail. Section 5 analyzes the results from regarding the different signals, positions, angles and other specifications. Finally section 6 summarizes the main conclusion of the paper.

2. Wave field synthesis principles

WFS has been introduced by Berkhout [6] as a concept for sound reproduction without the sweet-spot restrictions inherent in common multichannel systems, as illustrated in Figure 1. In two- (or more-) channel stereo playback the spatial properties of the reproduced field are determined by the characteristics of the loudspeakers. The source localization is correct only in a small area between the loudspeakers, shown by dashed lines in Figure 1(a). In WFS the wave patterns of the sources to be reproduced are correctly synthesized in time and space by an array of closely spaced loudspeakers such that their localization is correct for all listeners in the audience area, as depicted in Figure 1(b).

Based on the Huygens's principle, WFS reproduces an acoustic field inside a volume from the stored signals recorded in a given surface. Huygens's principle states that the wave front radiated by a source behaves like a distribution of sources that are in the wave front, named secondary sources, together creating the next wave front. In WFS the synthetic wave front is created by loudspeaker arrays that substitute the individual secondary sources. The ideal situation would be an area completely surrounded by loudspeakers and fed with signals that create a volumetric velocity proportional to the particle velocity normal component of the original wave front. The application of planar loudspeaker arrays, as prescribed by the Huygens's principle, would involve a very high number of loudspeakers and reproduction channels. However, the recreation of a true natural wave field can only be fulfilled with certain restrictions. Huygens's principle needs to be discretized in practice, which means that an infinite continuous secondary source distribution is replaced by a number of finite arrays of equidistant discrete loudspeakers. Therefore, practical WFS systems employ linear loudspeaker arrays that synthesize the field of 3D sources in the ear plane of the listeners, as depicted in Figure 2. The lack of continuity leads to a maximum usable frequency, known as spatial aliasing frequency, whereas the finiteness of the array causes some truncation effects. For example, a typical loudspeaker distance in the technical literature is 18–20 cm, which gives an aliasing frequency of about 1 kHz. A detailed description of these drawbacks within a listening room can be found in [7-8].

The main advantage of these systems is that the acoustic scene has no sweet spot since it recreates the wavefront of the virtual sources. When listeners move inside the listening area, the spatial sound sensation changes also in a realistic way according to its relative position to the virtual

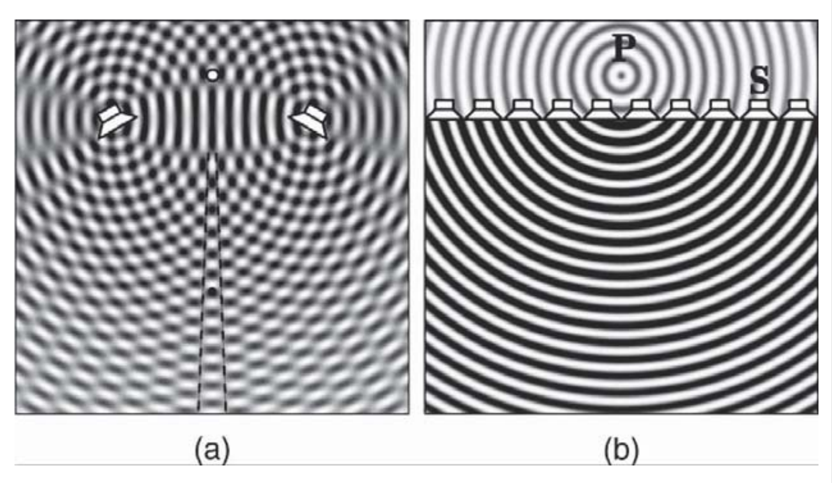


Figure 1. Illustration of basic difference between two-channel stereo and WFS. (a) Two-channel playback. Proper sound localization is shown between two dashed lines where sweet spot is located. (b) Wave field synthesis. P - primary source; S - secondary source.

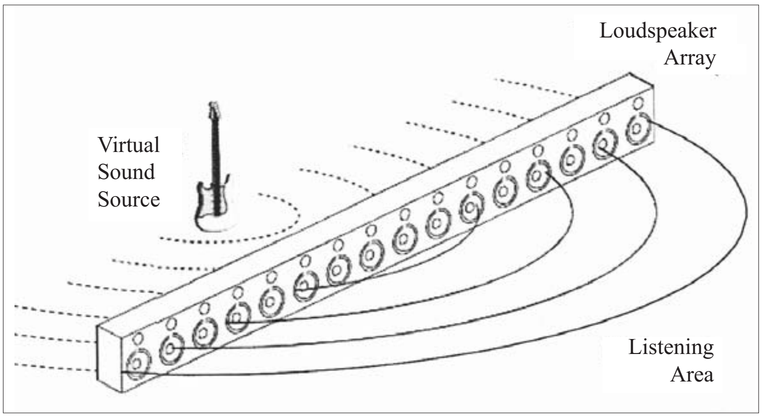


Figure 2. In WFS a linear loudspeaker array synthesizes a virtual sound source in the horizontal plane inside the listening area.

source. In addition to virtual sources behind the loudspeaker array, it is also possible to synthesize sources inside the listening area, the latter known as focused sources. A comprehensive derivation of the underlying mathematics can be found in several studies, such as [6] [9].

3. Hypothesis and objective

As commented before, there are contradictory studies about the possibility of WFS to produce better distance perception cues than other systems. Despite the fact that the curvature of a wavefront has a more relevant effect in the HRTF at short distances, slight movements of the listener might provide some parallax information that they may use to extrapolate distance even when the source is in the far field. In this context, it is considered that the absolute spatial resolution in azimuth is about 5°, but a relative resolution is about just 1° or less, which is enough to notice a difference between two source directions [3].

The objective of this work is to find out if WFS provides better distance perception than other general approaches based on the phantom effect through amplitude panning. VBAP [10] was selected because it provides analytic equations for the multiple loudspeaker case. In the experiments, the listeners were seated, but they were able to move slightly their heads in order to be sensitive to some hypothetical parallax effects. The influence of different factors in the perception was studied: type of sound, listening angle and reverberation at different distances.

4. Test description and procedure

In order to evaluate the perception of sound distance by means of WFS and VBAP, a direct scaling by interval test [11] was selected. The test was based on a comparison between the sounds synthesized by WFS and VBAP for a virtual source at different distances from the listener, having a real sound source placed at a fixed distance.

Figure 3 shows a scheme of the set-up employed in the experiments. An octagonal array of 64 loudspeaker units separated 18 cm apart was deployed around the listener to reproduce WFS and VBAP. Additionally, a reference

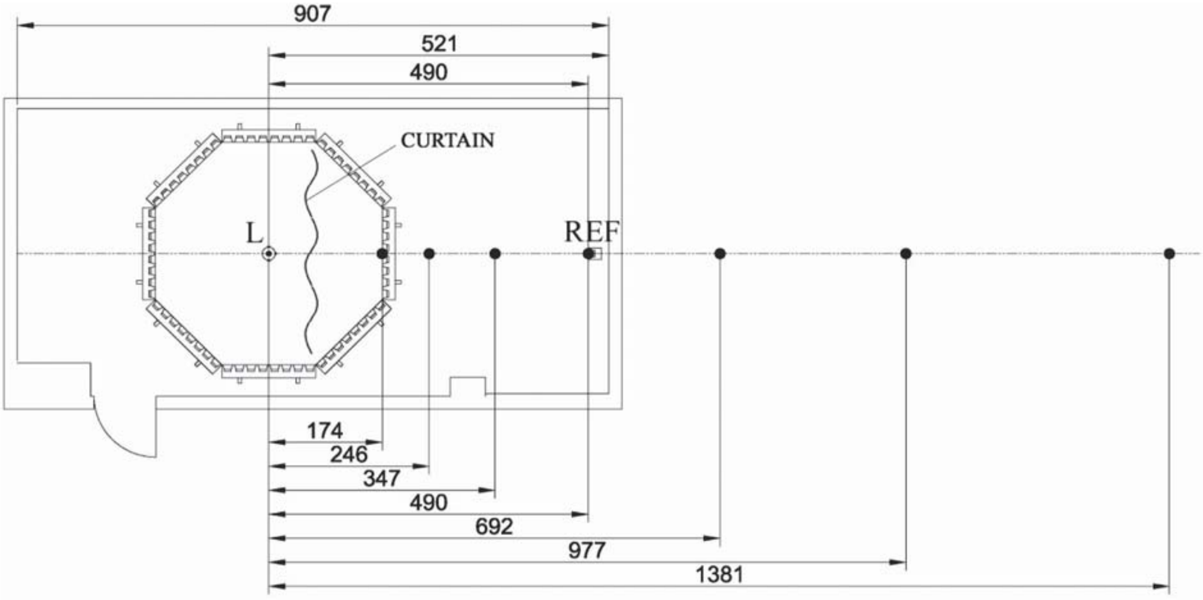
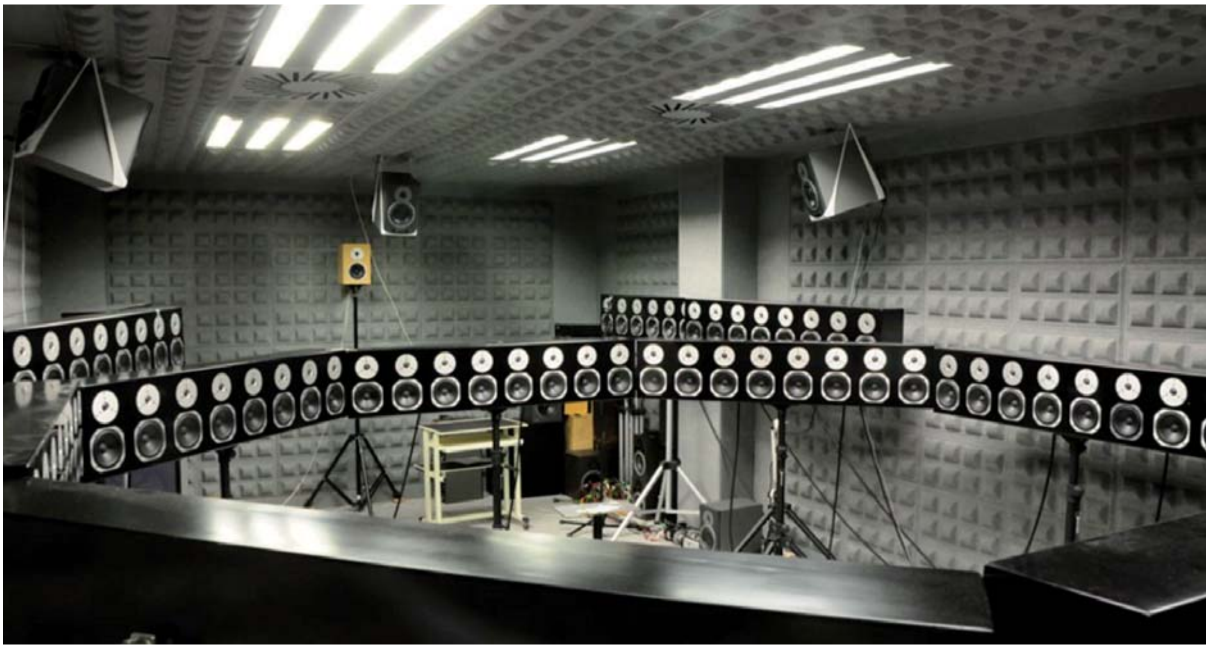
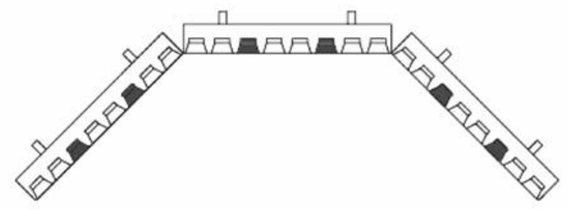


Figure 3. Speaker arrays and reference speaker (REF) set up with synthesized distances (cm).



■ **Figure 4.** Wave-Field Synthesis set-up, where the octagon with the 64 loudspeakers can be appreciated and also the reference loudspeakers at the back in brown color.



■ **Figure 5.** Array detail. The dark speakers are the ones employed for VBAP.

loudspeaker was placed at a distance of 4.9 m. Figure 4 shows a panoramic of the room with the WFS arrays used in the experiments. For testing VBAP, only some loudspeakers of the array were used, marked in grey in Figure 5. To avoid possible influences by visual cues, an acoustically transparent curtain was placed in front of the listener. The test was performed in a dedicated and acoustically treated listening room [12] with a T60 at 1kHz < 0.25s with a volume of 96 m3, (Figure 3). This room was located in the facilities of ITEAM at the Polytechnic University of Valencia.

The standards and recommendations [12-13-14] related to subjective evaluation of sound where fulfilled in the experiments.

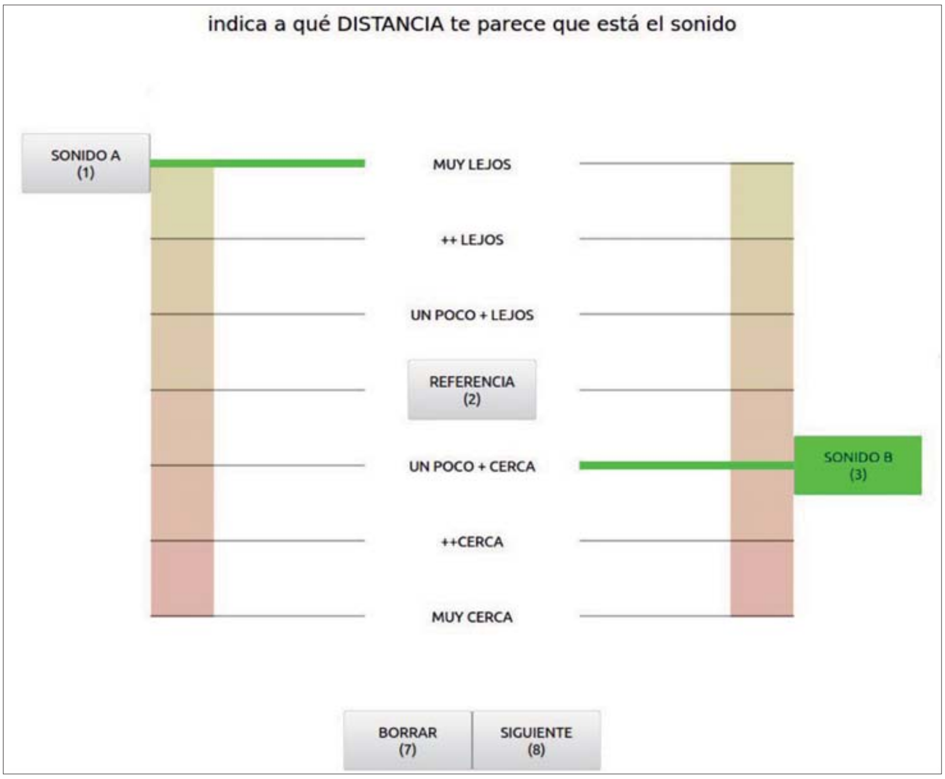
Different effects and factors were taken into account during the test:

1. **Seven distances** (synthesized source positions) were used to compare with the reference position: three ahead and three behind the reference distance at 1.74 – 2.46 – 3.47 – 4.9 – 6.92 – 9.77 – 13.81 meters, with 4.9 meters as reference position, (see Figure 3).
2. Four different **types of sounds** were considered: pink noise, speech, guitar and door closing. These sounds were interesting for localization according to their different spectral and temporal features.
3. Additionally, the effect of synthetic **early echoes** was also analyzed. The same stimuli were presented to the listeners with and without echoes. The added reverberation time was approximately 20 ms, generated by means of four additional first order reflection sources from four virtual walls, calculated by the image-source method [15].
4. Finally, two **listening angles** were studied, 0° and 90° azimuth. Note that 90° azimuth was chosen because it produces maximum interaural time differences.

Combining all the factors to be studied, a total of 224 stimuli per participant were needed. Besides these, 8 hidden references for each listening angle were added, resulting in a total of 240. All these stimuli were randomly presented.

The set-up was calibrated and equalized in order to match the frequency response of the reference speaker with the WFS and VBAP reproduction. The sound pressure level for pink noise at the reference position was 69 dBA.

A total of 25 people participated in the test, 15 male and 10 female all of them with normal audition with ages from 24 to 35. To perform the test, a graphic interface was developed, presented in a computer screen. This interface was easily controlled by the participant by means of a videogame joystick, Figure 6. The participants were seated in a high stool to have their ears at the same elevation than the loudspeakers. The acoustic curtain hides



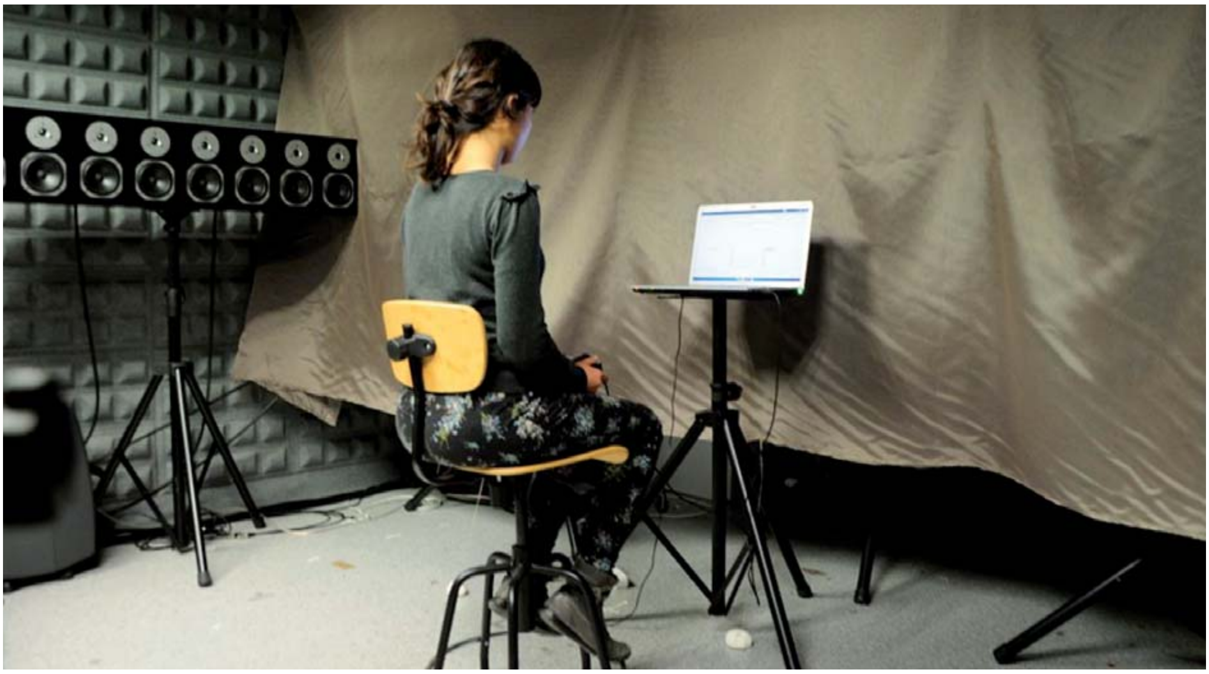
indica a qué DISTANCIA te parece que está el sonido

■ **Figure 6.** Graphic user interface of the perceptual test.

the reference loudspeaker and the frontal array loudspeakers to avoid visual cues or influences, Figure 7.

Before performing the test, some previous information was given to each participant. Moreover, the participants took part in a training phase before the test. They were able to listen to the different types of sound (pink noise, voice, guitar, and door) and also to the distance limits provided by the farthest and closest distances. Since the scale presented to the subjects was completely subjective, this experience allowed them to get an idea of the total range in which the distances would fall.

The test was performed in two phases. First, the subjects evaluated the stimuli for the 0º angle. Then, in the second part they proceeded with the 90º stimuli. The average execution time of each part was 25 minutes with a 15 minutes break between the two parts. The test procedure was



■ **Figure 7.** Participant seated in the listening position ready to start the test. The laptop computer with the GUI, the joystick and the acoustic curtain can be appreciated.

quite simple; the two sounds to assess were presented one after another, followed by the reference sound. For each test signal, the subjects evaluated the perceived distance over the subjective scale with respect to the reference sound, been possible to listen both several times.

5. Analisys of results

5.1. Reliability

A Cronbach's alpha was calculated with all the participants' answers. The alpha reliability of the distance answers is α=0,992 (for N=25 participants) which indicates a very high reliability. Besides, the analysis of the responses about the included hidden references confirmed a high degree of consistency in the responses.

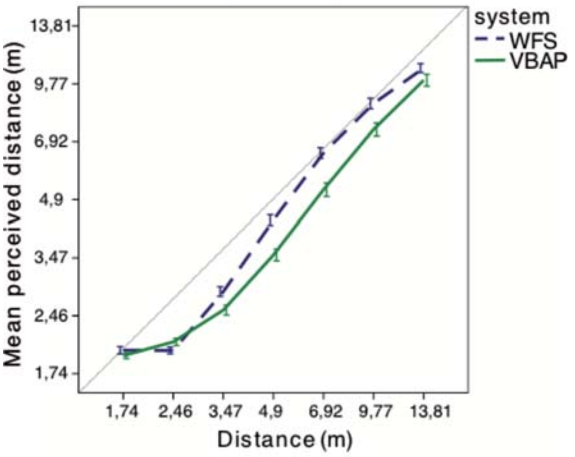
5.2. Aggregated results

The mean of all the answered distances for all the stimuli and participants and classified for each system is showed in Figure 8. As illustrated by this figure, both systems have coherent results. The graph also shows that both systems tend to perceive closer the synthesized sources. WFS seems to approximate better the ideal behavior (diagonal).

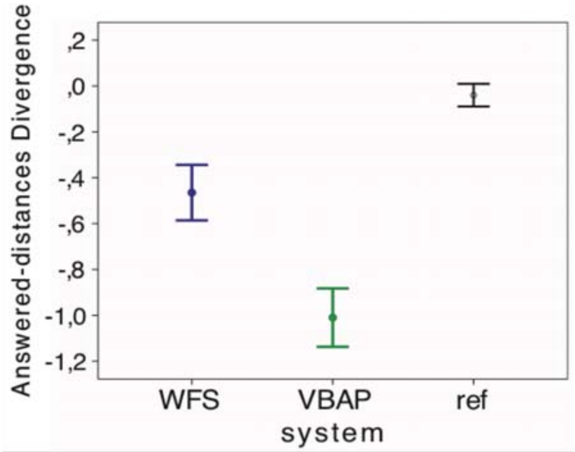
Figure 9 illustrates the divergence (perceived error) of WFS and VBAP for the reference position (4.9 m) and also for the hidden references. It is observed that the hidden reference has a very small deviation, indicating a nearly ideal behavior. In contrast, WFS and VBAP show a greater divergence. However WFS presents a better error average than VBAP, having a statistically significant difference, as shown by the 95% confidence intervals (C.I.).

An overall comparison for all positions is showed in Figure 10, where the total divergences of the aggregated answered distances for WFS and VBAP are represented. In these graphs, the general behavior of each system can be compared, indicating that WFS provides lower divergence.

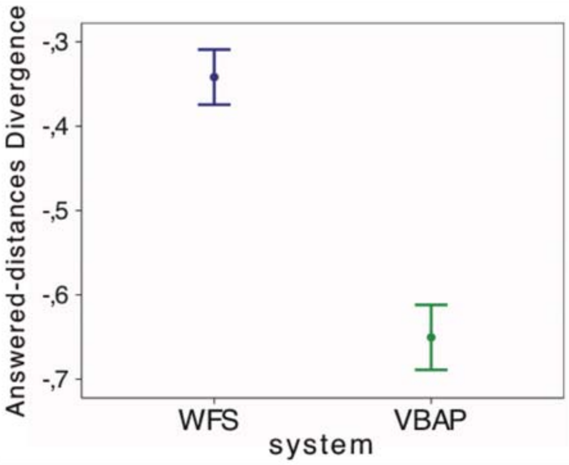
To study the influence of the system in the test, a paired samples t-test was performed. It showed that the system (WFS vs VBAP) has a highly significant influence in the answered distances (t=16, gl=2799, p<0.001).



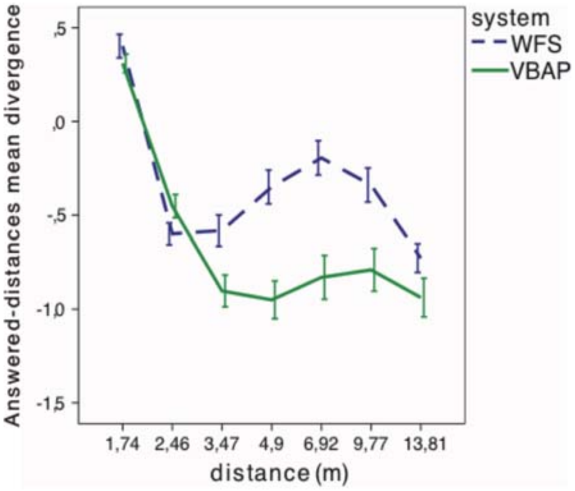
■ **Figure 8.** Mean (N=25) of the answered distances for each system (WFS and VBAP). 95% confidence intervals (CI).



■ **Figure 9.** Divergence of the answered distances at reference position (4.9m) for WFS, VBAP and hidden references. 95% CI.



■ **Figure 10.** Divergence of all answered distances for WFS and VBAP. 95% CI.



■ **Figure 11.** Divergence for each of the answered distances for WFS and VBAP. 95% CI.

Finally, the divergence for each distance is showed in Figure 11, where a similar distribution is observed. Note that WFS provides better performance than VBAP, especially for distances located around the reference and behind it.

5.3. Influence of the early echoes

The added four early reflections from 4 virtual walls introduce no noticeable effect, as shown in Figure 12. To measure this effect an analysis of variance (ANOVA) was performed to yield that the influence of the added reverberation was not significant (p=0,388).

5.4. Influence of the type of sound

Figure 13 shows that the guitar sound provides the best results, followed by the closing door sound. The voice sound is the next obtaining more correct assessments, followed closely by pink noise. It is worth to note the differences in distance divergence for pink noise and voice according to the system, where WFS outperforms VBAP. A one-way ANOVA was performed, founding that the type of sound has a significant influence (F=131.62, gl=3, p<0.001). There is also a cross relation between the type of sound and the system, with a very high significance (F=8.6, gl=3, p<0.001), corresponding to better results of WFS in general, and especially for pink noise and voice.

A direct scaling by interval test was performed to compare between the sounds synthesized by WFS and VBAP. ANOVAs and other statistical analyses were then carried out.

5.5. Influence of the listening angle

The test was performed at 0° and 90° by rotating the listener, with the same number of stimuli at each listening angle. Figure 14 shows that better results are obtained for frontal listening. A one-way ANOVA shows that the listening angle has a very high significance (F=56.08, gl=1, p<0.001), but its cross relation with the system is not significant (F=2.55, gl=1, p=0.110).

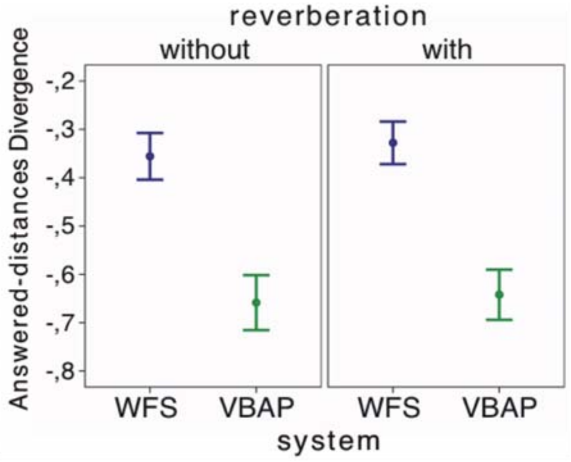


Figure 12. Divergence of the answered distances for WFS and VBAP, with and without reverberation. 95% CI.

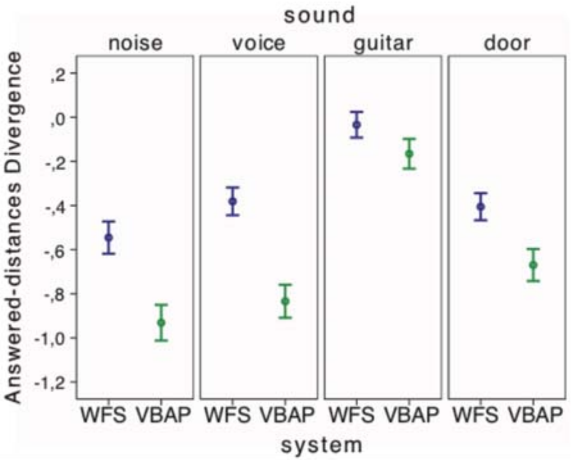


Figure 13. Divergence of the answered distances for WFS and VBAP, according to the type of sound. 95% CI.

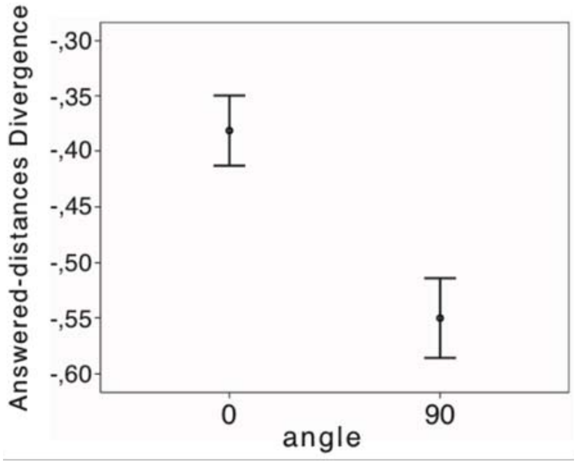


Figure 14. Divergence of the answered distances for the listening angles at 0° and 90°. 95% CI.

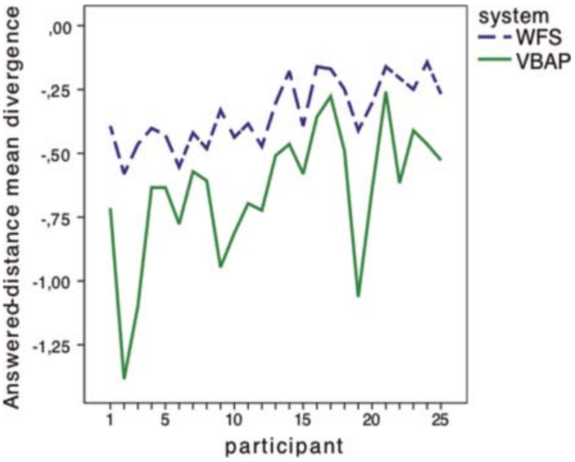


Figure 15. Divergence of the mean answered distance for each participant, considering the system, WFS or VBAP.

WFS has been shown to have a better overall capability to reproduce sound distance than VBAP.

5.6. Influence of the participant

A very high significant cross-relation between the participant in the test (the listener) and the system can be seen with a one-way ANOVA (F=2.33, gl=24, p<0.001). In a graphic representation of the divergence of the answered distances for each participant, considering separately WFS and VBAP, the difference between both systems stands out. Some of the participants (2, 9, 19) have good results with WFS and poor results with VBAP, as observed in Figure 15.

6. Conclusions

According to the subjective perception tests carried out in this work, some conclusions can be listed:

- Both spatial sound systems (WFS and VBAP) have been shown capable of simulating a certain sound sense of distance based on the attenuation caused by distance.

- The type of sound is a determining factor in the perception of distance, getting better results for impulsive sounds. Moreover, WFS is capable of reproducing better sound distance than VBAP for other sounds that are not impulsive.

- The listening angle has a high influence in sound distance perception, but its relation with the system (WFS or VBAP) is not determinant.

- The first order reflections did not provide a substantial improvement in distance perception. More experiments are needed to evaluate to what extent the introduction of multiple order reflections enhance this feeling.

- From the test results, it can be concluded that WFS has been shown to have a better overall capability to reproduce sound distance than VBAP, at least for sources placed in front of the listener.

Despite the pressure level is the main factor responsible for distance perception; our tests have concluded that, at least for this set-up, WFS is better at producing distance perception cues than VBAP, as confirmed by the statistical analysis of the results.

References

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

[3] Blauert, J. Spatial Hearing: The psychophysics of human sound localization, MIT Press, Cambridge, 1996.

[4] 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.

[5] Nogues, M; Corteel E. “Monitoring distance effect with Wave Field Synthesis” Proc. of the 6th Int. Conf. Digital Audio Effects, London, UK, 2003.

[6] A. J. Berkhout, “A Holographic Approach to Acoustic Control”, Journal of the Audio Engineering. Society (Engineering Reports), Vol. 36, pp. 977–995, 1988.

[7] Spors S, Buchner H, Rabenstein R, Herbordt W. “Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering”, Journal of the Acoustical Society of America, Vol. 122(1), pp. 354–69, 2007.

[8] Gauthier P-A, Berry A. “Objective evaluation of room effects on wave field synthesis”, Acta Acustica united with Acustica, Vol. 93(5), pp. 824–36, 2007.

[9] Berkout, A. J; de Vries, D; Vogel, P. “Acoustic Control by Wave Field Synthesis” Journal of the Acoustical Society of America, Vol.93, pp. 2764-2778, 1993.

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

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

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

[13] Rec ITU-R BS.1116-1. Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, 1997

[14] Rec ITU-R BS.1284. General methods for the subjective assessment of sound quality, 2003

[15] 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.

Biographies

Pablo Gutiérrez-Parera was born in Córdoba, Spain in 1982. He received a telecommunications engineer degree in 2008 from the Universidad Politécnica de Madrid. In 2010 and 2013 he obtained a M.S. degree in digital postproduction and a European graduate in telecommunication systems, sound and image engineering, both from the Universitat Politècnica de Valencia. Currently, he is a PhD grant holder from the Spanish Ministry of Economy and Competitiveness under the FPI program and is pursuing his PhD degree in telecommunications at the Institute of Telecommunications and Multimedia Applications (iTEAM) working in the field of spatial audio.

José Javier López was born in Valencia, Spain, in 1969. He received the telecommunications engineer degree and the Ph.D. degree, both from the Universitat Politècnica de València, Valencia, Spain, in 1992 and 1999, respectively. Since 1993, he has been involved in education and research at the Communications Department, Universitat Politècnica de València, where he is currently a Full Professor. His research activity is centered on digital audio processing in the areas of spatial audio, wave field synthesis, physical modeling of acoustic spaces, efficient filtering structures for loudspeaker correction, sound source separation, and development of multimedia software in real time. He has published more than 160 papers in international technical journals and at renowned conferences in the fields of audio and acoustics and has led more than 25 research projects. Dr. Lopez was workshop co-chair at the 118th Convention of the Audio Engineering Society in Barcelona and has been serving on the committee of the AES Spanish Section for nine years, currently as secretary of the Section. He is a full ASA member, AES member and IEEE senior member.

Emanuel Aguilera received a telecommunications engineering degree in 2004 and a M.S. degree in Artificial Intelligence, Pattern Recognition and Digital Image in 2011, both from the Universitat Politècnica de València, Spain. He is a researcher and senior programmer at the Institute of Telecommunications and Multimedia Applications (iTEAM), where he has been working since 2006 on the area of digital signal processing for audio, multimedia, virtual reality and mobile devices applications. He is interested in wave-field synthesis, image processing, real-time multimedia processing for telecommunications and audio applications for mobile platforms.