WAFE-FIELD SYNTHESIS: STATE OF THE ART AND FUTURE APPLICATIONS

Jose J. Lopez, Maximo Cobos, Basilio Pueo

Universidad Politécnica de Valencia,
Institute of Telecommunications and Multimedia Applications
Camino de Vera, Edif 8G, 46022 Valencia, Spain
jjlopez@dcom.upv.es

ABSTRACT

Wave-Field Synthesis (WFS) has become one of the most promising spatial sound reproduction systems. The most basic difference of WFS in comparison to other available systems is that the acoustic field is accurately synthesized using loudspeaker arrays in a broad area, suppressing the sweet spot that characterizes conventional surround systems and giving an accurate and deep immersion for all the listeners. In this paper, a review of the main concepts related to WFS, from its fundamentals to the latest applications and developments is presented. The limitations and drawbacks of WFS are listed and succinctly described, giving also the main proposal to overcome these limitations. Among them, a solution to include elevation in WFS is presented and some recent techniques to perform stereo to WFS up-mixing are also commented. Finally, applications of WFS to different engineering areas such as immersive videoconference and auditory display systems are presented.

1. INTRODUCTION

From a physical point of view, we can state that humans can hear because tiny auditory hair cells in the inner ear detect vibrations due to sound and convert them to nerve signals. However, we also hear because, throughout evolution, the sense of hearing has helped our survival. As with many other mammals, the sense of hearing has played a major role in hunting and avoiding being hunted. Our sense of hearing enables us to identify dangers or targets in the environment; first by identifying their position in space and later by classifying them (finding out the type of animal or thing that generated the sound).

The accuracy achieved by humans in these two tasks cannot be compared to any artificial development, as it is very difficult to emulate these capacities by means of computational methods. The auditory centers of the brain are responsible for interpreting the different sound signals that arrive at our two ears. These centers learn and are trained until reaching maturity. For example, babies are not able to localize sounds until they are five months old. Once these capacities are consolidated in the brain, the subject makes use of them without being aware of it. When an animal or a human being detects a hazard by a strange or uncommon sound, the brain automatically discharges in the bloodstream a load of adrenaline that warns the subject of this emergency situation. These involuntary actions make up the survival functions of the human auditory system.

The objective of three-dimensional spatial sound systems is to accurately recreate the acoustic sensations that a listener would perceive inside a particular room or in an environment with certain acoustic properties. This concept implies a series of physical and technological difficulties that are a current research issue in sound engineering.

Stereo sound systems, considered as the simplest approximation to spatial sound, have been utilized throughout the last 50 years as an added value in sound recordings, particularly for music material. Used in theaters since the mid-1970s, surround sound systems have evolved and entered homes to give a better sensation than stereo by using more reproduction channels (5.1, 6.1, 7.1, 10.2). Surround mixes are mainly intended to enhance the experience in video projections by adding artificial effects in the rear loudspeakers (explosions, reverberation, or ambient sound). The optimal listening position, known as the sweet spot, is almost limited to the central point in the loudspeaker set-up and the spatial sensation degrades considerably outside the central zone.

Another much more realistic strategy is to reproduce directly in the ears of the listener, via headphones, the signal that he would perceive in the acoustic environment to be simulated. The perception of the simulated scene depends on the fidelity of the reproduction. This strategy is widely known as binaural reproduction. The signals to be reproduced with headphones can be recorded with an acoustic head or artificially synthesized by using a measured head related transfer-function (HRTF). The future of HRTF-based techniques is promising, since a significant amount of music material is listened to over headphones using mobile devices. There are still some issues to be solved regarding the HRTF variability among different subjects and active research lines are centered on this aspect of binaural reproduction. In addition, the incompatibility in the reproduction of dummy head signals over loudspeakers is another classical problem due to the introduction of crosstalk.

On the other hand, the most promising spatial sound system currently is called wave-field synthesis (WFS). The most basic difference of this system in comparison to surround sound systems is that the acoustic field is accurately synthesized using loudspeaker arrays in a broad area, suppressing the sweet spot that characterizes conventional surround systems. Research in WFS has been very active in Europe in the last decade and several research groups are pioneers of this emerging sound system.

In this paper we are going to review the main concepts related with WFS, from its fundamentals to the latest applications and developments. In the next section an overview of the basic principles and mathematical foundations of WFS is presented. In section 3, the main limitations and drawbacks of WFS are listed and succinctly described. The following sections are devoted to new state-of-the-art developments in WFS that, in most cases, try to overcome some of the limitations commented in section 3. In sections 8 and 9 applications of WFS to immersive videoconference and to auditory display systems are presented. Finally, the conclusions of this paper are summarized in section 10.

2. WAVE-FIELD SYNTHESIS FUNDAMENTALS

2.1. Origin of WFS

The origins of WFS come back from 1953 when Snow published an overview of stereophonic techniques and discussed the acoustic curtain as the ideal stereophonic reproduction technique. It was aimed at transporting the acoustic of the recording venue to a reproduction room using microphone and loudspeaker arrays. Due to technical constraints at that time, it was not feasible to put his ideas into practice. As a compromise, they applied three-channel stereophony, accepting that the original aim of recreating the real sound field would no longer be fulfilled.

The intuitive acoustic curtain concept was replaced later by a well founded wave theory. In the late 80s, the Wave Field Synthesis (WFS) concept was introduced by the Technical University of Delft. The origin of this theory was published in “Applied Seismic Wave theory” [1] and “A holographic approach to acoustic control” [2]. The term “acoustical holography” was used, not yet called WFS, and the system was designed to be the ultimate tool for acoustical control systems in theaters. These publications introduced the physical basis of WFS by applying algorithms known from seismics to the field of acoustics. The basic work on WFS was continued by Berkhout in [3] and [4]. Since then, a number of publications have appeared to complement and improve this basic theory.

2.2. Mathematical background

The theory of WFS is related to Huygens’ principle, formulated in 1678. This principle states that each element of a wave front propagating through a particular medium may be seen as the center of an individual spherical wave. Consequently, the wave front generated by a primary sound source can be seen as a series of elementary, secondary sources. It is not very practical to position the acoustic sources on the wavefronts for synthesis. By placing the loudspeakers on an arbitrary fixed curve and by weighting and delaying the driving signals, an acoustic wavefront can be synthesized with a loudspeaker array. Figure 1 illustrates this principle.

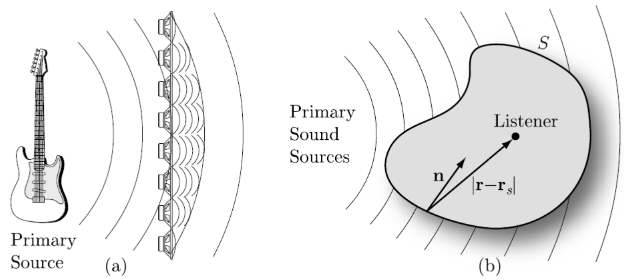

Figure 1. (a) Basic principle of WFS. (b) Parameters used for the Kirchhoff-Helmholtz integral.

According to this principle, an arbitrary acoustical wave field can be recreated within a source-free volume V by secondary sound sources distributed on a closed boundary surface S in the so-called Kirchhoff-Helmholtz integral:

$$P(\mathbf{r},\omega)=\oint_S \frac{1}{4\pi}\left[P(\mathbf{r}_s,\omega)\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|}\right)-\frac{\partial P(\mathbf{r}_s,\omega)}{\partial \mathbf{n}}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|}\right)\right]dS \tag{1}$$

where $P(\mathbf{r},\omega)$ is the Fourier transform of the sound pressure $p(\mathbf{r},t)$, k is the wave number, $\mathbf{r}$ is the coordinate vector of an observation point and $\mathbf{r}_s$ is the coordinate vector of the integrand functions on the surface S.

In practice, the Kirchhoff-Helmholtz integral states that at any listening point within the source-free volume V, the sound pressure $P(\mathbf{r},\omega)$ can be calculated if both the sound pressure and its gradient are known on the surface enclosing the volume. This can be used to synthesize a wave field within the surface S by setting the appropriate pressure distribution $P(\mathbf{r}_s,\omega)$. This fact is used for WFS-based sound reproduction that, in his discretized form, can be expressed as:

$$P(\mathbf{r},\omega)=\frac{jk\rho c}{2\pi}\sum_n u_n(\mathbf{r}_n,\omega)\,\frac{e^{-j|\mathbf{r}-\mathbf{r}_n|}}{|\mathbf{r}-\mathbf{r}_n|}\Delta x\Delta y \tag{2}$$

Equation (2) expresses the field produced by a plane of loudspeakers separated Δx and Δy in each axis. In [4] is demonstrated than instead of a plane, a line of loudspeakers can be employed to reproduce the acoustic field in a horizontal area. The principle is maintained except for a small error in the attenuation of pressure with the distance that usually remains unnoticed by the listener in practice.

The last interested equation in WFS theory is the driving signal function that provides the driving signal for each loudspeaker in order to synthesize a virtual source. It can be approximated by using stationary-phase representation [5]. Physically, this approximation means that the wavefront is synthesized by all the loudspeakers of the array, but a dominant contribution is given by the loudspeaker positioned at the point of stationary phase. According to the geometry of Figure 2 and after substantial mathematical manipulation, the driving signal $Q(\mathbf{r}_n,\omega)$ of the n^{th} loudspeaker can be found:

$$Q(\mathbf{r}_n, \omega) = S(\omega)\frac{\cos\theta_n}{G(\theta_n, \omega)}\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{|z - z_1|}{|z - z_0|}}\,\frac{e^{-jk|\mathbf{r}_n - \mathbf{r}_m|}}{\sqrt{|\mathbf{r}_n - \mathbf{r}_m|}} \tag{3}$$

Now, let's analyze the 5 terms that compose equation (3). The first one is the excitation signal; the nominator of the second one reflects the gain to be applied to the *n*-th loudspeaker regarding the angle with the virtual source position and the denominator is the directivity of the loudspeaker. The second term is a filter that emphasizes high frequencies at +3dB/oct. The third term is usually fixed for a centered listener position and then becomes a constant. The numerator of the fifth term is a pure delay related to the distance of the virtual source to the loudspeaker and the denominator is a gain term also related to this distance.

As can be concluded from equation (3) and the above explanation, WFS does not imply a great computer power but mainly a huge quantity of hardware resources.

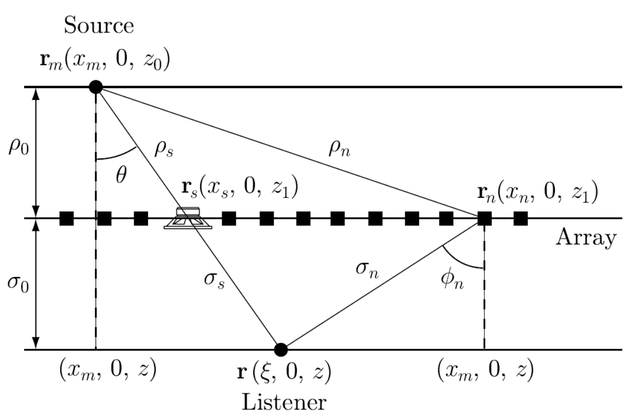

Figure 2. Configuration for WFS . Loudspeaker array at $z = z_1$ synthesizes wavefield of a source at $\mathbf{r}_m$ in the receiver plane at $z > z_1$.

2.3. WFS features

Definitely, the main advantage of WFS systems is that the acoustic scene has no sweet spot. When listeners move inside the listening area, the sound pressure level changes also in a realistic way according to their relative position to the virtual source, Fig. 3.

In practice, it is not necessary to surround completely the listener by a surface in three dimensions; it is enough to consider a linear loudspeaker array located in front of the listener. In Figure 3, a typical WFS configuration is presented, where a virtual sound source is synthesized in the location of the listener by using a loudspeaker array. However, unlike stereo systems, the synthesized field is not only valid for one listener, but also for all the listeners in the room.

Figure 4 shows a simulation of the rendered acoustic field produced by a virtual source consisting of a pure tone situated at the middle point. In the figure, it is possible to compare the resulting acoustic field of pressure employing a stereo system

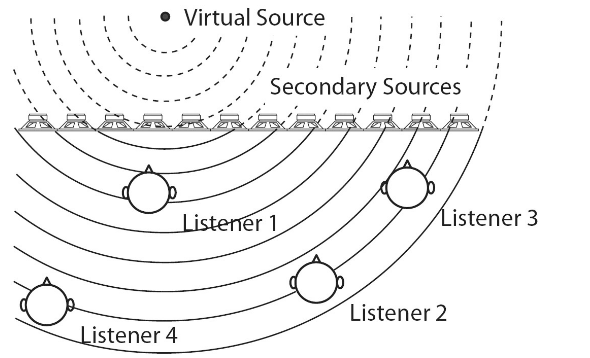

Figure 3. Several listeners in an extended area perceive a virtual source with spatial fidelity.

and a WFS array. The synthesized field in the stereo case in not as perfect as in the WFS case and it is only valid in a trapezoidal area in the middle of the loudspeakers. Additionally as frequency increases, this area becomes narrower.

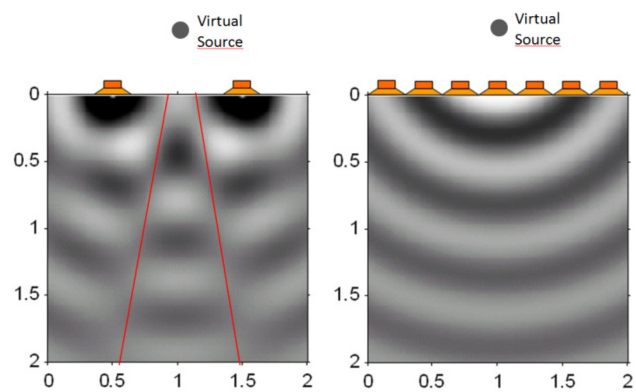

Figure 4. Simulation of the sound field rendered by a stereo system and a WFS system for a pure tone.

Another advantage of the WFS systems is that it allows the listeners not only to perceive with high precision the angle of a source (i.e. the direction of arrival), but also it allows them to notice the distance to the points where the sources are located. This effect is better detectable when the listener moves around the room and the angle of the source changes according to its distance to the center of the room. As a result, WFS is considered a 2.5D spatial sound system, despite the fact that it does not provide elevation cues (it provides distance cues).

Finally, as opposed to conventional surround sound systems (5.1, 6.1, …) that are channel-oriented, WFS provides a clearer advantage because it is a source-oriented system. It means that a production developed for WFS can be reproduced in any WFS system, regardless its size, topology or number of channels. This fact implies a considerable change in the philosophy of audio production, since it is necessary to change from channel-oriented mixing to object-oriented production. During the adaptation, it is possible to reproduce surround material in WFS set-ups, obtaining better results than in the case of using classical surround.

3. LIMITATIONS OF WFS

In the previous sections, we presented a series of properties that make WFS the most powerful spatial sound reproduction system. However, creating a copy of a sound field is not completely possible due to some practical constraints. The main intrinsic physical constraints of WFS are:

- The discretization of an ideal continuous secondary source distribution to a loudspeaker array leads to spatial aliasing, resulting in both spatial and spectral errors in the synthesized sound field at high frequencies.
- The finiteness of the array leads to truncation effects, resulting in diffraction waves that cause after-echoes and pre-echoes.
- The restriction to a line loudspeaker array in the horizontal plane instead of a planar array leads to amplitude errors.

Additionally, WFS presents some more difficulties and drawbacks when trying to put this theory to work in practice. Some of these are:

- A huge number of loudspeakers are needed, besides the associated hardware for audio amplification and digital signal processing. This leads very to expensive systems.
- Restriction to the horizontal plane as a consequence of using linear arrays instead of 2D arrays does not allow elevation effects, not being a true 3D audio system.
- Reflections in the reproduction room distort the field created by the WFS array.
- The vast majority of musical recordings are stored and supplied in a two channel “stereo” format, whereas WFS needs separated sources to synthesize acoustic scenes.

Regarding the three main intrinsic physical constraints, in the last years, none of them has supposed a major obstacle in the development of the technique. Spatial aliasing has been studied mathematically in different works in order to optimize the sound field and to reduce spatial artifacts, especially in some parts of the listening area. However, subjective studies have found that the effect of spatial aliasing is very difficult to be noticed by average listeners. The advantages introduced by the increased spatial impression hide and chip away the effects of aliasing, which are unnoticeable for the majority of the listeners.

The truncation effects, although they cannot be completely eliminated, they can be significantly reduced by windowing the loudspeakers closer to the end of the array and avoiding square angles in array shapes different from straight lines. Additionally, the amplitude errors due to the use of point sources instead of cylindrical ones are quite small. For example, as stated in [6], if we calculate the system for a listener situated 2 meters away from the array, the amplitude error for a listener situated at 4 meters is less tan 1dB. This value is below the detection threshold of the listeners, making the error unnoticeable.

4. ELEVATION IN WFS

WFS has been shown to provide excellent localization accuracy in the horizontal plane. However, it is restricted to azimuth only localization. This is a clear disadvantage of WFS in comparison to other spatial sound systems that provide elevation, as Ambisonics or the 10.2 cinema surround sound system. For these reasons, different solutions have already been proposed to overcome this problem. For example, putting a linear array on the ceiling or using two parallel linear arrays located at different elevation angles. Unfortunately, the phantom effect does not work in elevation as good as in azimuth so these systems do not always provide the desired quality [7]. It was also reported that subjects did not perceive a well-defined phantom image between the two loudspeakers but either heard the sound coming from one of the two loudspeakers or perceived an unclear image.

More sophisticated and accurate localization cues are those provided by Head-Related Transfer Functions (HRTF) [8], which describe how a given sound is filtered by diffraction and reflection properties of the head, pinna, and torso, before the sound reaches the eardrum and inner ear. Pre-filtering effects are very dependent on the direction of arrival of the incoming sound. These effects play a very important role in source localization, particularly in the determination of source elevation. [9]. The idea of using HRTF cues for providing an enhanced reproduction of sound including elevated sources is nothing new. In the mid 70s, Jens Blauert patented a binaural synthesis method that made use of HRTFs [10]. Moreover, he continued working on the idea, studying the influence that some specific frequency ranges had on the perception of elevation [11].

The authors proposed in [12] the use of HRTF spectral elevation cues in conjunction with WFS with the aim of producing the sensation of elevated virtual sources. Azimuth localization is achieved with the usual WFS system, but elevation is simulated by means of a filtering stage prior to WFS rendering, as seen in Figure 5. Different elevation filterbanks have been computed from several HRTF databases. Also, a normalization step for removing cues which are not directly related to elevation has been performed in order to obtain pure elevation cues.

Several listening tests were carried out using a panel of 12 subjects in order to validate the system. The experiments were carried out using a 24 loudspeaker WFS array where sources with different elevation were presented to the jury. An acoustically transparent curtain was used to hide the loudspeakers and avoid the ventriloquism effect. The virtual sound sources used in the experiments were rendered in WFS as point sources. The test signals consisted of a set of three broad-band pink noise bursts of 1 second duration with silence intervals of 0.5 seconds in between. The loudness of the test signal was adjusted to be 65 dB(A) at the listening position when elevation was 0. The experiments below were repeated for each of the considered HRTF elevation responses (IRCAM, CIPIC, RSS and PEAK). For more information about these 4 responses see [12]. The desired HRTF filtering is performed directly in the sources before applying the WFS algorithm.

In the first experiment, the discrimination capability between upwards and downwards moving sources was studied. For each of the databases considered, four stimuli were generated. They consisted of the reference test signal (pink noise) moving in the following directions:

1. Upwards from the horizontal plane to 40º.
2. Downwards from 40º to the horizontal plane.
3. Upwards from the horizontal plane to -40º.
4. Downwards from -40º to the horizontal plane.

Figure 6 shows the results obtained for this experiment, where the bars indicate the grade of success obtained by the listeners in identifying correctly the direction of movement of the stimulus 1 to 4 of the above list. It can be observed that all the systems performed significantly well. For movements above the horizontal plane, the results are always above 95% for all HRTF banks. In the case of movements below the horizontal plane, the results are slightly worse, but also provided about 90%. These results are quite encouraging since at least, it can be distinguished whether a source is moving up or down in the proposed system, which would allow such systems to use special effects in movies or video games.

In the second experiment, the capability of identifying which sound is at a higher position from two successive sounds filtered with different θ responses was studied. To accomplish this, the jury must have the same sound excitation, but filtered for two different elevations. Then, the subject has to answer which of the two seems to be higher, the first or the second one. The sounds were again two pink noise bursts corresponding to different values of angle θ. In the experiments all the pair combinations of available elevations in the range of ±45 or ±40 degrees were used, but differentiating two tests, one for positive and another for negative elevations. Combinations were chosen randomly for each subject using a software test that automates the process.

Figure 7 shows different matrices representing the hit rate obtained for each pair of stimuli in the different elevation systems. Rows indicate the θ position of the first sound and columns indicate the θ position of the second sound. White cells indicate that all the subjects successfully perceived which sound was at upper and which was at lower positions. Note that smaller rates are achieved when the relative distance of the two sounds is small, as the filters used for simulating elevation are more similar in these cases.

Limiting the test at elevations no higher than 45 degrees, the graphs show quite encouraging results, although there is confusion in some stimuli for static sources. Although not directly studied, it seems reasonable to assume that, following the experiments of sources in motion, a subject could readily identify the elevation of a static source if it has previously had an ascending or descending movement. Therefore, better results are expected if the sources move in elevation before becoming static.

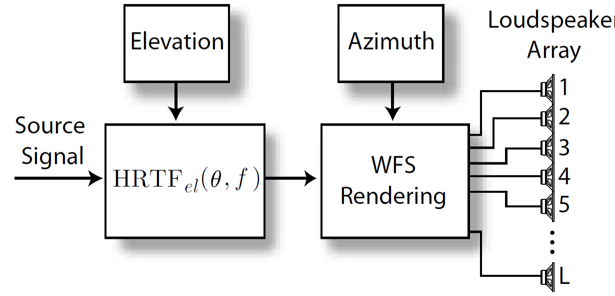

Figure 5: Block diagram of the proposed HRTF-WFS hybrid system for add elevation in WFS systems.

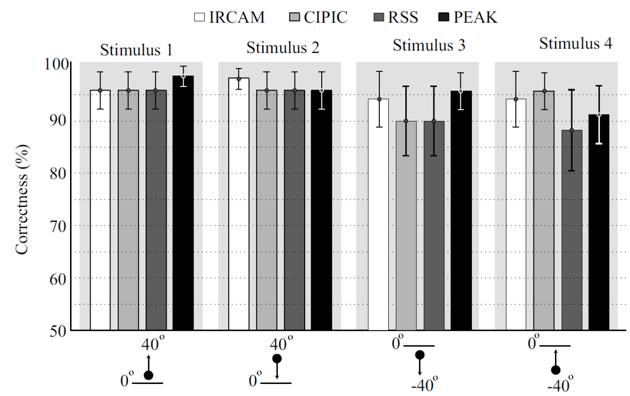

Figure 6: Results for the Upwards/Downwards Discrimination Experiment. Hit rates for the different elevation systems (95% confidence intervals).

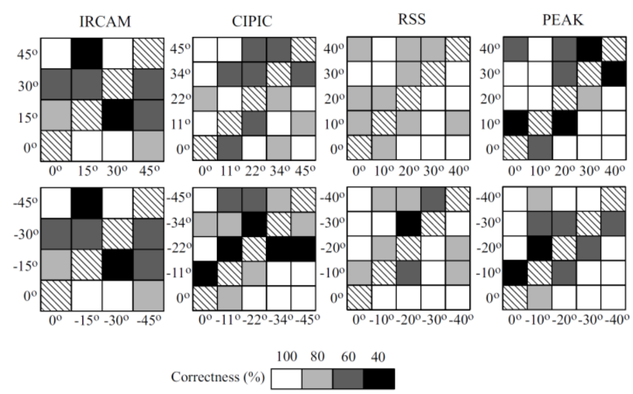

Figure 7: Hit rate for the Higher / Lower discrimination experiment. For each matrix, rows indicate θ position of the first sound and columns indicate the θ position of the second sound.

5. RESYNTHESIS OF SOUND SCENES

Despite all of the advances made in spatial sound reproduction over the last few years, the vast majority of musical recordings are stored and supplied in a two channel "stereo" format, making it necessary to listen to them on a two-loudspeaker reproduction system. In this context, audio signal processing systems for converting stereo recordings into four or five channels are gaining attention. These up-mixers are used for reproducing conventional stereo recordings with spatial reproduction systems, taking advantage of the spatial properties of multichannel audio reproduction. The goals of these kinds of processors relate to modifying the stereo listening experience to create a source image with a spatial quality that is similar to the original mix, with natural-sounding ambiance and a better evaluation by listeners [13]. Stereo to-5.1 up-mixers are usually based on a matrix scheme, which generates the additional channels by simply adding and subtracting the input channels with altered gain and phase.

As WFS systems are not yet widely deployed, up-mixing processors fully designed for converting stereo recordings into synthesized scenes have rarely been discussed in the literature. The main objective of stereo-to-WFS up-mixers would be the same as those developed for five-channel up-mixing: to enhance the spatial quality of conventional stereo recordings. However, the spatial properties of WFS, which are ideally suited to be combined with virtual and augmented reality systems and other applications, open a new door to go further than the conventional home-theater-oriented up-mixing. From this point of view, more sophisticated up-mixing schemes based on blind audio source separation (BASS) algorithms are considered. Algorithms for source separation have been shown to be very useful in many fields, ranging from biomedical applications to music information retrieval. Applications to audio and speech are widespread, although many different approaches are taken, depending on the problem under consideration. In this context several BASS algorithms for extracting musical instrument and speech sources from stereo recordings have been developed over the last few years with acceptable results, although not fully satisfying.

In [14] a work with this objective was presented by the authors. The purpose of this paper is to evaluate the subjective quality and spatial attributes of synthesized acoustic scenes in WFS when the virtual sources are generated using separated tracks from stereo mixtures.

A diagram of this up-mixing system is depicted in Fig. 8. The left and right channels are the input signals of a BASS algorithm, which extracts a set of separated tracks corresponding to estimations of the original sources that were added in the stereo mixdown. These tracks feed the WFS rendering algorithm, which drives the excitation signals corresponding to each unit of the loudspeaker array.

Using this up-mixing system, 4 different separation algorithms (DUET, MuLeTS, ADRess and PIW) were used to up-mix different stereo materials to WFS. A series of listener tests were conducted in order to evaluate the quality of different spatial sound attributes after the up-mixing using these separation algorithms and comparing also with the original separated sources as with the stereo mix. For that purpose, a setup as shown in Figure 9 was employed. The spatial attributes evaluated were source locatedness, source widthness, sound quality, localization accuracy and ensemble aperture.

In Fig. 10 the mean values obtained for all of the stimuli and listeners are represented jointly in order to show the overall performance of each system/separation method considered. Source locatedness and widthness are represented in a 5-grade scale and the sound quality conserves the MUSHRA scale (0 to 100). The localization accuracy shows the mean run standard deviation in degrees. The ensemble aperture is represented as the deviation from the expected aperture value (in degrees as well). This figure is intended to summarize all the results in a single graph. Therefore, although the area enclosed by each polygon can be loosely related to the perceived audio quality, these results should not be interpreted in a strict way. The DUET and MuLeTS algorithms achieved the best results, both being very similar. The PIW and ADRess methods had acceptable results, but not as good as the other two separation algorithms. More details and interpretation of the results can be found in [14].

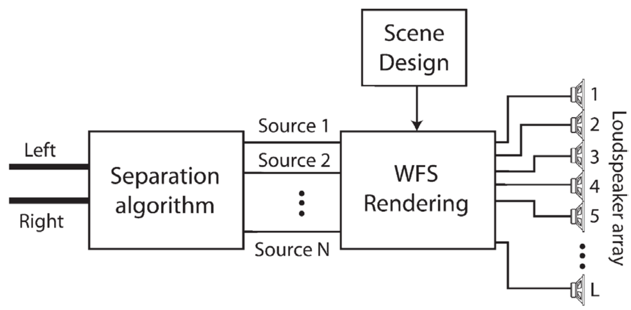

Figure 8: Stereo to WFS up-mixing scheme.

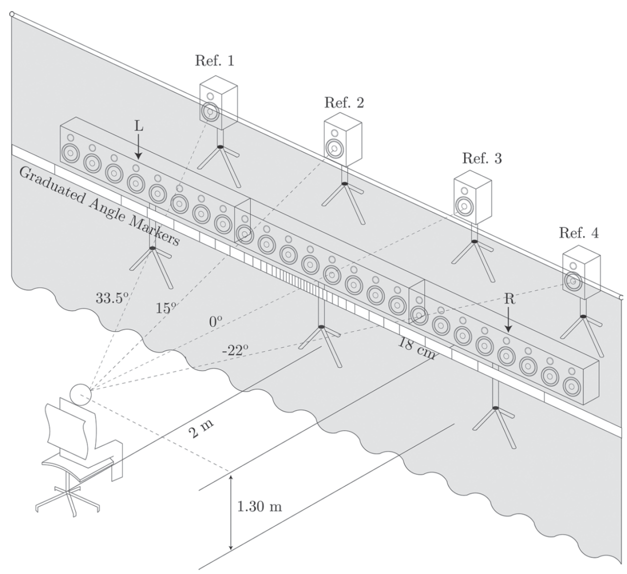

Figure 9: Experimental setup for subjective evaluation of spatial sound attributes.

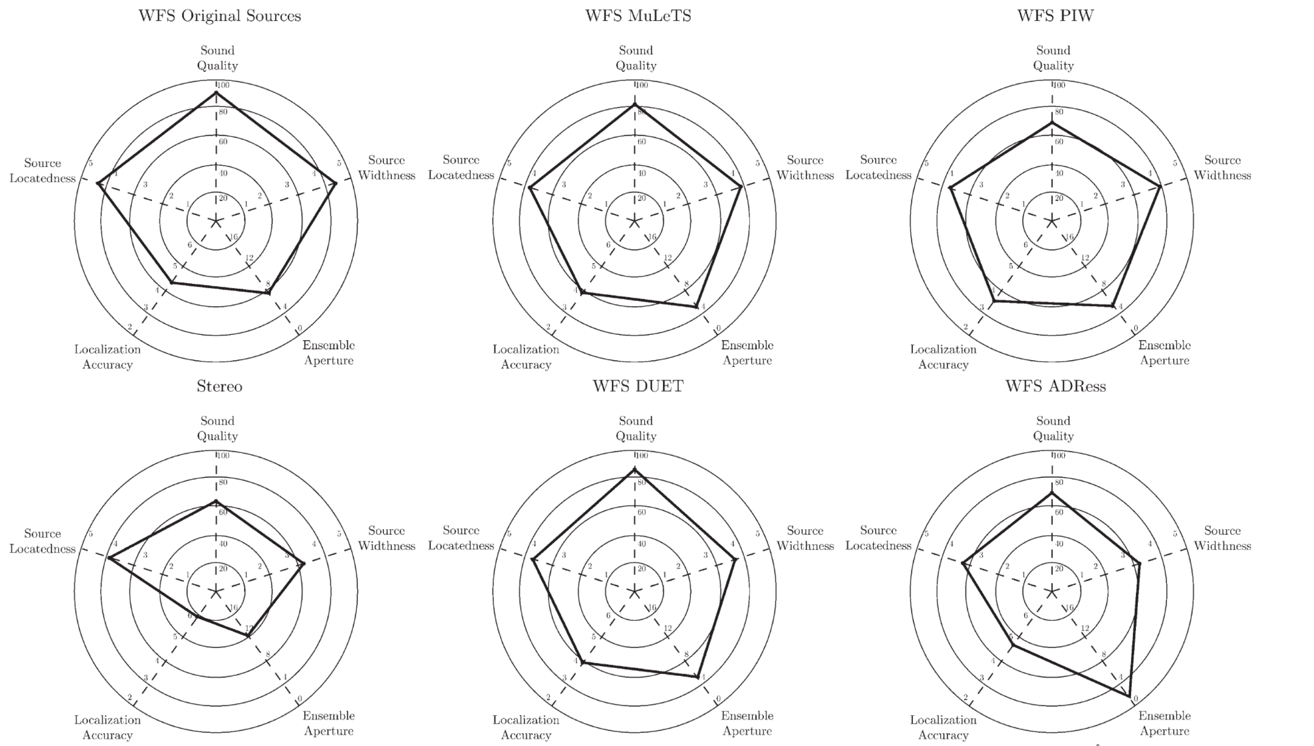

Figure 10: Overall system performance.

6. DISTRIBUTED MODE LOUDSPEAKERS

One on the advances in the hardware related to WFS is the introduction of a kind of flat loudspeakers that give some benefits for certain applications. In the next subsection, the fundamentals of these loudspeakers and the application to projection over them will be explored.

6.1. Fundamentals of DML

The DML essentially consists of a thin, stiff panel that vibrates in a complex pattern over its entire surface by means of a electro-mechanic transducer called exciter. The exciter is normally a moving coil device, which is carefully positioned and designed to excite the natural resonant modal structure of the panel optimally. In Fig. 11, a graphical representation of a DML is presented, which shows panel, exciter and housing.

DMLs are panels of finite extent deploying bending waves. The DML relies on the optimization of its eigenmodes to produce a modal density that is sufficiently high to give the impression of a continuous spectrum [15]. The excitation of bending waves on panels results in sound radiation with distinct qualities with regard to the pistonic motion of typical dynamic loudspeakers. A traditional loudspeaker acts for the most part of its radiation as a phase coherent radiator, and thus, it has a correlated output. However, the uncorrelated output of a DML produces an omnidirectional directivity response over the major part of the audio frequency band [16]. In addition to this, DML sources produce reflections that are less correlated to the direct sound than those radiated from piston sources and thus, constructive and destructive interference of sound is minimized.

One of the practical advantages of DMLs is their ease to mount directly on the wall surface. Besides, they are light-weight loudspeakers with a small back housing that can be get unnoticed as part of the decoration. Since the panel surface can be large and the vibration is low enough to be imperceptible to the human eye, they can be integrated into a room interior and simultaneously used as projection screens [7]. In this way, image and sound are fully integrated for multimedia applications. Furthermore, the cost of DMLs is generally lower than that of dynamic loudspeakers on baffles. These features make DMLs very suitable for WFS reproduction.

Encouraged by the positive results on sound localization, the applicability of single-exciter DMLs for WFS reproduction was tested for the first time in [17], reporting that individual panels reconstructed the wave field correctly. However, the secondary sources spacing required by the WFS algorithm to acquire a reasonable useful bandwidth, forced the size of panels to be very low. This conferred DMLs weak bass response due to the lack of excited modes in the low frequency region.

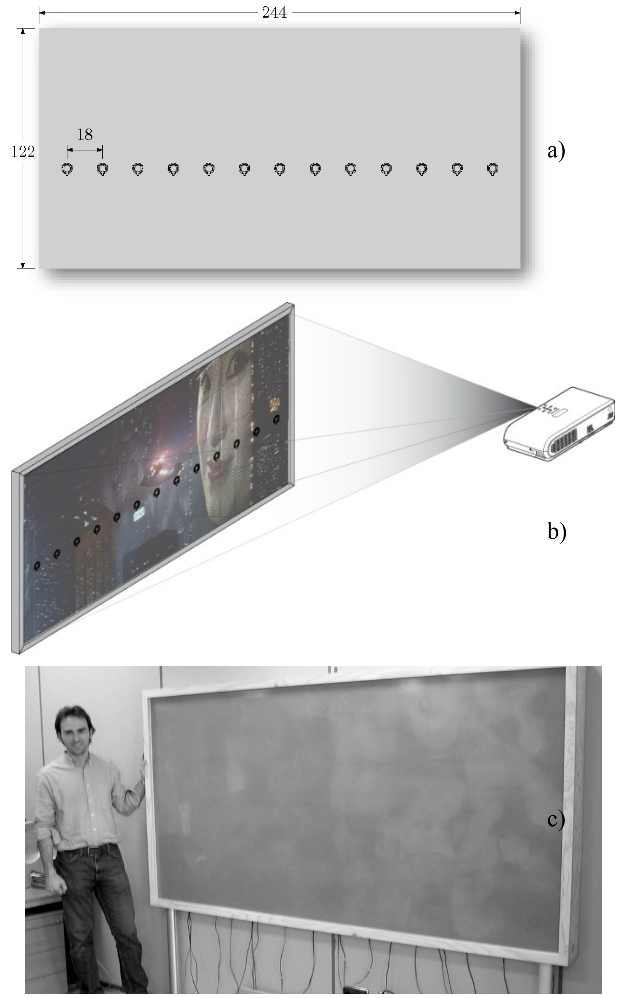

Figure 11: Large MAP, a) block diagram and measures, b) employment in conjunction with a projector, c) photograph of the resulting prototype panel assembled and ready to use.

In [18], Boone proposed to extend the DML technology to a panel with multiple exciters, each driven with a different signal. Such a configuration would act as a WFS array if every exciter on the panel would excite only a small part of the panel around the exciter position, which was experimentally confirmed in [19]. Since exciters in a DML operate by converting electrical signals into mechanical movement which is applied to the panel, these panels are also known in the technical literature as Multiactuator Panels (MAP). There are some benefits for MAPs to be used in WFS reproduction. They can easily be integrated into a living room because of its low visual profile. Furthermore, the vibration of the surface is almost negligible so that it can be used as projection screens.

6.2. Large MAPs for WFS and Projection

One of the first developments of large MAPs suitable for projection were presented by the authors in [20]. The projection on these panels can be even in stereo (left eye/right eye) in order to produce 3D images that match with 3D sound produced by the WFS.

The well-known 3D displays that require the viewer to wear special glasses present two different images in the same display plane. The glasses select which of the two images is visible to each of the viewer's eyes. Technologies for this include polarization, shuttering or anaglyph. In the prototype presented in [20] we selected the shuttering technology were a double frame-rate is employed (left and right eye emitted alternatively) in combination with a shutter glasses that block the opposite image. The projector employed was an InFocus DepthQ working at 120 Hz with DLP technology.

For the projection screen, a large MAP was especially designed and built, fig. 11, to meet the demands of immersive audio applications. For that purpose, it includes a horizontal line of exciters composed of 13 exciters with 18 cm spacing, presenting an aliasing frequency of approximately 1 kHz.

The panel is a sandwich of polyester film bonded to an impregnated paper honeycomb 5 mm thick using a thermoplastic adhesive (cell size = 4.8 mm). Its bending rigidity is 4.23 and 2.63 Nm in the x and y directions respectively and has an areal density of 0.51 kg/m2. Due to its size, frequencies until 100 Hz can be reproduced successfully. More about the acoustic performance and audio quality of this panel was analyzed and previously presented by the authors in [20].

7. APPLICATIONS OF WFS IN IMMERSIVE VIDEOCONFERENCE

Videoconference systems have been around the market for long time. Their more ambitious aim has always been to avoid the need for having physical presence of people for carrying out meetings. However, their impact in the market has not been as important as many people expected. The reason for that is, essentially, that the sense of realism was far away from the expected. In order to improve this, the main efforts actually are focused on making the feeling of being there as real as possible. The goal is to achieve a video screen that appears to be a virtual window to the other side of the conference.

The concept of virtual windows should include image and sound. In both cases, the human perception system is able to obtain a 3D sense of the space using two eyes and two ears. That is why an acoustic window should not only provide a realistic image perception of the space, but spatial sound as well.

Two main technologies are usually employed to produce stereoscopic images: the systems where the user must wear special glasses (polarized, shuttered or anaglyph), and the autostereoscopic displays that provide 3D perception without the need for special glasses or other headgear, [21]. In the case of sound, Wave-Field is a good method to complement stereo vision in order to increase the immersion of the participants in the conference.

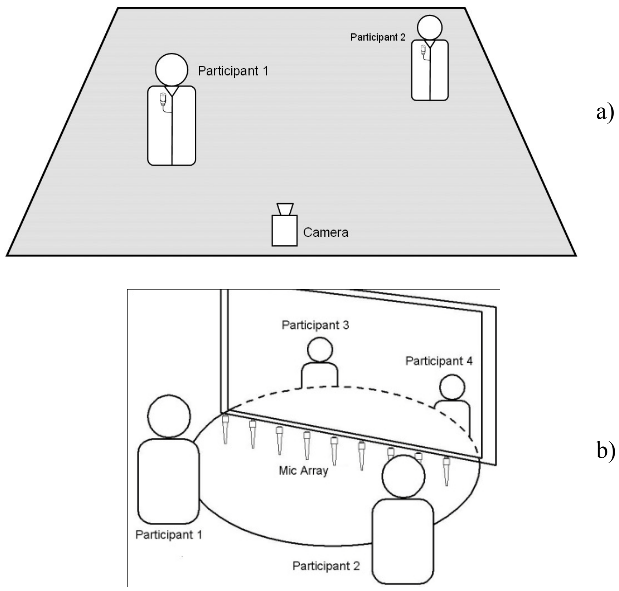

Figure 12:. Sound capture system, a) using spot-mics and the support of a videocamera and b) using a microphone array and beamforming.

Figure 13: Prototype combining two WFS arrays of cone loudspeaker with a 42” autostereoscopic screen.

Figure 14: Prototype combining a V-shaped WFS array with two autostereoscopic screens.

Some experiences have been carried out by the authors combining the WFS reproduction with screens (stereoscopic and non stereoscopic) in order to obtain highly immersive acoustic windows for videoconferencing applications.

As a previous step to sound reproduction, the sound capture stage should be carried out. Figure 12 shows two typical set-ups for sound capturing. In the first one, 12a), each participant wears a spot-mic that captures close sound. The position on each participant, needed in the reproduction stage, is obtained from image analysis using the same video-camera used for image transmission. In the second one, 12b), an array of microphones captures and analyzes the sound field, obtaining direction of arrivals and the location of the sources. This information, audio for each participant and position, is sent to the other side of the videoconference for spatial rendering using WFS.

Figure 13 shows a 42” autostereoscopic screen with two WFS arrays, one above and another below the screen in order to provide not only accurate source localization in the horizontal plane, but also to provide elevation effects on the screen and discriminate the vertical source position on the screen

Figure 14 shows a V-shaped WFS array that covers a more extended area that includes two screens. This set-up allows a wider scenario that the previous set-up. In this case no elevation effects are produced, but the more stretched aspect ratio improves immersion.

8. APPLICATIONS TO THE AUDITORY DISPLAY

In many auditory display applications spatial sound plays an important role. In this kind of applications, not only the nature of the sound to be displayed is important, but also its position in the space around the listener [22].

Binaural sound systems have been often employed for this purpose successfully. When the sound has to be situated only in front of the listener (±90º) and no elevation is needed, the system performs almost perfectly. However, when it is necessary to differentiate auditory events coming from the front and back of the listener or to situate objects above or below the horizontal plane, binaural systems based on HRTF do not perform as good as in the frontal source case. This is because the well known effect of HRTF individualization. Each person has his own HRTF and using a non-personalized one, generally produces back to front confusions and considerable errors in the perception of source elevation.

Surround sound systems (as the ones employed in the cinema industry) can cope with the need to place sources behind the listener. However, the placement obtained is not very precise, especially in 5.1 systems, where the spatial resolution at the side and the back is quite disappointing. Other derivate systems, such as 7.1, improve this resolution but at the expense of having and managing more channels. Another drawback of surround systems is that, although the sweet spot is better than in stereo systems, it is not as large as it would be desirable for some applications.

WFS can be an alternative to surround systems when very precise localization is needed or the auditory display is employed in a multi-user experience. In these cases, a large sweet-spot covering the entire listening room is mandatory. Otherwise, the excessive cost and infrastructure of a WFS set-up does not justify its employment.

9. CONCLUSIONS

In this paper, a review of the fundamentals of WFS and the state of the art of the technology has been presented. Despite the great economic investment that represents a WFS installation, this technology is progressively being established for special applications where a real acoustic spatial immersion is needed. The advantage that WFS brings suppressing the sweet spot that characterizes conventional surround systems opens spatial sound systems to new applications different from cinema, giving an accurate and deep immersion for all the listeners in the show room. Applications where WFS has been successfully integrated include high immersion videoconference, movie theaters, virtual reality set-ups, spatial music reproduction, etc.

Moreover, in this paper, limitations and drawbacks of WFS have been listed and succinctly described, commenting also the main approaches that have been proposed to overcome these limitations. Among them, a solution to include elevation in WFS has been presented, describing additionally some recent techniques to up-mix from stereo to WFS. Next, applications of WFS to different engineering areas such as immersive videoconferencing were discussed. Finally, some comments on the possible application of WFS to auditory displays were given, reasoning its advantages and comparing to classical surround systems.

10. ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Science and Innovation under project TEC2009-14414-C03-01, by the University of Alicante under project GRE09-33 and by the Generalitat Valenciana under project GV/2011/042.

11. REFERENCES

[1] A. Berkhout, Applied Seismic Wave Theory. Amsterdam: Elsevier Science, 1987.

[2] A. Berkhout, “A holographic approach to acoustic control,” J. Audio Eng. Soc., vol. 36, no. 12, pp. 977–995, Dec. 1988.

[3] A. Berkhout, D. de Vries, and P. Vogel, “Wave front synthesis: a new direction in electro-acoustics,” in 93th Conv. Audio Eng. Soc., no.3379, San Francisco, USA, Mar. 1992.

[4] A. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by Wave Field Synthesis,” J. Acoust. Soc.America, vol. 93, no. 5, pp. 2764–2778, May 1993.

[5] N. Bleistein, Mathematical Methods for Wave Phenomena. Academic, 1984.

[6] D. de Vries, “Wave Field Synthesis”, AES Monograph, New York, USA: Audio Engineering Society, 2009

[7] W. de Brujin, M. Boone: Application of wave field synthesis in life-size videoconferencing. Proceedings of the 114th AES Convention.

[8] C. I. Cheng, G. H. Wakefield: Introduction to head-related transfer functions (HRTF’s): Representations of HRTF’s in time, frequency, and space (invited paper). Proceedings of the 107th Convention of the Audio Eng. Soc., New York, USA, 1999.

[9] P. M. Hofman, J. G. A. Van-Riswick, J. Van-Opstal: Relearning sound localization with new ears. Nature Neuroscience 1 (1998).

[10] J. Blauert, G. Boerger, L. Peter: Method and arrangement for controlling acoustical output of earphones in response to rotation of listener’s head. United States Patent 3962543, 1976.

[11] J. Blauert: Spatial hearing. MIT Press, Cambridge, MA, 1997.

[12] J.J. López, M. Cobos, B. Pueo, Elevation in Wave-Field Synthesis using HRTF cues, Acta Acustica, 96 (2), pp 340-350, 2010

[13] J. Usher, “Design Criteria for High Quality Upmixers,” in Proc. AES 28th Int. Conf. (Pitea, Sweden, 2006 June).

[14] M. Cobos, J.J. López, Resynthesis of Sound Scenes in Wave-Field Synthesis from Stereo Mixtures using Sound Source Separation Algorithms, Journal of the Audio Engineering Society, 57 (3), pp 91-110, 2009

[15] J. W. Panzer and N. Harris, “Distributed-mode loudspeaker radiation simulation,” in 105th Conv. Audio Eng. Soc., no. 4783, San Francisco, USA, Sept. 1998.

[16] A. Angus, “Distributed mode loudspeaker polar patterns,” in 107th Conv. Audio Eng. Soc., no. 5065, New York, USA, Sept. 1999.

[17] M. Boone and W. de Brujin, “On the applicability of distributed mode loudspeaker panels for Wave Field Synthesis based sound reproduction,” in 108th Conv. Audio Eng. Soc., Paris, France, Feb. 2000.

[18] M. Boone, “Multi-actuator panels (MAPs) as loudspeaker arrays for Wave Field Synthesis,” J. Audio Eng. Soc., vol. 52, no. 7–8, pp. 712–723, July 2004.

[19] M. Kuster, D. de Vries, D. Beer, and S. Brik, “Structural and acoustic analysis of multiactuator panels,” J. Audio Eng. Soc., vol. 54, no. 11, pp. 1065–1076, Nov. 2006.

[20] J.J. Lopez, B. Pueo, M. Cobos, Conventional and Distributed Mode Loudspeaker Arrays for the Application of Wave-Field Synthesis to Videoconference, IWAENC 2008, Seattle, WA (USA), 2008

[21] N.A. Dodgson, “Autostereoscopic 3D Displays”, Computer, vol 38, no. 8, pp. 31 – 36, Aug. 2008

[22] D. Brock, J.A. Ballas, B. McClimens, Perceptual issues for the use of 3D auditory displays in operational environments, Proceedings of ISICT’03, Dublin, 2003