REAR AND SIDE REPRODUCTION OF ELEVATED SOURCES IN WAVE-FIELD SYNTHESIS

Jose J. Lopez, Maximo Cobos and Basilio Pueo

Institute of Telecommunications and Multimedia Applications (iTEAM), Universidad Politécnica de Valencia
Camino de Vera s/n, 46022, Valencia, Spain
phone: + (34) 616251395, fax: + (34) 963879583 , email: jjlopez@dcom.upv.es, mcobos@iteam.upv.es

ABSTRACT

Wave-Field Synthesis (WFS) is a spatial sound reproduction technique that has attracted the interest of many researchers in the last decades. Unfortunately, although WFS has been shown to provide excellent localization accuracy, this property is restricted to sources located in the horizontal plane. In order to deal with this problem, a spectral-filtering-based solution has been recently proposed for achieving slight elevation effects. This paper reports new experiments regarding side and rear reproduction of elevated virtual sources in WFS. Results show that elevation can be also perceived in these cases, confirming the validity of the spectral-filtering approach.

1. INTRODUCTION

Wave-Field Synthesis (WFS) is a spatial sound rendering technique capable of producing a realistic acoustic field in an extended area by means of loudspeaker arrays. WFS achieves excellent localization accuracy in the horizontal plane but, unfortunately, it is not possible to reproduce virtual sources located above or below this plane. This is a clear disadvantage of WFS in comparison to other spatial sound reproduction systems, as for example, Ambisonics or 10.2 Surround. In this context, 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. However, the phantom effect does not work in elevation as good as in azimuth and these systems do not always provide the desired quality [10]. Recently, the authors proposed to use Head-Related Transfer Function (HRTF) spectral elevation cues in conjunction with WFS for producing in the listeners the sensation of elevated virtual sources [14]. In this hybrid system, azimuth localization is achieved with the usual WFS system, but elevation is simulated by means of a filtering stage prior to WFS rendering. Several elevation responses were computed from different HRTF databases and a set of listening tests were conducted to assess localization of elevated sources in the median plane. Results showed that most listeners could correctly perceive differences between two sources located at different elevation angles if their angular spacing was not very small ($> 10^{\circ}$).

In this paper, new listening tests are carried out for assessing listeners' discrimination capability between sources located at different elevation angles but considering two extreme cases: rear reproduction and side reproduction. These two cases reflect how elevation effects can be perceived for sources located at any azimuth position, thus, noticeable in the whole WFS area. The paper is structured as follows. In Section 2, a brief introduction to WFS is presented. Section 3 describes azimuth and elevation localization cues and Section 4 explains how to achieve elevated sound sources in WFS with spectral-filtering. Experiments and results are discussed in Section 5. Finally, the conclusions of this work are summarized in Section 6.

2. WAVE-FIELD SYNTHESIS

One of the most important cues in spatial perception of sound is localization. Generally, sound is perceived in all three dimensions (width, height and depth) and all of them are necessary to achieve a natural perception of sound [16]. Over the last few decades, surround sound has played an increasingly important role in the entertainment industry, as well as in the field of multimedia, reproducing sound in a way which is more "natural" with the aim of enhancing the listening experience. Although five channel systems are a consolidated standard in multichannel audio today, there is increasing interest in emerging reproduction systems based on sound field rendering. The most popular of these systems is Wave-Field Synthesis (WFS), a spatial sound reproduction technique capable of synthesizing an acoustic field in an extended area by means of loudspeaker arrays. This property makes it possible to synthesize a sound scene with the correct spatial characteristics for each listener [4][5][7]. Therefore, every listener perceives his own perspective of the reproduced scene and the experience is closer to natural hearing. However, creating a copy of a sound field is not completely feasible due to some practical constraints:

- 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 synthetisized 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 and restricts localization to the horizontal plane.

Several methods for dealing with the problems mentioned above can be found in the literature [17]. However, of all these problems, the third one is the most important (and difficult to solve) in terms of source localization, as it prevents the audience from having a full 3D experience.

3. AZIMUTH AND ELEVATION LOCALIZATION CUES

The HRTF describes how a given sound is filtered by the 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 sound and play a very important role in the neural determination of source location, particularly the determination of source elevation [12].

Binaural localization is based on the comparison of the auditory inputs from our two ears. Azimuth localization is performed via the "interaural time difference" (ITD) and the "interaural level difference" (ILD) cues [6]. The primary biological binaural cue is ITD and it is related to the time delay that appears when sound reaches the near ear and when it reaches the far ear. The ILD cue, less significant, is given by the reduction in loudness when the sound reaches the far ear. These cues will only aid in localizing the sound source in the azimuth plane, not in elevation. To get instantaneous localization in more than two dimensions from ITD and ILD more than two detectors are required. However, there are complex variations in the degree of attenuation of a sound in travelling from the source to the eardrum. These frequency-dependent attenuations are related to both azimuthal angle and elevation and can be summarized in the HRTF. As a result, an estimation of the source location in azimuth and elevation can be performed with wideband sounds. Additional information can be found by moving the head, so that the HRTF for both ears changes in a way known by the subject. In [15], a spatial audio system was proposed based on the assumption that the spectral cues are common in any sagittal plane (see Figure 1). In that system, sound localization in any direction was successfully achieved by combining HRTF filtering in the median plane and interaural differences. However, subjects' own HRTFs were used in the listening tests and therefore, not a single HRTF model was considered for their experiments.

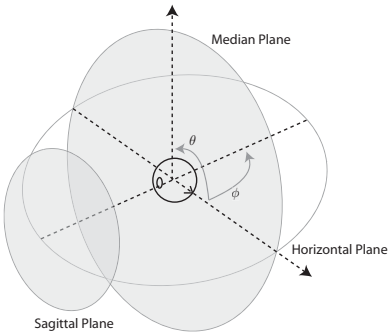

Figure 1: System Geometry

Several attempts have been made to model HRTFs, both to understand their behavior and to simplify the binaural synthesis process. The task of modeling HRTFs has always encountered four major problems [11]:

1. the difficulty of approximating the effects of wave propagation and diffraction by simple, low-order parameterized filters.
2. the complicated joint dependence of the HRTFs on azimuth, elevation and range.
3. the lack of a quantitative criterion for measuring the accuracy of an approximation
4. great person-to-person variability of HRTFs.

Batteau [3] showed that the pinnae play a critical role in determining elevation, and he conjectured that two major ridges in the outer ear act like reflecting surfaces, producing multipath echoes whose timing gave the cues for elevation. Various models have been proposed in the literature [8][9][18]. However, while azimuth effects are readily modeled, accurate modeling of elevation cues is still a challenge.

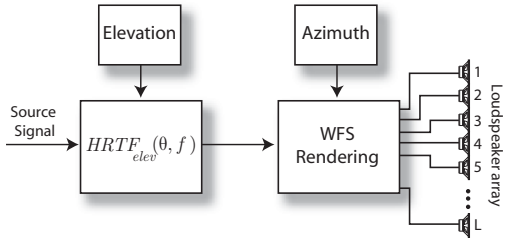

Figure 2: Proposed HRTF-WFS Hybrid System

4. WAVE-FIELD SYNTHESIS WITH ELEVATION

Despite the difficulty of getting a universal HRTF model, the authors proposed in [14] a WFS system that includes elevation after examining an extensive set of HRTFs. The following aspects were considered in its design:

1. elevation should be perceived in any location inside the listening area in order to be consistent with the main WFS advantage,
2. the computational cost must be acceptable,
3. elevation effects should be perceived by all individuals,
4. the basic WFS field should remain the same if no elevation is considered in the positioning of the virtual sources.

The block diagram of the proposed hybrid HRTF-WFS system is depicted in Figure 2. As can be seen, elevation is achieved by filtering the virtual source signals according to elevation cues measured in the median plane. In order to fulfill the specified requirements, the following issues were examined:

1. if the sources have an azimuth angle different from $\phi = 0$, still some elevation will be perceived. This is due to the fact that elevation spectral cues are common in any sagittal plane [15].
2. Also, as these spectral cues are the same in any sagittal plane, only a reduced set of filters is required. In addition, the sampling in elevation does not need to be very accurate, so there is no need for excessive storage resources.
3. Several HRTF databases were analyzed in order to look for common spectral features in the median plane. Although the aim is not to provide an accurate universal HRTF that works perfectly for any individual, some common features may cause a noticeable elevation effect.
4. Filtering does not affect the sources when they are located in the horizontal plane, and so no modifications to the original WFS are introduced in this case.

4.1 Selection of HRTF elevation cues

In our previous paper, we examined two public HRTF databases provided by IRCAM [13] and CIPIC [2] in order to look for common elevation cues. Also, a set of impulse

responses from the *Roland Sound Space Processor RSS-10*, were measured for comparing synthetic filtering approaches to real-measured responses. By analyzing these responses, a peak-filtering approach using a chain of second order IIR filters for achieving elevation cues was also introduced and discussed. These synthetic set of responses are hereafter denoted as PEAK. Next, we briefly describe how elevation filters were computed from this set of responses.

4.1.1 Removing azimuth cues

For all the HRTF databases, only responses for the median plane ($\phi = 0$) were considered. Assuming far field conditions, the HRTF is a function of the direction-of-arrival of the source (ϕ, θ) and frequency f, expressed as $HRTF(\phi, \theta, f)$. Moreover, although there is a different HRTF for the right and left ears, they can be considered to be symmetric in ϕ. It is important to note that, even in the case of a source located in $\phi = \theta = 0$, the HRTF is not a neutral function of frequency, but it has a filtering effect for that direction. Although this effect is important for headphone binaural reproduction, it is not of interest for systems with frontal loudspeaker reproduction, as this effect will be naturally produced in the sound path from the loudspeakers to the listener. Therefore, it becomes necessary to eliminate the filtering effect produced by a head exposed to a front coming sound and retain only filtering effects due to elevation cues. For this purpose, HRTFs were normalized as follows:

$$HRTF_{elev}(\theta, f) = \frac{HRTF(0, \theta, f)}{HRTF(0, 0, f)} \quad (1)$$

where $HRTF_{elev}(\theta, f)$ is now a neutral function of frequency for sources with $\theta = 0$, $(HRTF_{elev}(0, f) = 1)$. Taking into account the diagram shown in Figure 2, this function results totally coherent with the hybrid system involving HRTF and WFS. Notice, that the original WFS system remains the same when the sources are considered to be located in the horizontal plane.

5. EXPERIMENTS

In our previous work, we obtained satisfying results regarding frontal reproduction for source locations in the median plane. Experiments to evaluate listeners' capability of identifying the direction of moving sources in the median plane (from -40° to 40°) were conducted for the 4 different filterbanks (IRCAM, CIPIC, RSS and PEAK). In addition, the capability to differentiate the highest source from two successive sounds corresponding to different elevation angles was also assessed. The outcome of these experiments showed that elevation could be perceived for all the databases considered. Although differences could be easily perceived for angular separations greater than 20°, it became difficult to notice differences for angular spacing below 10°. In this new work, it becomes necessary to examine if these results are also in concordance for those cases in which the subject is not directly in front of the virtual source. Moreover, the capability to decide which is the highest source when two simultaneous stimuli are present is also studied. Note that these two experiments are carried out for all the databases considered in [14] (IRCAM, CIPIC, RSS and PEAK).

Next, we describe two experiments carried out for testing how elevation of virtual sources is perceived in the cases of side and rear reproduction. Listening tests were carried out for assessing the locatedness of elevated virtual sources in the cases of side and rear reproduction. Also, a higher/lower discrimination test between two simultaneous stimuli located at different azimuth positions was carried out. These stimuli consisted of uncorrelated full band pink noise bursts sampled at 44.1 kHz. A panel of 12 subjects took part in these tests, including students and people involved in audio research with ages between 23 and 38. The tests were conducted using a 72 loudspeaker WFS array. This array is placed inside our recording studio, which is acoustically treated to get a T_{60}@1000Hz < 0.25 s. The volume of this room is 96 m^3 and its floor size is 4 by 9.1 m. The background noise inside the studio was below 25 dBA. Each element of the array is a two-way system, using a 5"$\frac{1}{2}$ woofer and a 1" tweeter. The loudspeaker separation of the array is 180 mm. The spatial aliasing frequency for this arrangement is about 1 kHz in the worst case. No special aliasing improvement techniques were used in the rendering process. The coloration introduced by spatial aliasing artifacts has a very irregular pattern, which is supposedly averaged by the human auditory system [1]. These artifacts have been shown to have little influence on the subjective assessment of sound quality and localization of virtual sources in the horizontal plane, and it is expected that the influence on elevation spectral cues is as well negligible.

5.1 Experiment 1: Locatedness in Elevation

The definition of locatedness, according to Blauert [6], is the degree to which an auditory event can be said to clearly be in a particular location. We preferred to evaluate locatedness instead of localization accuracy for having preliminary results regarding the performance of the different databases considered.

For the evaluation of locatedness, filtered pink noise for $\theta = -30^\circ$ and $\theta = 30^\circ$ was presented at 5 different azimuth positions between $\phi = 90^\circ$ and $\phi = 270^\circ$, asking the listener about the ease to localize (in elevation) the presented source. A 5-grade scale was used, ranging from 1 (very bad) to 5 (very good). *How well can you localize the source in elevation? How well can you assign a particular direction to the perceived source below or above the horizontal plane?*

1. very bad
2. bad
3. fair
4. good
5. very good

Figure 3 shows the locatedness of elevated sources located at rear and side positions for the different databases considered. In Figure 3a., results are given for a source with elevation $\theta = -30^\circ$ and in Figure 3b., for a source with elevation $\theta = 30^\circ$. We can observe these main differences:

- Locatedness for the case $\theta = 30^\circ$ is better than for $\theta = -30^\circ$. A similar conclusion was also observed in our previous work, where elevation for sources above the horizontal plane was more evident for the listeners.
- Locatedness of elevated sources seem to be worse for side reproduction than for rear reproduction, as can be seen in the results for azimuth positions $\phi = 90^\circ$ and $\phi = 270^\circ$.
- There are no significant differences between the

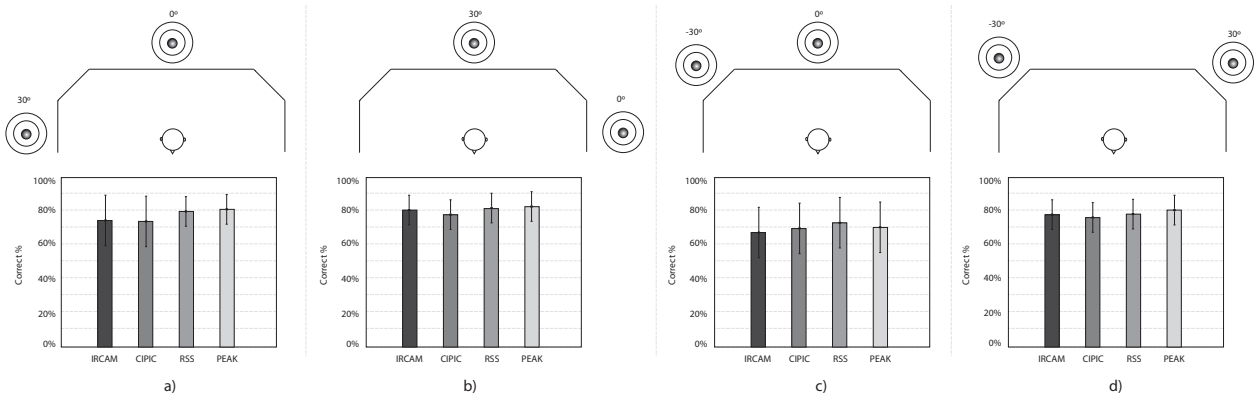

Figure 4: Hit rate for the Higher/Lower discrimination experiment for each database. Each panel describes one of the four situations described in Section 5.2

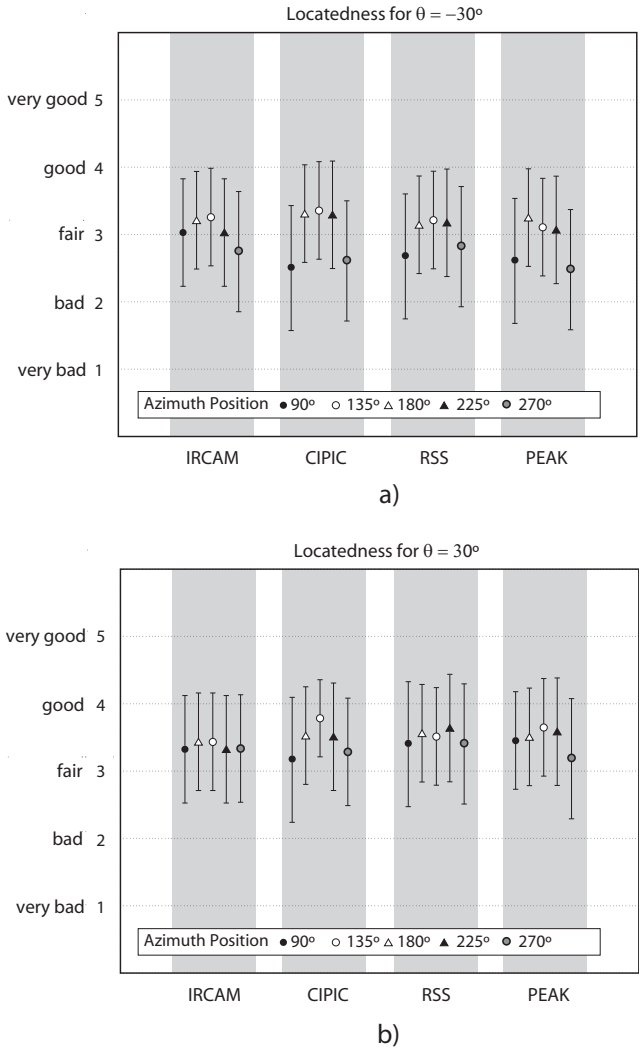

Figure 3: Subjective assessment of locatedness, showing mean and 95% confidence intervals for each system and azimuth position. a)Virtual source at $\theta = -30^\circ$. b) Virtual source $\theta = 30^\circ$

databases considered, and means are almost always in the fair-good range.

5.2 Experiment 2: Higher/Lower Discrimination of Simultaneous Sources

In this experiment, the capability of identifying which source is at a higher position from two simultaneous sounds filtered with different θ responses was studied. The sounds were again two filtered pink noise bursts corresponding to different values of θ, and were positioned in the WFS system at different azimuth angles between $\phi = 90^\circ$ and $\phi = 270^\circ$. The following situations were presented to each listener:

1. **Source 1**: $\phi = 270^\circ$ $\theta = 30^\circ$. **Source 2**: $\phi = 180^\circ$ $\theta = 0^\circ$.
2. **Source 1**: $\phi = 180^\circ$ $\theta = 30^\circ$. **Source 2**: $\phi = 90^\circ$ $\theta = 0^\circ$.
3. **Source 1**: $\phi = 225^\circ$ $\theta = -30^\circ$. **Source 2**: $\phi = 180^\circ$ $\theta = 0^\circ$.
4. **Source 1**: $\phi = 225^\circ$ $\theta = -30^\circ$. **Source 2**: $\phi = 135^\circ$ $\theta = 30^\circ$.

These situations were randomly chosen and repeated 4 times. Listeners had to indicate approximately the direction of the source that was higher.

The results for this experiment are shown in Figure 4, showing the rate of correct answers for all the situations described in Section 5.2. The following observations can be made:

- For all situations, the hit rate remains in the range between 60% and 80%.
- Although there are not very big differences between databases, the RSS and PEAK filter banks show better performance than IRCAM and CIPIC.
- Results seem to be better when the higher source is above the horizontal plane. This observation may also reveal again the fact that elevation is more clearly perceived for $\theta > 0$.

6. CONCLUSION

In this paper, we have further explored the possibilities of producing elevation effects in Wave-Field Synthesis Systems. A recently proposed system that includes a filtering stage previous to WFS rendering was shown to be able to simulate virtual sources above and below the horizontal plane. The filtering stage is essential for producing this sensation and is based on common HRTF elevation cues extracted from public databases. This hybrid system (HRTF-WFS) has been tested in two new aspects that play an essential role in WFS: rear and side reproduction of elevated

virtual sources. Several databases and synthesis filters have been considered for that purpose and two experiments have been carried carried out. The first one is related to the locatedness in elevation of the sources. The second one is aimed at assessing listeners' ability to discriminate in elevation in the presence of two simultaneous sources. Results show that, although it is difficult to perceive a clear direction in elevation for side and rear reproduction, it is feasible to produce slight elevation effects for an audience in an extended area, even when multiple sources are active. Future works will consider coloration of the sources introduced by this filtering stage and will evaluate localization accuracy for complex sound scenes of real recorded sources.

REFERENCES

[1] J. Ahrens and S. Spors. Alterations of the temporal spectrum in high-resolution sound field reproduction of different spatial bandwidths. In *Proceedings of the AES 126th Convention*, Munich, Germany, May 2009.

[2] V. R. Algazi, O. R. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *In Proceedings of IEEE WASPAA'01*, New Paltz, NY, USA, October 2001.

[3] D. W. Batteau. The role of the pinna in human localization. *Proceedings of the Royal Society of London*, 168:158–180, 1967.

[4] A. J. Berkhout. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36:977–995, 1988.

[5] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93:2764–2778, 1992.

[6] J. Blauert. *Spatial Hearing*. MIT Press, Cambridge, 1997.

[7] M. M. Boone, E. N. G. Verheijen, and P. F. van Tol. Spatial sound field reproduction by wave field synthesis. *Journal of the Audio Engineering Society*, 43(12):1003–1012, 1995.

[8] C. P. Brown. Modeling the elevation charactersitics of the head-related impulse response. Master's thesis, San Jose State University, May 1996.

[9] C. P. Brown and R. O. Duda. An efficient HRTF model for 3-D sound. In *Proceedings of the IEEE Workshop on ASSP*, New Paltz, NY, USA, 1997.

[10] W. P. J. deBrujin and M. Boone. Application of wave field synthesis in life-size videoconferencing. In *Proceedings of the 114th AES Convention*, Amsterdam, The Netherlands, 2003.

[11] R. O. Duda. Modeling head related transfer functions. In *Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers*, Asilomar, CA, USA, October 1993.

[12] P. M. Hofman, J. G. A. Van-Riswick, and J. Van-Opstal. Relearning sound localization with new ears. *Nature Neuroscience*, 1(5), September 1998.

[13] IRCAM. LISTEN HRTF database. available online at: http://http://recherche.ircam.fr/equipes/salles/listen/, 2003 (last update).

[14] J. J. Lopez, M. Cobos, and B. Pueo. Elevation in wave-field synthesis using HRTF cues. *Acta Acustica united with Acustica*, 2009. Submitted.

[15] M. Morimoto, M. Itoh, and K. Iida. 3-D sound image localization by interaural differences and the median plane HRTF. In *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, July 2002.

[16] F. Rumsey. *Spatial Audio*. Focal Press, 2001.

[17] H. Wittek. *Perceptual differences between wavefield synthesis and stereophony*. PhD thesis, School of Arts, Communication and Humanities, University of Surrey, October 2007.

[18] W. Zhang, T. D. Abhayapala, and R. A. Kennedy. Modal expansion of HRTFS: Continuous representation in frequency-range-angle. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009.