

Adding 3D Sound to 3D Cinema: Identification and Evaluation of Different Reproduction Techniques

Cédric André, Jean-Jacques Embrechts, Jacques G. Verly
*Laboratory for Signal and Image Exploitation (INTELSIG),
Department of Electrical Engineering and Computer Science,
University of Liège, Liège, Belgium*
C.Andre@ulg.ac.be

Abstract

Very little research has been conducted so far into the general problem of producing a 3D soundscape consistent with the visual content of a 3D-stereoscopic movie. First, the following 3D sound reproduction techniques are reviewed: Vector Base Amplitude Panning (VBAP), binaural and transaural techniques, Wave Field Synthesis (WFS) and Ambisonics. Second, the new challenges of 3D cinema are introduced. Third, we reconsider each 3D audio technique in the light of these challenges. We find that, at least in theory, a completely personalized soundscape is needed and that, due to various technical reasons, only binaural reproduction through headphones is able to accurately produce a 3D soundscape consistent with a 3D-stereo movie in a theater environment.

1. Introduction

1.1. Why 3D audio ?

3D audio aims at reproducing the correct sensation of direction and distance to the sound source as well as the room effect. The acoustic sweet spot, where sound reproduction is controlled, can either be a portion of the space or the listener's ears.

Today's cinema sound systems are designed to immerse the spectators in sound and audio sources that are confined to the horizontal plane at the height of the listener's ears [1]. 3D audio techniques are able to produce such a rendering, but can also do much more in terms of localization. It is thus interesting to investigate the use of 3D audio in a movie theater environment.

As a first step, this paper focus on the audio rendering technique that would best fit this type of environment, by extending the ideas presented in [2].

1.2. 3D audio and virtual reality

Few systems combine 3D-stereoscopic video with 3D sound capabilities. Systems that do are often virtual reality systems such as in [3] and [4].

Virtual reality systems occasionally allow several users to experience together video and audio. However, this small number of users is not comparable to the large number of (static) persons a movie theater system should be able to handle. The significant number of persons in a theater room is thus an important characteristic of the problem considered here.

As a consequence, attention should be paid to the space in which audio has to be rendered because it will typically be larger for a movie theater application than it is for a virtual reality system.

For all these reasons, we can deduce that a technology used in virtual reality systems may not necessarily be the best choice for movie theater applications, so that a review of the existing techniques would be useful.

This is the reason why Section 2 is dedicated to the analysis of the major, existing 3D sound reproduction systems. The challenges of 3D cinema are then listed in Section 3. With these challenges in mind, each reproduction technique is reconsidered in Section 4 to select possible candidates for the movie theater application. Finally, in Section 5, the results of the discussion are summarized, and suggestions are made for future research.

2. 3D audio technology

In this section, we review the following 3D audio reproduction techniques: Vector Base Amplitude Panning (VBAP), binaural techniques, transaural techniques, Wave Field Synthesis (WFS) and Ambisonics.

2.1. Vector Base Amplitude Panning (VBAP)

Stereophony is based on an artificial placement of virtual sound sources between two loudspeakers equidistant from the listener. One of the most important techniques to achieve this is the amplitude panning technique [5]: by applying different amplitude gains to a monophonic recording, the sound appears to be coming from a location between the two loudspeakers.

VBAP [6] extends this idea by adding a third loudspeaker to the stereo pair, allowing virtual sources to be created above or below the horizontal plane, but still

between the loudspeakers. A complete 3D soundscape is reproduced using more than three loudspeakers arbitrarily placed on a sphere surrounding the listener. The listening room is assumed to be not too reverberant.

The source is restricted to lie outside of the sphere but can be created along any direction by the three closest loudspeakers. Each loudspeaker gain is simply obtained by linear algebra involving only its direction with respect to the listener and the direction of the phantom source.

If $\mathbf{l}_i = [l_{i1} \ l_{i2} \ l_{i3}]^T$ is the unit-length vector corresponding to the direction of the i^{th} loudspeaker and the desired phantom source direction is the unit-length vector

$$\mathbf{p} = [p_1 \ p_2 \ p_3]^T, \quad (1)$$

then we can write

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2 + g_3 \mathbf{l}_3, \quad (2)$$

and we can solve for the gain factors $\mathbf{g} = [g_1 \ g_2 \ g_3]$ of the three loudspeakers,

$$\mathbf{g} = \mathbf{p}^T L_{123}^{-1}, \quad (3)$$

where $L_{123} = [\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3]^T$ and its inverse L_{123}^{-1} exists if the three loudspeakers define a basis that spans the three dimensional-space.

2.2. Binaural techniques

Binaural techniques aim at producing a correct sound field directly at the listener's ears. Headphones are therefore the natural reproduction system for binaural signals as they offer complete channel separation and a relative acoustic protection from the environment (although this may not be entirely desirable as will be seen later).

Sound spatialization, however, is achieved based on cues that are mostly derived from the interaction of sound with the listener's head and torso. Since using headphones bypasses this interaction, it is necessary to make sure that spatialization is included in the signal fed to each ear.

This can be achieved in two different ways. On the one hand, we can record two signals with two microphones placed in a subject's ears or with two microphones that simulate ears on a dummy head. In this case, the recorded signal contains the spatial information. On the other hand, it is possible to process a monophonic signal that does not contain any spatial information, to accurately simulate the effect of propagation of the signal from the source to the listener. This processing consists of a convolution with the Head Related Impulse Responses (HRIRs), or, in the frequency domain, of a multiplication with the Head Related Transfer Functions (HRTFs) which are the Fourier transforms of the HRIRs. In the first case, the listener hears through someone else's ears, whereas, in the second case, it is possible to hear through one's own ears, provided that personalized HRTFs were measured.

HRTFs consist of a set of two functions, one for each ear, that are the mathematical expression of the interaction of sound with the listener's head and torso. They depend on the position of the source, which we express here in terms of a radius R to the listener, an azimuth angle θ , and an elevation angle ϕ . They vary significantly from one subject to the other, and using someone else's HRTFs when listening to a binaural recording can lead to errors in spatial localization judgment [7].

The complete recording and playback chain has to be calibrated in order to reproduce the correct pressure at the eardrum. When the recording is made outside of blocked ear canals, the equalization function G can be expressed as [8]

$$G = \frac{1}{M \cdot \text{PTF}} \cdot \text{PDR}, \quad (4)$$

where M is the recording microphone's transfer function, PTF is the headPhone Transfer Function, measured at the position in the ear canal where the recording is made, and PDR is the Pressure Division Ratio.

Headphones rarely present a uniform frequency response. Their PTF has smooth variations at low frequency and rather marked peaks at high frequencies [9]. To ensure that the headphones reproduce the correct pressure at the eardrums, it is necessary to compensate for their behavior.

PDR is best understood using an electrical equivalent for the free-field sound transmission to the external ear (Fig. 1). The Thévenin equivalent pressure source P_{Th} and its impedance Z_{Th} model the complete sound field outside the ear canal. When the ear canal is blocked, P_{Th} and P_{entrance} are respectively measured outside and inside the blockage. The ear canal is modeled by a two-port element loaded by the eardrum impedance. The impedance seen from the entrance of the ear canal is called $Z_{\text{ear canal}}$.

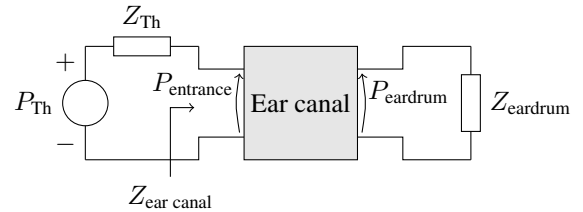


Figure 1. Electrical equivalent for the model of free-field sound transmission to the human external ear (after [8]).

During recording, the source impedance Z_{Th} is the radiation impedance $Z_{\text{radiation}}$ that is seen from the entrance of the ear canal looking out into the free field. During reproduction, Z_{Th} is the headphone impedance $Z_{\text{headphone}}$. PDR is the ratio of the entrance pressures

during recording and reproduction. We have

$$\text{PDR} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}}. \quad (5)$$

Determining PDR would thus theoretically require three acoustic impedance measurements. Headphones that have a PDR equal to unity are called Free-air Equivalent Coupling (FEC) headphones. Although not every headphone exhibit this property, Møller *et al* [9] have shown that, in practice, the deviation induced by the PDR is small compared to the PTF.

2.3. Transaural techniques

We use the term transaural when the binaural recording is reproduced through loudspeakers that act as “virtual headphones.” In this case, channel separation is lost and interchannel crosstalk cancels the localization cues present in the binaural signal. An adequate processing is therefore required.

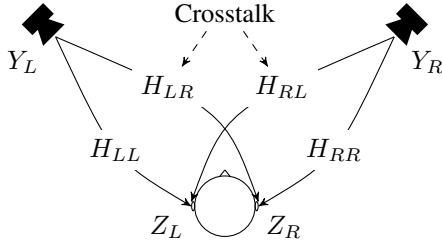


Figure 2. Crosstalk in transaural reproduction (after [10]).

According to Fig. 2, the signals Y_L and Y_R produced by the loudspeakers are different from the ear signals Z_L and Z_R because of crosstalk [10]. Mathematically, we have

$$Z_L = H_{LL} \cdot Y_L + H_{RL} \cdot Y_R \quad (6)$$

$$Z_R = H_{RR} \cdot Y_R + H_{LR} \cdot Y_L, \quad (7)$$

where the functions H_{XY} are HRTFs normalized with respect to the free-field response at the center of the head, but with no head present [11]. The effects of loudspeaker responses and propagation to the ears are not taken into account in these relations.

Solving (6) and (7) for Y_L and Y_R yields the expression

$$Y_L = \frac{H_{RR}}{H_{LL}H_{RR} - H_{LR}H_{RL}} Z_L - \frac{H_{RL}}{H_{LL}H_{RR} - H_{LR}H_{RL}} Z_R \quad (8)$$

and a similar one for Y_R .

This solution is stable only in the angle spanned by the loudspeakers. Furthermore, this solution requires

that the listener be located at a particular position as the sweet spot is only a few centimeters wide in the left – right direction [12]. To allow for, say, head rotation, head tracking is necessary as well as the use of adaptive filtering [11].

2.4. Wave Field Synthesis (WFS)

According to Huygens’ principle, each point of a wavefront can be considered to be a secondary source. By picking up all these secondary sources with a microphone array, it is possible to reconstruct the whole wavefront with a loudspeaker array. This is the idea behind Wave Field Synthesis (WFS), originally suggested in [13]. Here, we are interested in solutions of the form

$$p(\vec{r}, t) = P(\vec{r})e^{j\omega t}. \quad (9)$$

When no sound source is present at \vec{r} , P satisfies the homogeneous Helmholtz wave equation

$$\nabla^2 P(\vec{r}) + \left(\frac{\omega}{c}\right)^2 P(\vec{r}) = 0 \quad (10)$$

in a closed volume Ω surrounded by a surface Λ .

If a sound source is located at r_0 outside of Ω , then the pressure at a point \vec{r} inside the volume is given by

$$P(\vec{r}, \vec{r}_0) = \int_{\Lambda} \left[G(\vec{r}|\vec{r}_q) \frac{\partial P(\vec{r}_q)}{\partial \vec{n}_q} - P(\vec{r}_q) \frac{\partial G(\vec{r}|\vec{r}_q)}{\partial \vec{n}_q} \right] d\Lambda, \quad (11)$$

where \vec{r}_q denotes a generic position on Λ and \vec{n}_q the outward pointing normal at \vec{r}_q , and $G(\vec{r}|\vec{r}_q)$ is the Green function, which is the solution of the inhomogeneous Helmholtz equation

$$\nabla^2 G(\vec{r}|\vec{r}_q) + \left(\frac{\omega}{c}\right)^2 G(\vec{r}|\vec{r}_q) = -\delta(\vec{r} - \vec{r}_q), \quad (12)$$

where δ is the Dirac “delta function” of appropriate dimensionality. In three dimensions, we have

$$G(\vec{r}|\vec{r}_q) = \frac{e^{-jk|\vec{r}-\vec{r}_q|}}{4\pi|\vec{r}-\vec{r}_q|} + F, \quad (13)$$

where $k = \omega/c$ and F is any function satisfying the Helmholtz wave equation (10).

Considering that most information is included in the horizontal plane at the height of the listener’s ears, Berkhout *et al* [14] have suggested the use of linear arrays of loudspeakers.

However, a linear array produces unwanted diffraction waves because of its finite length. Furthermore, diffraction waves appear at corners between two arrays. Correcting factors for the loudspeaker pressures are therefore provided in [14].

The main drawback of WFS is the number of loudspeakers needed and their small size. As long as the

frequency is below a certain aliasing frequency f_{Nyq} , the reproduction is stable in the whole playback area. However, because of spatial sampling, if the successive loudspeakers along the array are separated by a distance Δx , then the reproduction is incorrect at frequencies above

$$f_{\text{Nyq}} = \frac{c}{2\Delta x}. \quad (14)$$

This implies a spacing Δx of at most 8.5 mm to achieve a correct reproduction up to 20 kHz. WFS systems often count more than a hundred loudspeakers, each with its own channel.

2.5. Ambisonics

Ambisonics is a technique based on Gerzon's work [15]. It is designed to reproduce plane waves coming from any direction. Instead of using integral theorems as in WFS, the solution to (10) is given in polar coordinates using a Fourier-Bessel series expansion of the pressure [16].

In this case, the function $P(\vec{r})$ is expressed as

$$P(\vec{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l P[l, m] j_l(kr) y_l^m(\theta, \phi) \quad (15)$$

where the spherical Bessel functions $j_l(kr)$ describe the radial dependance and the spherical harmonics $y_l^m(\theta, \phi)$ describe the angular behavior of the acoustic field. The coefficients $P[l, m]$ must be identified for a given sound field, and they are used as input to the reproduction system.

The value of l , called the order of the expansion, has to be limited in practice. This value determines the number of channels L required for playback according to

$$L = (l + 1)^2. \quad (16)$$

Consequently, the first order expansion ($l = 1$) in 3D has four components: the absolute pressure (W) and the three pressure gradients in three orthogonal directions (X , Y , and Z). This expansion is known as the B-Format. The SoundField Microphone natively captures a sound field in B-Format [17].

Truncating the expansion reduces the sweet spot when the frequency increases. However, the representation remains valid whatever the frequency, provided that the listener is at the sweet spot.

Note that the spherical wave expansion is independent of the reproduction system. For example, it is possible to use Ambisonics coding with a binaural system [18].

3. New challenges of 3D cinema

This section addresses different questions that arise when one considers adding 3D sound to 3D movies.

We first focus on the issues related to the sound volumes and image volumes. We then establish the difference in localization of visual cues between viewers.

3.1. Sound volumes versus image volumes

A first observation concerning today's average cinema layout is that the volumes in which visual objects and auditory objects appear do not perfectly overlap (Fig. 3).

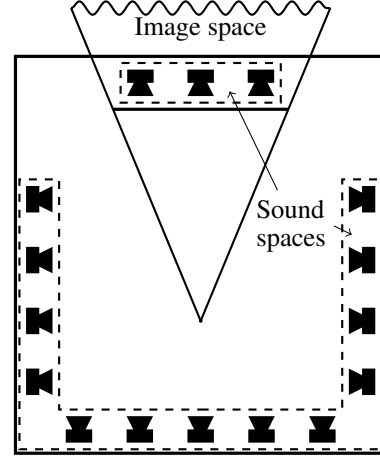


Figure 3. The sound and image spaces (after [19]).

Sound, on the one hand, is typically produced by three loudspeakers behind the screen, acting as a three-channel stereo source, and by a set of three linear arrays located on the sides and in the back, acting as a surround source to give the audience a sense of immersion.

Vision, on the other hand, is limited to our field of view. It is roughly a truncated cone with the apex at the viewer that extends towards infinity behind the screen. In this volume, there is a certain zone in which the two stereo images can be fused by the viewer's brain to give a visual perception of a single 3D object. This zone is called the Panum's fusional zone [20]. Outside of this zone, double vision occurs, causing discomfort.

Ideally, one would want to render sound sources in a volume at least corresponding to the audience's field of view. The question of whether a theater sound system should be able to render sound in the whole 3D sphere needs to be addressed by the people involved in the process of 3D movie making.

Historically, when Dolby Labs introduced their version of surround sound, attention was mostly given to the sound events in the horizontal plane, because they are far more likely than events above or under the audience and also because a movie director avoids taking the audience attention away from the screen [21]. The fourth surround channel was introduced to immerse the spectators in the movie they are watching, by producing a sound that has no particular direction of arrival. This channel has always served as a way to provide ambience as well as special sound effects. Since 3D audio is capable of precise localization, the use of

this channel in conjunction with a 3D movie seems contrary to this trend. Therefore, there is a need for psychological studies to show whether or not a realistic soundscape improves the perceptual immersion, i.e. the attention given to the movie, as compared to a physical immersion, such as provided by Dolby's fourth channel.

If movie makers prefer to limit 3D sound to objects inside the field of view, one can imagine a hybrid solution that keeps this ambience channel. In that case, 3D audio would be used only to render the material that was previously sent to the three front channels, including the dialogs that are never spatialized for lip-sync reasons. The ambience channel can then be rendered either by the existing loudspeaker arrays or by virtual loudspeakers rendered through a full 3D-capable audio system.

3.2. Visual localization

By essence, the illusion of depth perception in 3D cinema is created by presenting a different image to each eye. Each image is physically located on the screen and all the viewers look at the same image pair. The above-mentioned truncated cone thus follows each spectator and, when one compares the visual perception of two spectators seated at different locations in the room, one concludes that the objects of the scene are not rendered at the same physical location in the room.

In Fig. 4, two viewers *A* and *B* look at the same stereoscopic image of the sketch of a loudspeaker. This loudspeaker appears in front of the screen for both viewers, but their visual cue, i.e. the locations of their visual perception, are located at two different positions. Note that, for clarity, only one line segment of each visual cue is represented in the figure.

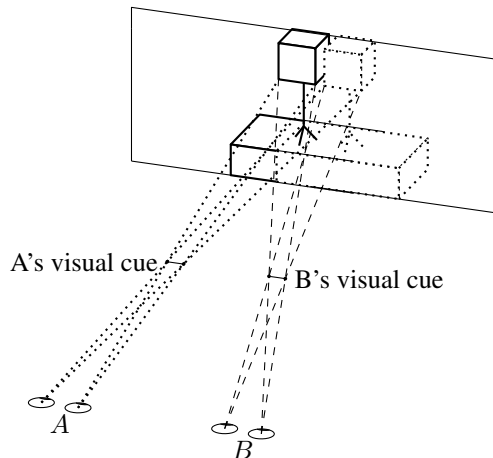


Figure 4. Visual localization for two viewers *A* and *B* seated at two different locations.

The seating location in the room is not the only factor that impacts the visual localization [22]. The distance from the viewer to the object is dependent on the viewer's interocular distance, i.e. the distance

between his/her two eyes. The interocular distance mainly depends on ethnicity, gender, and age.

For example, imagine a child, a woman, and a man, each having the average interocular distance of their category, sitting in the middle of the theater at 20 m from the screen and looking at a bell that appears in front of it. If the man sees the bell at 14 m from the screen, then the child sees it at 15.08 m (closer to him), and the woman at 13.55 m (farther from her).

From these disparities in visual localization, we argue that the soundscape should be personalized to take into account the position of the viewer and his physical characteristics.

4. 3D audio and cinema

4.1. Vector Base Amplitude Panning

The use of VBAP techniques for 3D sound reproduction in a 3D-movie theater would necessarily be limited since it is unrealistic to enclose the room in a sphere of loudspeakers. VBAP could instead be used as an upgrade of the stereo pair behind the screen, provided that at least one loudspeaker is added above the others. Even then, the distance cue would be poor and the sources would be limited to lie behind the screen. Such a setup would thus suit a 2D-video content better.

Adding the fact that VBAP reproduction is constrained to a sweet spot and cannot be fully personalized, we conclude that VBAP is not a right choice for the application of 3D cinema.

4.2. Binaural techniques

Binaural techniques allow one to reproduce at the listener's ears a complete 3D soundscape from a pair of headphones. The reproduction can be personalized and integrated in a multi-user system. This technology is therefore a right choice for the application considered here. This conclusion is not surprising, considering that binaural reproduction is the audio equivalent of 3D-stereo for video.

Combining HRTFs with monaural recording has an important advantage: it is indeed common to record sound in mono on a movie set [23].

Headphones also allow for a hybrid solution combining loudspeakers by using a solution such as the one suggested by Martin *et al* in [24] for augmented reality. The loudspeaker system could be used for low frequency effects, difficult to render through headphones.

A few difficulties have yet to be overcome. It is illusory at this point to provide every person with his/her own HRTFs. Psychoacoustic experiments should therefore be performed to evaluate the localization performance by the members of the audience when presented with both 3D-stereo video and binaural audio.

Head tracking should be provided to ensure that the sound field remains fixed with respect to the room and not with respect to the listener himself. Indeed the listener should be able to turn his head without the

soundscape following this movement. The associated real-time processing highly increases the required computing power.

Lastly, the audience will need to be convinced that being able to experience a 3D movie with 3D sound is worth the addition of headphones to the required pair of 3D glasses.

4.3. Transaural techniques

By comparison with binaural techniques, transaural techniques free the audience from headphones and simplify the filtering as listeners use their own ears. However, the limited sweet spot forces one to provide each spectator with his own transaural system (two or four loudspeakers), and the associated crosstalk would not only have to be dealt with for all users, but it would also have to be dealt with between the different users.

Furthermore, even if crosstalk could really be avoided, there would be no guarantee that a listener would not be disturbed by his neighbors' soundscapes. This solution is not adapted to a multi-user environment.

While binaural techniques should be chosen as a standard for 3D audio in cinema, transaural techniques would probably be the right choice for the corresponding 3D audio at home. This can be compared with the fact that the 3D-stereo reproduction in a movie theater is almost necessarily based on glasses, while manufacturers are starting to offer autostereoscopic HDTVs for home cinema applications.

4.4. Wave Field Synthesis

For full 3D reproduction, Wave Field Synthesis (WFS) requires that the volume be entirely surrounded by loudspeakers. To achieve the highest aliasing frequency possible, the secondary sources need to be placed extremely close to each other. This calls for a significant number of very small loudspeakers along with high computing power to generate the corresponding signals. Because of this, WFS would probably have to be limited today to a 2.5D sound reproduction system using a linear array of loudspeakers with the sound sources in the horizontal plane.

A good match can be achieved between visual and auditory volumes. Indeed, virtual sound sources are not limited to the volume outside the loudspeaker array. It is possible to generate concave wave fronts that focus inside the playback area [25]. Focus however can only occur when there is no obstacle between the array and the desired position of the source. Hence the area in a theater between the screen and the first row can be considered a part of the whole sound volume and sound sources can appear both in front and behind the screen.

WFS has already been successfully used with 2D video content on a large screen [26]. There is also a problem of localization when two viewers look at the same 2D image from two different seats: their perspective is different. As a soundscape personalization is

not possible with WFS, the actual depth of the virtual sound source is chosen according to an ideal viewpoint (A in Fig. 5). From another seat, for example at B , the viewer expects the sound to come along the direction between himself and the image. As he is not sitting at the ideal viewpoint, there is an angle of reproduction error (γ in Fig. 5). The value of this angle is computed for each seat in the theater room and the maximal spatial depth of sound sources is limited to guarantee that the sound always appears to come from the visual object.

With 3D-stereo video however, depth is defined by the content itself and should not be limited for technical reasons. Therefore, WFS is a good choice but not the best, since binaural techniques do not suffer the limitation described here.

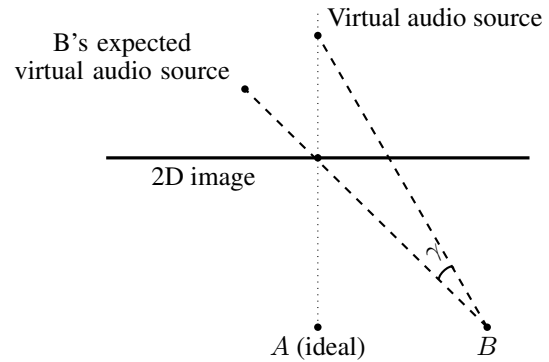


Figure 5. Visual localization with a 2D video content (after [26]).

4.5. Ambisonics

As stated earlier, Ambisonics is independent of the reproduction system: either headphones or loudspeakers can be used for playback.

For cinema applications, however, directly recording in Ambisonics format is intractable because the sound field on a set is not entirely controlled. Indeed there are numerous sources of noise such as the equipment, the crew, and the environment.

The Fourier-Bessel series expansion is nonetheless interesting for its properties as a coding system. It allows one to decouple the recording and reproduction processes. Any sound field that includes the localization cues can be expressed in the Fourier-Bessel series formalism. The reproduction system can then be chosen, depending only on the order of the decomposition that is used.

Since the reproduction through headphones is the best choice for 3D-stereo movies, the reproduction using Ambisonics coding would be based on

$$L \geq (l + 1)^2 \quad (17)$$

virtual loudspeakers at order l . Only the HRTFs corresponding to these virtual loudspeaker positions would

be needed because the original signal already contains the spatial information. Furthermore, in this formalism, a rotation of the soundscape is achieved by a simple linear combination of the original signal components. Therefore, the virtual loudspeakers can be fixed at all times. This drastically reduces the processing at the listener end of the reproduction chain.

5. Conclusions

This paper describes various 3D audio techniques capable of reproducing the correct sensation of direction and distance to the sound source as well as the room effect. All these techniques have been reviewed to address the problem of adding 3D sound to 3D-stereo movies.

The match between the sound volumes and the image volumes, together with the ability to personalize the sound field in a highly multi-user environment, were shown to be the most important elements of the problem considered here.

The reviewed 3D audio techniques were Vector Base Amplitude Panning, binaural techniques using headphones and loudspeakers, Wave Field Synthesis, and Ambisonics. Among these, only binaural techniques through headphones offer the technical possibility to tackle all the challenges of 3D-stereo movies.

In addition, the computing cost associated with the use of HRTFs and with head-tracking could be mitigated using the Fourier-Bessel formalism, which exhibits interesting mathematical properties.

Future research on this topic should aim at evaluating the audience localization performance in the case of 3D-stereo projection and 3D audio reproduction using the suggested solution of Ambisonics-based binaural audio. Further work could be carried out on a hybrid solution consisting of a combination of an augmented-reality audio system and loudspeakers.

References

- [1] C. Kyriakakis, P. Tsakalides, and T. Holman, "Surrounded by sound," *IEEE Signal Process. Mag.*, vol. 16, no. 1, pp. 55–66, 1999.
- [2] C. André, "Adding sound to movies: historical trends and new challenges of 3D," in *3D Stereo MEDIA Conf.*, Liège, Belgium, Dec. 2009.
- [3] T. Lentz, D. Schröder, M. Vorländer, and I. Assenmacher, "Virtual reality system with integrated sound field simulation and reproduction," *EURASIP J. on Advances in Signal Process.*, vol. 2007, no. 1, pp. 187–187, 2007.
- [4] X. Amatriain, J. Kuchera-Morin, T. Hollerer, and S. T. Pope, "The AlloSphere: immersive multimedia for scientific discovery and artistic exploration," *IEEE Multimedia*, vol. 16, no. 2, pp. 64–75, 2009.
- [5] R. Streicher and F. A. Everest, *The new stereo soundbook*, 3rd ed. Pasadena: Audio Engineering Associates, 2006.
- [6] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, Jun. 1997.
- [7] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?" *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, Jun. 1996.
- [8] J. Blauert, *Communication acoustics*. Springer, Jul. 2005.
- [9] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer characteristics of headphones measured on human ears," *J. Audio Eng. Soc.*, vol. 43, no. 4, pp. 203–217, Apr. 1995.
- [10] T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *J. Audio Eng. Soc.*, vol. 54, no. 4, pp. 283–294, Apr. 2006.
- [11] W. G. Gardner, "3-D audio using loudspeakers," Ph.D. dissertation, MIT, Boston, MA, Sep. 1997.
- [12] M. Vorländer, *Auralization*. Springer, 2008.
- [13] A. J. Berkhout, "A holographic approach to acoustic control," *J. Audio Eng. Soc.*, vol. 36, no. 12, pp. 977–995, Dec. 1988.
- [14] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," vol. 93, no. 5, pp. 2764–2778, May 1993.
- [15] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, Nov. 1985.
- [16] A. Laborie, R. Bruno, and S. Montoya, "A new comprehensive approach of surround sound recording," in *114th Conv. Audio Eng. Soc.*, Amsterdam, The Netherlands, Mar. 2003, preprint 5717.
- [17] K. Farrar, "Soundfield microphone," *Wireless World*, pp. 48–50, Oct. 1979.
- [18] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3D Ambisonic based binaural sound reproduction system," in *24th Int. Conf. Audio Eng. Soc.*, Jun. 2003.
- [19] B. Mendiburu, *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, Apr. 2009.
- [20] P. Burt and B. Julesz, "Modifications of the classical notion of panum's fusional area," *Perception*, vol. 9, no. 6, pp. 671–682, 1980.
- [21] I. Allen, "Matching the sound to the picture," in *9th Int. Conf. Audio Eng. Soc.*, Feb. 1991.
- [22] L. Verduci, "3D audio and 3D images for 3D movie theater," in *3D Stereo MEDIA Conf.*, Liège, Belgium, Dec. 2009.
- [23] J. Borwick, Ed., *Sound recording practice*, 4th ed. Oxford: Oxford University Press, May 1996.
- [24] A. Martin, C. Jin, and A. van Schaik, "Psychoacoustic evaluation of systems for delivering spatialized augmented-reality audio," *J. Audio Eng. Soc.*, vol. 57, no. 12, pp. 1016–1027, Dec. 2009.
- [25] E. W. Start, "Direct sound enhancement by wave field synthesis," Ph.D. dissertation, TU Delft, The Netherlands, Jun. 1997.
- [26] F. Melchior, S. Brix, T. Sporer, T. Röder, and B. Klehs, "Wave field synthesis in combination with 2D video projection," in *24th Int. Conf. Audio Eng. Soc.*, Jun. 2003.