Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis

Jean-Marie Pernaux (1)
Patrick Boussard (2)
Jean-Marc Jot (3)
(1) and (2) Steria/Digilog S.A, Aix-en-Provence France
pernaux@ircam.fr patrick.boussard@steria.fr
(3) IRCAM, Paris France
jmj@ircam.fr

Abstract

The aim of this article is to compare two multi-speaker spatialization techniques - Vector Base Panning (VBP) and Ambisonics - on a particular loudspeaker layout. Theory is presented for configurations in the horizontal plane. A way to deal with the elevation effect for such layouts is proposed. A new "local" panning method is introduced. VBP and Ambisonics are compared on a 5.1 configuration by means of objective simulation and preliminary listening tests on the real-time DSP system Genesis. This study could find recording and mixing applications in home cinema and multimedia.

1 Introduction

In the past years, many spatialization systems have been designed. The aim of these systems is to reproduce at the listener's ears various indices that bring information about the position of reproduced sound sources. Among all sound spatialization systems, we will discuss systems using several loudspeakers in the horizontal plane and most particularly the 5.1 format loudspeaker configuration [1]. This five loudspeaker configuration became popular in motion picture theaters and was accepted as a standard for HDTV. The principle of these systems is to give the listener the illusion of a virtual sound source placed at a desired position, by feeding the loudspeakers with source signal, multiplied by particular gain factors.

The aim of this paper is to compare two multi-speaker techniques, the Vector Base Amplitude Panning (VBAP) introduced by V. Pulkki [2] and the Ambisonic technique developed M. A. Gerzon in [3] and [4].

In a first part, the theoretical aspects and the connections between the two techniques are reviewed, for loudspeaker configurations in the horizontal plane. Extensions of these techniques are described for dealing with the elevation variable over 2-D layouts. The connections between the two approaches suggest a new "Vector Base Intensity Panning" (VBIP) method, derived from Gerzon's high frequency localization model. VBAP and VBIP techniques form a frequency-dependent Vector Base Panning method (VBP). Hence, VBP and Ambisonics are applied to the 5.1 format loudspeaker configuration. Objective comparison is performed by calculating localization criteria and subjective comparison is made through preliminary listening tests using an implementation on the real-time DSP system Genesis.

2 Theoretical bases

We will consider here a *N* loudspeaker layout placed in the horizontal plane at the same distance from the listener. In a coordinate system were the *x* axis is pointing forward and the *y* axis pointing to the left, let θ_i be the azimuths of the N loudspeakers with. $\{\theta_i\}_{i=1..N}$. We will first present the encoding equations, which consist in storing significant values of a sound field in a various number of channels [5] and are quite similar for VBP and Ambisonics. Then, we will discuss the decoding process which attempts to optimize the diffusion of these indices over a particular speaker layout, taking into account the different mechanisms of our perception of localization at low and high frequencies according to [3].

2.1 Encoding

VBP and Ambisonic encoding are very close. In the Ambisonic system, the B-format encoding is the most widely used [6]. This format allows to represent sounds situated in the horizontal plane with 3 signals *W*, *X* and *Y*. These signals can be obtained by recording with a « SoundField » microphone or synthesized by signal processing. For a plane wave of amplitude P and incidence θ, B format encoding equations can be written :

$$W = P$$

$$X = \sqrt{2}.P.\cos\theta$$

$$Y = \sqrt{2}.\,P.\sin\theta$$

The factor $\sqrt{2}$ is used so that all signals have equal energy for typical sound fields. More generally, it can be shown that Ambisonic encoding is the result of the sound field projection over a base of space functions called spherical harmonics [5] and [7]. B format is an approximation of the sound field with spherical harmonics of order 0 (*W)* and 1 (*X* et *Y*).

In the VBP technique, the encoding is made using a position vector $p = [p_1 \;\; p_2]^T = [\cos(\theta) \;\; \sin(\theta)]^T$.

One can see that, except for the factor $\sqrt{2}$, this is similar to B-format encoding. Another difference is that the *W* signal is not used. As in B-format, the sound field is approximated over 1[st] order spherical harmonics.

2.2 Decoding

It consists in finding the *N* loudspeaker gains g_i to satisfy constraints on virtual source position and localization quality [3]. One tries to optimize criteria taking into account two aspects of our perception of localization depending on the frequency of the source signal. We will consider four criteria taken from the Velocity model (below 700 Hz) and from the Energy model (for higher frequencies) [3]:

- θ_V low-frequency perceived virtual source position.

- r_V velocity vector amplitude . It is a quality index.

-of the virtual images at low frequencies. It must equal 1 for optimum quality.

- θ_E high-frequency perceived virtual source position.

- r_E energy vector amplitude. It is a quality criterion of the virtual images at high frequencies. It must be as close to 1 as possible for optimum quality.

The decoder produces the loudspeaker gains g_i for every encoding azimuth θ under various constraints [3]:

- low frequency constraints (< 700 Hz):

(A) sum of velocity vectors of the *N* loudspeakers points towards the desired direction ($\theta_V = \theta$)

(B) amplitude of this sum vector equals 1 ($r_V = 1$)

- high frequency constraints (> 700 Hz):

(C) sum of intensity vectors of the *N* loudspeakers points towards the desired direction ($\theta_E = \theta$)

(D) amplitude of this sum vector is as close to 1 as possible (r_E maximum)

- energy preservation constraints so that the volume of the virtual source is constant for every θ (E) sum of squared gains equals 1.

2.2.1 Ambisonics

Ambisonics could be considered as a "global" panning method because every loudspeaker is fed whatever the position of the virtual source. The design of an Ambisonic decoder (associated with B-format) is to find a 3 to *N* channel distribution matrix under the above constraints. As the constraints vary with frequency, two different decoders are defined: a low frequency one and a high frequency one. The signal processing includes low pass filters before the LF matrix and high pass filters before the HF matrix [8]. Low frequency decoder design means finding the scattering matrix to verify constraints (A) and (B). This matrix then satisfies $\theta_V = \theta$ and $r_V = 1$. For the high frequency decoder design, we search the 3×*N* coefficients of the matrix to verify conditions (C) and (D). Then the HF matrix satisfies $\theta_E = \theta$ and maximizes r_E . These characteristics aim at achieving a perceived direction equal to the encoding direction and the best localization quality (for low and high frequencies).

2.2.2 Vector Base Panning

Unlike Ambisonics, VBP decoding is a "local" panning technique, where the decoder matrix depends on the panning position: for each position of the virtual source, a pair of loudspeakers (a base) is selected and these two loudspeakers are fed with the source signal scaled by gain factors. The gains of all the other loudspeakers are set to zero. The process of selecting the loudspeaker base is detailed in [2]. The VBAP technique is shown below to satisfy the low frequency localization model. To obtain a general Vector Base Panning formalism, we will introduce a high frequency local panning method called Vector Base Intensity Panning (VBIP). Base selection being the same for VBAP and VBIP, we will only present the gains calculation, depending on frequency range.

Vector Base Amplitude Panning (< 700 Hz)

Let θ_1 and θ_2 be the azimuths of the two selected loudspeakers. Assuming $G_i = g_i / r_V$ we can write condition (A):

$$\begin{pmatrix} \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

It is then possible to compute the non normalized gains by [2]: $\begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \cos\theta_2 \end{pmatrix}^{-1} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$.

Writing condition (E) ${r_V}^2 \sum_{i=1}^{2} G_i^2 = 1$, we obtain the loudspeaker gains by [2]: $g_i\big|_{VBAP} = G_i \,/\, \sqrt{\sum_{i=1}^{2} G_i^2}$.

Vector Base Intensity Panning (> 700 Hz)

Here, we try to satisfy conditions (C) and (E).

Taking $G_i = g_i^2 / r_E$, the expression for the non normalized power gains is the same as for VBAP. Writing condition (E) ${r_E}^2 \sum_{i=1}^{2} G_i^2 = 1$, we obtain the loudspeaker gains by: $g_i\big|_{VBIP} = \sqrt{G_i \,/\, \sum_{i=1}^{2} G_i}$.

2.3 Dealing with the elevation variable over 2-D horizontal layouts

To allow for virtual source positions upwards from the listener (which is not theoretically possible with a layout in the horizontal plane), we describe for each technique possible panning strategies integrating the elevation variable over 2-D reproduction layouts.

2.3.1 Ambisonics

It is possible to modify the 3-D B format encoding equations [3, 6] to include the energy of the elevation channel (*Z*) into the *W* channel. With φ the elevation of the virtual source, the equations become:

$$as\ Z = \sqrt{2}.P.\sin\varphi$$

$$\Rightarrow \begin{cases} W = \sqrt{W^2 + Z^2} = P.\sqrt{1 + 2\sin^2\varphi} \\ X = \sqrt{2}.P.\cos\theta.\cos\varphi \\ Y = \sqrt{2}.P.\sin\theta.\cos\varphi \end{cases}$$

For $\varphi = 90°$ (virtual source at zenith), all speakers are fed, giving the impression of a source placed above the listener.

2.3.2 Vector Base Panning

A first solution would be to assume elevated loudspeakers (say at $\varphi_0 = 20°$ for instance) and to implement the 3-D VBP equations as described in [2]. However, since only 3 loudspeakers radiate at any time, virtual sound images then tend to be "jumpy" for interior positions. An alternative approach is to consider a virtual loudspeaker situated at the zenith position ($\varphi_0 = 90°$) with the 3-D VBP method. To simulate this virtual speaker, we can distribute its gain factor $G_{90°}$ over the *N* real loudspeakers, before normalizing the gains (condition (E)). This technique provides smoother transitions between elevated positions.

3 Application to a 5 loudspeaker layout (5.1 format)

Ambisonics and Vector Base Panning techniques were compared in the case of a 5 loudspeaker layout in the 5.1 format (see *Figure 1*).

Φ_F was chosen to be 45° and Φ_B was fixed at 50° (see [8] for the Ambisonic decoder coefficients to be used with such values).

For this particular layout, four panning laws were compared:

- low frequency Ambisonics, satisfying $\theta_V = \theta_E = \theta$ and $r_V = 1$ for frequencies below 700 Hz

- high frequency Ambisonics, satisfying $\theta_V = \theta_E = \theta$ and maximizing r_E for frequencies above 700 Hz

- VBAP (< 700 Hz) and VBIP (> 700 Hz)

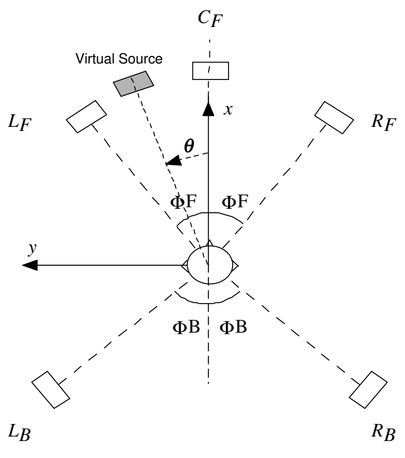

Figure 1. 5.1 Format loudspeaker layout

3.1 Objective comparison of the two techniques by means of psychoacoustical localization criteria

The four localization criteria ($\theta_V, r_V, \theta_E, r_E$) can be derived from each of the above panpot laws and are plotted against the desired azimuth of the virtual source θ in *Figure 2* and *3*.

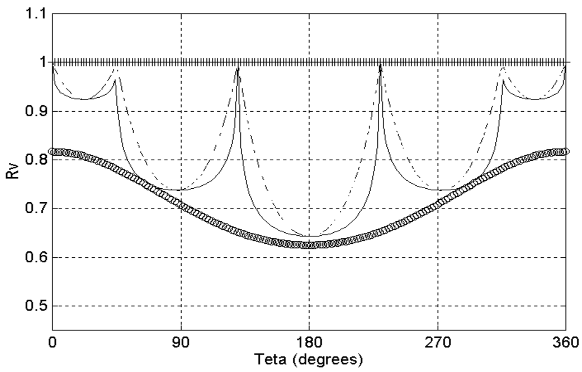

Figure 2 Low frequency localization quality (r_V)
'+': LF Ambisonics - 'o': HF Ambisonics dashed: VBAP - solid: VBIP

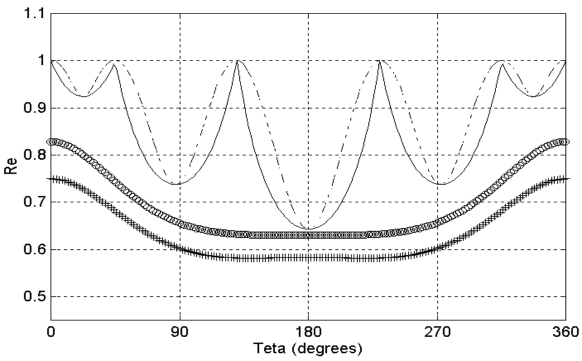

Figure 3 High frequency localization quality (r_E)
Same legend as Figure 3

Four main observations can be made: a) VBAP guarantees $\theta_V = \theta$, while VBIP ensures $\theta_E = \theta$; b) Frontal localization is better for all four techniques, which may be of importance for applications such as HDTV; c) At low frequencies, the Ambisonic decoder is the only one to achieve optimal localization for all θ; d) At high frequencies, VBP techniques yield better results than Ambisonics for all θ. Focusing on r_E rather than r_V seems preferable (see [3]), which gives a noticeable advantage to VBP techniques for the positioning of fixed sources. However, the strong variations of r_E and r_V for VBP might recommend Ambisonics for the simulation of a moving source.

3.2 Subjective comparison of both techniques in preliminary listening tests

Informal subjective tests were carried out on 7 subjects with an implementation of both techniques on the real-time digital signal processing system GENESIS developed by Steria/Digilog, which provides up to 16 analog outputs in its base configuration. For each of the four panning methods (LF and HF Ambisonics, VBAP and VBIP), two stimuli were used: a low-frequency stimulus (bass drum) and a mid/high-frequency stimulus (snare drum). The virtual source was positioned at a given location (0°-22,5°-90°-180°) and listeners were asked to choose the method allowing the easiest localization. They could switch in real-time from one method to the other at will. The same was asked for the source rotating around the listeners.

The answers given confirm the expectations of section 3.1: VBP techniques allow a more accurate localization of fixed sources, while Ambisonics is preferred in the case of moving sources since the loudspeakers seem "less present".

4 Perspectives

So as to extend the comparison of Ambisonics and VBP to other 5.1 layouts, it is necessary to design Ambisonic decoders for other values of Φ_F and Φ_B. The design of such decoders according to [8] is relatively complex. It would also be desirable to optimize the base selection algorithm used in VBP [2] to make VBP as computationally efficient as Ambisonics. Lastly, in order to compare both techniques more accurately and to evaluate the localization criteria, more extensive listening tests must be carried out, allowing a more quantitative statistical analysis.

5 Conclusion

The simulations and the implementation on the Genesis system allowed evaluating and to comparing the performances of two techniques for positioning sound sources in the horizontal plane on a particular 5-loudspeaker layout according to the 5.1 format. A variation of V. Pulkki's VBAP technique was introduced, designed to satisfy high-frequency intensity-based localization criteria and thus termed VBIP. Whereas Ambisonics decoder design is quite tedious for irregular configurations such as the 5.1 format, it is easy to apply VBP techniques to any layout.

Objective comparison and preliminary listening tests indicate that VBP gives better results for fixed sources. Although Ambisonics seems to be preferred for moving sources, VBP appears as an interesting alternative to Ambisonic for non regular loudspeakers layouts, such as 5.1.

References

[1] G. Steinke, "Surround Sound - The Next Phase : An Overview", presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 651 (1996 July/Aug.), preprint 4286.

[2] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", *J. Audio Eng. Soc*., vol. 45, no. 6, pp. 456-466 (1997 June).

[3] M.A. Gerzon, "General Metatheory of Auditory Localisation", presented at the 92nd Convention of the Audio Engineering Society (Vienna, 1992), preprint 3306.

[4] D. G. Malham and A. Myatt, " 3-D Sound Spatialization using Ambisonic Techniques", *Computer Music J*., vol. 19 no. 4, pp. 58-70 (1995).

[5] M.A. Gerzon, "Periphony: With-Eight Sound Reproduction", *J. Audio Eng. Soc*., vol. 21, no. 1, pp. 2-10 (1973 January/February).

[6] R.K. Furness, "Ambisonics - An Overview", presented at the 8th International Conference of the Audio Engineering Society.

[7] J. S. Bamford, "Ambisonic Sound For Us", presented at the 99th Convention of the Audio Engineering Society (New York, 1995).

[8] M.A. Gerzon, "Ambisonic Decoders for HDTV", presented at the 92nd Convention of the Audio Engineering Society (Vienna, 1992), preprint 3345.