Creating Auditory Displays with Multiple Loudspeakers Using VBAP: A Case Study with DIVA Project

Ville Pulkki
Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology
P.O. Box 3000, FIN-02015 HUT, Helsinki, Finland, Ville.Pulkki@hut.fi

Tapio Lokki
Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Helsinki, Finland, Tapio.Lokki@hut.fi

Abstract

Auditory displays containing an arbitrary number of loudspeakers in any positioning are created. The sound signals are positioned to the display using Vector Base Amplitude Panning (VBAP). The VBAP method is reviewed and a new automatic loudspeaker setup division routine is presented. The VBAP is used to create auditory display to the Digital Interactive Virtual Acoustics (DIVA) virtual environment system.

1 Introduction

Auditory displays that produce immersive listening experiences have been investigated lately. Often the display is created using HRTF-processed sound in a pair of loudspeakers or headphones. The greatest advantage of HRTF processing is that a full 3-D soundscape can be produced using only two reproduction channels. Some drawbacks make the system however unpractical in certain situations. In loudspeaker listening the spatial effect collapses outside of the listening area, and the listening conditions have to be ideal. In headphone listening the performance may be spoilt if the used HRTF functions are not personal and a head tracker is not used.

The use of multiple loudspeakers around the listener solves some of these problems. The listening area is larger, and the listening perception does not collapse totally outside the optimal listening area. Also the head movements do not ruin the experience. The drawback in loudspeaker reproduction is, of course, the greter amount of loudspeakers.

However, for long time the only generic method to position virtual sources to any number of loudspeakers was the periphonic Ambisonics [1], which is a matrixing system for audio recording and reproduction. The Ambisonics system is not ideal in all cases, its performance is at its best with limited number of loudspeakers in certain positioning.

Recently published Vector Base Amplitude Panning (VBAP) is a new approach to multi-channel 3-D panning problem [2]. The VBAP enables positioning of virtual sources with any number of arbitrarily placed loudspeakers. When the number of loudspeakers is increased, the virtual sources get more precisely localized and the colorizations of virtual sources decrease. This is due to the fact that when the sound signals are panned with VBAP, maximally three adjacent loudspeakers emanate the sound signal.

In this paper we review the theory of VBAP and propose implementations for system initialization routines and runtime computation. We describe also the use of VBAP in DIVA virtual audio reality system and report results of some informal listening tests.

2 Vector Base Amplitude Panning

Amplitude Panning. In the simple amplitude panning method two loudspeakers radiate coherent signals, which may have different amplitudes. The listener perceives a virtual sound source [3]. The relation of amplitudes of emanating signals controls the perceived direction of the virtual source.

A typical loudspeaker configuration is illustrated in Fig. 1a. Two loudspeakers are positioned symmetrically with respect to the median plane. Amplitudes of the signals are controlled with gain factors of loudspeakers. The virtual

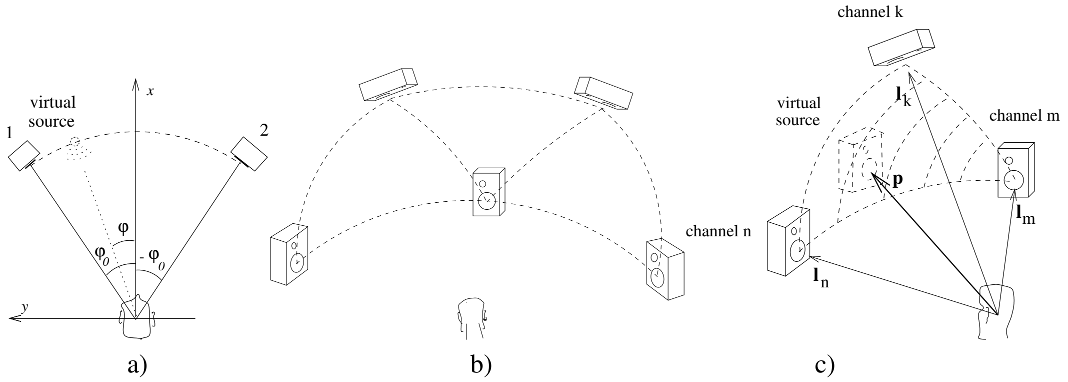

Figure 1: a) A standard stereophonic listening configuration. The virtual source can be positioned between the loudspeakers. b) A 3-D loudspeaker configuration is divided to loudspeaker triangles. c) The standard stereophonic listening configuration being extended with third loudspeaker. The listener perceives a virtual source that is located inside the triangle.

source can be placed on a two-dimensional sector between the loudspeakers. The relation between the ratio of gain factors and localized azimuth angle of the virtual source can be described with the stereophonic law of sines or tangents. Sine law is valid if the listener's head is pointing directly forward [4]. If the listener turns his/her head towards the virtual source, the tangent law is more correct [5].

The loudspeaker pair can also be in other directions than in front of the listener. There exists, however, some limitations in loudspeaker placement on sides. The loudspeakers should both be in front or in back of the listener, or the other loudspeaker directly to left or right from the listener [6]. If pairs that include both back and front loudspeakers are present, there exists a gap on directions to where the virtual sources cannot be positioned.

When multiple loudspeakers are applied to horizontal plane, a virtual source can be positioned to any direction on the plane using two adjacent loudspeakers surrounding the virtual source. This method is called pair-wise panning paradigm. With this method virtual sources are positioned to a plane on which the loudspeakers are situated. If there exists loudspeakers outside a plane, the pair-wise panning paradigm is insufficient. To formulate such cases, the loudspeaker setup can be divided to triangles (loudspeaker triplets), to one of which the signal is panned at one time, as in Fig 1b. There has, however, not existed a generic method for triplet-wise panning.

Vector Base Amplitude Panning. In VBAP a loudspeaker triplet is formulated with vectors as in Fig. 1c. The unit-length vectors $\mathbf{l}_m, \mathbf{l}_n$ and $\mathbf{l}_k$ point from listening position to the loudspeakers. The direction of the virtual source is presented with unit-length vector $\mathbf{p}$. Vector $\mathbf{p}$ is expressed as a linear weighted sum of the loudspeaker vectors

$$\mathbf{p} = g_m\mathbf{l}_m + g_n\mathbf{l}_n + g_k\mathbf{l}_k. \tag{1}$$

Here g_m, g_n, and g_k are called gain factors of respective loudspeakers. The gain factors can be solved as $\mathbf{g} = \mathbf{p}^T\mathbf{L}_{mnk}^{-1}$, where $\mathbf{g} = [g_m\ g_n\ g_k]^T$ and $\mathbf{L}_{mnk} = [\mathbf{l}_m\ \mathbf{l}_n\ \mathbf{l}_k]$. The calculated factors are used in amplitude panning as gain factors of the signals applied to respective loudspeakers after suitable normalization, e.g. $||\mathbf{g}|| = 1$.

When the number of loudspeakers is greater than three, the loudspeaker setup is divided to triangles forming a triangle set, as in Fig. 1b. During the panning process a single triangle from the set is chosen to be used in panning. The selection can be made by calculating the gain factors in each loudspeaker triangle in the triangle set and selecting the triangle that produced non-negative factors. If the triangles in the set are non-overlapping, the selection is unambiguous.

The division to triangles must be performed before the panning process is started. It cannot be performed as simply as in the 2-D case, where the adjacent loudspeakers form the loudspeaker pairs. In 3-D case there are many ways to divide the loudspeaker setup to triangles, however most of the triangles are not proper ones to be used in panning process. We present here a method which enables automatic division.

Dividing the Loudspeaker Setup to Triangles. The triangularization method includes four steps, which are presented in Fig. 2. In *step 1* triangles of all combinations of loudspeakers are formed. In *step 2* triangles that have

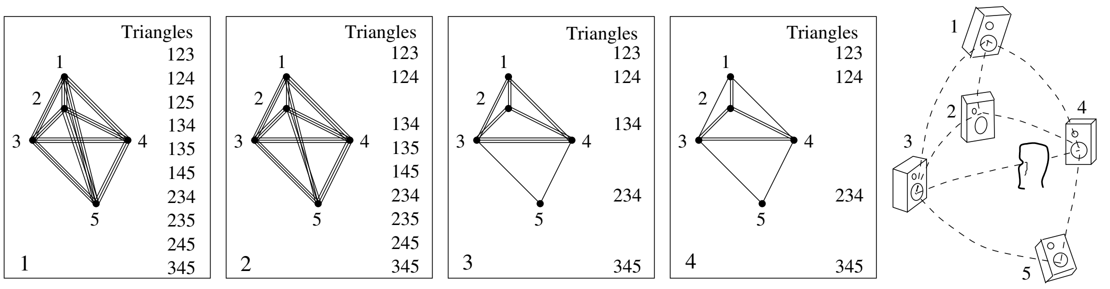

Figure 2: Initial triangularization process for five loudspeakers in 3-D setup. *Step 1*: All triangles are formed. *Step 2*: Too narrow triangles are removed. *Step 3*: Crossing sides are resolved. *Step 4*: Triangles that include loudspeakers are removed. The result of the process is seen in the rightmost figure.

small area when compared to total length of sides are deleted. In *step 3* all crossings of triangle sides are searched. When a crossing is found, the triangle that includes longer crossing side is removed. The crossing of two triangle sides (lines) is checked using simple vector calculation. Suppose that we want to check if lines between loudspeakers i and j, and between loudspeakers n and m cross, as in Fig. 3. The unit-length vectors $\mathbf{l}_i, \mathbf{l}_j, \mathbf{l}_n$, and $\mathbf{l}_m$, specify the directions of the loudspeakers. Both vector pairs $\{\mathbf{l}_i, \mathbf{l}_j\}$ and $\{\mathbf{l}_n, \mathbf{l}_m\}$, specify a plane. If the planes cross on the segment of line connecting the both loudspeaker pairs, the triangle sides cross. The two directions $\pm\mathbf{c}$ in where the planes cross is found using equation

$$\mathbf{c} = \pm\big((\mathbf{l}_i \times \mathbf{l}_j) \times (\mathbf{l}_n \times \mathbf{l}_m) \big), \tag{2}$$

where $\times$ denotes the vector cross product. If $\mathbf{c}$ or $-\mathbf{c}$ points to both triangle sides, the sides cross each other. This holds if $\angle(\mathbf{c}, \mathbf{l}_i) + \angle(\mathbf{c}, \mathbf{l}_j) = \angle(\mathbf{l}_i, \mathbf{l}_j)$ and $\angle(\mathbf{c}, \mathbf{l}_n) + \angle(\mathbf{c}, \mathbf{l}_m) = \angle(\mathbf{l}_n, \mathbf{l}_m)$ holds for either $\mathbf{c}$ or $-\mathbf{c}$. The operator $\angle$ denotes the smallest angle between specified two vectors.

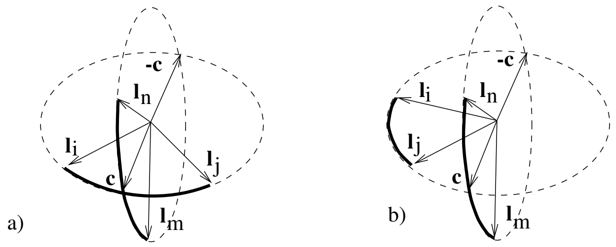

Figure 3: a) The line between vectors $\{\mathbf{l}_i, \mathbf{l}_j\}$ and between vectors $\{\mathbf{l}_n, \mathbf{l}_m\}$ cross, thus $\mathbf{c}$ points to both lines b) The lines do not cross, thus neither $\mathbf{c}$ nor $-\mathbf{c}$ points to both lines.

In *step 4* the triangles that include a loudspeaker are removed. For each triangle all loudspeakers are tested. The testing is performed by calculating gain factors to each loudspeaker direction vector using VBAP. The triangle must be deleted if all three gain factors of any loudspeaker are positive.

After these four steps a set of triangles is formed. The triangles are non-overlapping, and they have as equal-length sides as possible. The triangularization is performed only once, during initialization.

3 Digital Interactive Virtual Acoustics system with multi-channel sound reproduction

The Digital Interactive Virtual Acoustics (DIVA) research project has been introduced earlier in ICAD'96 and ICAD'97 [7, 8]. The auditory display part of the DIVA system consists of sound source and room acoustics modeling as well

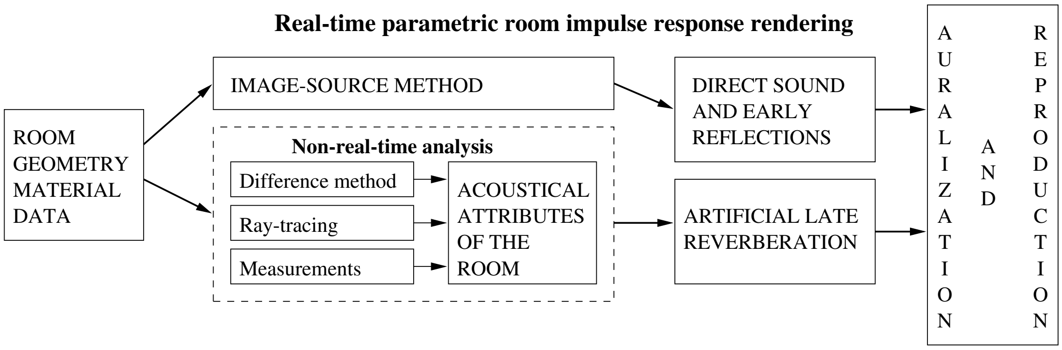

Figure 4: A parametric room impulse response rendering method used in DIVA system

as modeling of spatial hearing. With our system the natural sounding virtual auditory environment is created simultaneously with the visualization of virtual environment. The user can move interactively in a desired space and the visual and aural displays agree with movements. The room acoustics modeling is handled with *parametric room impulse response rendering* method (Fig. 4). This hybrid method models the direct sound and early reflections with image-source method [9] and diffuse late reverberation with a special case of feedback delay networks (FDN) [10].

We have added the multichannel reproduction by replacing the modeling of the spatial hearing (HRTFs) with the VBAP technique. The VBAP requires less computational capacity than HRTF filtering and dispenses with the measurement of individual HRTFs.

The image-source method replaces the sound propagation paths for direct sound and early reflections with image sources. Each image source is positioned to its direction with VBAP, image sources correspond then to virtual sources. The late reverberation algorithm we use consists of N feedback delay lines, where N is usually chosen according to the number of loudspeakers. In FDN like structure the outputs of delay lines are mutually incoherent and thus outputs of each delay line are connected to corresponding loudspeaker to render a diffuse soundfield [11].

When the VBAP system is initialized for the current listening setup, the directions of the loudspeakers are measured relative to the best listening position and loudspeaker triangles are formed automatically from adjacent loudspeakers. To speed up gain factor calculation during run time $\mathbf{L}_{mnk}^{-1}$ matrices are calculated for each triangle.

During run time each sample of each image source is multiplied with three gain factors and the results are added to audio stream going to the respective loudspeakers. If the image sources move (i.e. the listener or the sound source moves), the system performs also the following steps in an infinite loop: 1: new direction vectors $\mathbf{p}_{1,..,n}$ are defined according to the image-source calculation. 2: the right triangle for each image source is selected. 3: the new gain factors are calculated. 4: the old gain factors are cross faded to new ones and the loudspeaker triangles are changed if necessary.

To minimize the computational requirements of gain factor cross-fading, we use a two-level update algorithm. For each image source the direction vector $\mathbf{p}$ is updated at a rate of 20 Hz. When the image source is moving the cross-fading of gain factors has to be performed. However, the fading for every sample is unnecessary and computationally inefficient. In informal listening tests we have found that if the new gain factors are updated every 50 samples no disturbing clicks are perceived.

If the image source moves from a loudspeaker triangle to another, the loudspeaker(s) that are not existent in the target triangle are faded out, and simultaneously the new loudspeakers are faded in. Generally the loudspeaker gain factors that are faded out have small values, thus the fading in and out cannot be perceived.

4 Informal tests

Experiences with VBAP. In informal listening tests it has been observed that VBAP is able to create stable virtual sources. Even when the virtual sources are positioned in the middle point of a triangle, the localization is quite accurate.

The number and placing of the loudspeakers has a great effect on the localization accuracy. The lengths of the triangle sides should be as short as possible. This can be achieved by reasonable loudspeaker positioning and by increasing the number of loudspeakers. There exist also similar limitations on loudspeaker positioning as in pair-wise panning. The leftmost and rightmost directions should not be included in any loudspeaker triangle, since in such cases there will exist a hole where the virtual sources cannot be placed. This can be avoided similarly to the 2-D case by placing a loudspeaker to the rightmost and the leftmost direction.

Experiences with DIVA-VBAP. The DIVA system runs in Silicon Graphics (SGI) workstations. The tests were completed in a SGI Octane R5000 workstation, in which both the graphics and sound processing were performed in real time. The auditory display was produced with eight loudspeakers in the upper hemisphere, five loudspeakers in horizontal plane at directions $\pm 30°$, $\pm 90°$ and $180°$, and three at elevation $40°$ at azimuth angles $\pm 40°$ and $180°$.

In informal tests it was found that the system produced a natural sounding virtual auditory environment. In nonreverberant virtual environments the sound source could be localized distinctly in all directions. When the virtual room was reverberant, the direction of sound source could be perceived accurately near it, and inaccurately farther away, as in the real world.

The system has already been used to demonstrate architectural design by creating a simulated acoustic environment to the computer model of a concert hall. The “Marienkirche” demonstration film with a 5.1 channel surround sound track has been presented in Electronic Theater in SIGGRAPH’98 [12].

References

[1] Gerzon M. A. Periphony: With-height sound reproduction. Journal of the Audio Engineering Society, 1972; 21(1):2–10

[2] Pulkki V. Virtual source positioning using vector base amplitude panning. Journal of the Audio Engineering Society 1997; 45(6):456–466

[3] Blauert J. Spatial Hearing. The MIT Press, Cambridge, Massachusetts, 1997

[4] Blumlein A.D. U.K. Patent 394,325, 1931. Reprinted in Stereophonic Techniques, Audio Eng. Soc., NY, 1986

[5] Bernfeld B. Attempts for better understanding of the directional stereophonic listening mechanism. In: 44th Convention of the Audio Engineering Society, Rotterdam, Netherlands, February 1973

[6] Theile G and Plenge G. Localization of lateral phantom sources. Journal of the Audio Engineering Society, 1977; 25(4):196–200

[7] Huopaniemi J, Savioja L, and Takala T. DIVA virtual audio reality system. Proc. Int. Conf. Auditory Display (ICAD’96), Palo Alto, California, USA, Nov. 8-11, 1996, pp. 111-116.

[8] Savioja L, Huopaniemi J, Lokki T, and Väänänen R. Virtual environment simulation - Advances in the DIVA project. Proc. Int. Conf. Auditory Display (ICAD’97), Palo Alto, California, USA, Nov. 1997, pp. 43-46

[9] Borish J. Extension of the image model to arbitrary polyhedra. Journal of the Acoustical Society of America, Jun. 1984; 75(6):1827-1836

[10] Väänänen R, Välimäki V, Huopaniemi J and Karjalainen M. Efficient and parametric reverberator for room acoustics modeling. The Proc. of ICMC’97, Thessaloniki, Greece, Sep. 1997, pp. 200-203

[11] Gardner, W. G. Reverberation Algorithms. In: Applications of Digital Signal Processing to Audio and Acoustics. Kahrs, M. and Brandenburg, K. (Eds.) Kluwer Academic Publishers. pp. 85-131. 1998

[12] Takala T, Rousku E, Lokki T, Savioja L, Huopaniemi J, Väänänen R, Pulkki V, and Salminen P. Marienkirche - A visual and aural demonstration film. Shown in SIGGRAPH’98 Computer Animation Festival. Electronic Art and Animation Catalogue, 1998, pp. 149