A Comparative Analysis of Panning Techniques in Spatial Audio Applications

Midhun Palakkal Raghunath
Aalto Universtiy
Master's Programme CCIS / AAT

midhun.palakkalraghunath@aalto.fi

Abstract

Panning techniques serve the crucial role of positioning sound sources within stereo or multichannel audio configurations. This paper explores various techniques for spatial audio reproduction, including Vector Base Amplitude Panning (VBAP), Multiple Direction Amplitude Panning (MDAP), Distance Based Amplitude Panning (DBAP) and Volumetric Amplitude Panning. These techniques aim to enhance the localization and spatialization of sound sources in spatial audio environments. Amplitude panning methods such as VBAP and MDAP offer precise control over sound source positioning, leveraging vector-based mathematics to achieve directionally robust auditory event localization. DBAP introduces the consideration of distance between sound sources and listeners, improving realism and accuracy in three-dimensional space. Volumetric Amplitude Panning extends traditional methods by incorporating volume into the spatialization process, leading to smoother transitions and enhanced spatial width. Each technique offers distinct advantages in terms of precision, flexibility, and spatial accuracy, catering to diverse requirements in audio production and reproduction. Through this article, the focus will primarily be on elucidating the different types of techniques and comparing their respective characteristics.

Keywords — Amplitude panning, VBAP, virtual source

1 Introduction

In an era defined by rapid technological advancements, the realm of auditory experience is undergoing a profound transformation. The convergence of rapidly growing technology and the ever-increasing sophistication of audio devices has revolutionized the consumption of sound. Amidst this evolution, spatial audio emerges as a pioneering format poised to redefine how we perceive and engage with audio content. Spatial audio refers to a set of techniques and technologies used to create immersive auditory experiences by accurately reproducing the spatial characteristics of sound [1]. In essence, spatial audio aims to replicate how humans perceive sound in real-world environments, considering factors such as the direction, distance, and movement of sound sources. By leveraging innovative spatialization techniques, spatial audio promises to elevate the user's listening experience to unprecedented levels.

The acoustic environment is rich and complex, consisting of direct sounds and their reflections and refractions. Listeners process these sounds mentally, associating them with their sources. Traditional evaluation of virtual sound sources directional accuracy relies on listening tests, where participants verbally describe the perceived position. However, this method is time-consuming and lacks generalization. To address this, there's a growing need for an objective tool for spatial sound quality assessment, grounded in a deep understanding of spatial hearing mechanisms. In spatial hearing research, the perception of virtual sources is complex, as different directional cues may not always align. Understanding how humans prioritize these cues, particularly in conflicting scenarios, offers insights into spatial hearing. Additionally, studying elevation perception, especially with amplitude panning, unveils further nuances in spatial hearing mechanism [2].

Spatial perception is crucial for understanding our surroundings. Determining sound direction relies on two main cues: interaural level difference (ILD) and interaural time difference (ITD), which vary with frequency [3]. These cues result from differences in sound paths and shadowing by the listener's head. Humans also use spectral coloring to localize sound sources. Audio reproduction research aims to create immersive three-dimensional soundscapes. Early recordings were monophonic, but two-channel stereophonic reproduction expanded spatial depth. Horizontal-only (pantophonic) systems have been developed using different speaker configurations, often in a two-dimensional setup [4]. Efforts also focus on periphonic sound fields, employing three-dimensional speaker arrangements like holophony [5] or three-dimensional Ambisonics [6]. Periphonic sound fields can be created using two-channel speaker or headphone setups through digital models of head-related transfer functions (HRTFs) [7], which replicate the spectral characteristics of sound arriving at the listener's ear canal from various directions. However, this method is constrained by strict boundary conditions. While most systems have fixed speaker positions, Ambisonics systems offer flexibility in speaker number and placement. Orthogonal speaker placement maximizes localization accuracy, with little benefit from increasing speaker count.

The emergence of sound reproduction systems with multiple loudspeakers, both in theaters and domestic settings, signifies a shift towards immersive auditory experiences. However, reproducing sound with identical spatial attributes across different loudspeaker configurations remains a challenge. In multi-loudspeaker setups, a common technique is amplitude panning [8], where identical sound signals are distributed to several evenly spaced loudspeakers with varying amplitudes. This generates a virtual source perceived by the listener at a separate location from the actual sound sources. The goal is to accurately reproduce the direction and spread of the virtual source, but limitations in practice lead to imperfections. Some directions may not effectively produce virtual sources, and desired point-like sources may be perceived as spatially dispersed, diminishing the precision of the virtual source's directionality.

As a solution to this problem, different techniques for localization and positioning of virtual sources were introduced. Virtual sources can be located within speaker systems using a range of methods. In two-dimensional setups, panning is typically accomplished through pair-wise techniques [9]. For three-dimensional configurations, pair-wise panning can be expanded to triplet-wise approaches [10]. Generally, both pair-wise and triplet-wise panning offer satisfactory virtual source quality in large listening spaces. However, the quality of virtual sources may fluctuate based on the panning direction, as the number of speakers reproducing the same sound signal varies across directions. This variability can be seen as problematic, particularly when dealing with moving virtual sound sources.

In this manuscript, a comparative study on different types of panning technique available for spatial audio applications and their characteristics are reviewed. The different types of techniques discussed are Vector base Amplitude Panning (VBAP), Multiple Direction Amplitude Panning (MDAP), Direction Based Amplitude Panning (DBAP) and Volumetric Amplitude Panning (VAP).

1.1 Spatial Hearing and Spatial sound reproduction

Spatial hearing refers to the ability of humans to locate sounds in space and is based on the processing of various cues, including interaural level differences (ILDs) and interaural time differences (ITDs). The overview of spatial and directional hearing is provided in [3], focusing on loudspeaker listening in far-field environments. It introduces the duplex theory of sound localization, which identifies ITD and ILD as the primary cues for localizing sound sources. ITD is primarily utilized for frequencies below 1.5 kHz, representing a shift in carrier signals, while ILD is dominant above 1.5 kHz, indicating a shift in envelopes. The concept of the cone of confusion is introduced, defining it as a region where ITD and ILD remain constant regardless of sound source movement. Within this cone, there may be ambiguity in direction perception, mitigated by spectral cues and head rotation, which provide elevation and front-back information [11].

Spatial sound reproduction can be achieved through various methods, including direct-speaker playback, amplitude panning, and binaural synthesis. Direct-speaker playback involves playing back sounds through a spatially distributed array of loudspeakers, while amplitude panning involves changing the amplitude of the sound across the array to approximate the positions of sounds in a stereo or surround stage. Binaural synthesis is a technique that place the signal within a 360-degree field, and even change the perceived elevation, tilt, and more. Spatial sound reproduction can lead to perception discrepancies, such as front-to-back or back-to-front confusions, due to variations in the intended and perceived source locations. While both ITD and ILD cues are utilized across the audible spectrum, their relative importance remains uncertain. Wightman and Kistler's [12] proposition suggests that the auditory system prioritizes the most consistent cue when faced with distorted or conflicting information. A cue is deemed consistent if it consistently indicates the same direction across a broad frequency range.

2 Time Delay Panning and Amplitude Panning

Panning in spatial audio refers to the process of positioning sounds in the left to right spectrum of a stereo image or in 3D space. This is made possible by a pan pot or slider that allows adjusting the ratio of levels between the left and right channels. In spatial audio, panning can be used to place sound sources in 3D space relative to the listener and change their distance, azimuth and elevation angles, dimensions of a virtual room, the intensity of room reflections and reverb, and other properties. There are different types of panning such as amplitude panning, time delay panning, spectral panning, and phase panning, which can create complex stereo placement. Here we discuss about amplitude panning and time delay panning.

Time delay panning is a technique that uses time differences across the speakers, to create a more convincing stereo image. This is done by delaying the sound coming off one of the speakers between a small range, as little as 0 to 1ms. This technique has not been used much because it is quite easy to end up with some nasty phasing issues if the listener isn't in the right position. However, time-difference is one of the many factors used in binaural synthesis and panning.

Pulkki *et al.* [14] explore the use of time delay panning as a technique for creating virtual sound sources in spatial audio environments. Time delay panning involves introducing interaural time differences (ITDs) between the left and right channels to simulate the perception of directionality and spatial localization. The authors employ a binaural auditory model to analyze the perceptual attributes of virtual sound sources created using TDP. This model incorporates physiological and psychoacoustic factors to simulate how the human auditory system perceives spatial cues and directional information. Through their analysis, they evaluate various parameters of TDP, such as the magnitude and frequency range of the applied time delays, and their effects on perceived sound localization and timbre. They investigate how changes in these parameters influence the accuracy and realism of virtual sound source reproduction.

In amplitude panning [8] the same sound signal is applied to two or more loudspeakers equidistant from a listener with appropriate amplitudes. Amplitude panning is a technique used to approximate the positions of sounds in a stereo or surround stage by changing the amplitude of the sound across the array. This is the most common form of panning used in audio production. It is based on the principle that our ears and brain use a combination of factors to localise sounds, including inter-aural time difference, inter-aural level difference, reflections off the body, early reflections, reverberation, amplitude and more.

Amplitude panning, also known as intensity panning, and in this method, multiple loudspeakers are positioned at different directions and equal distances from the listener. Each loudspeaker receives the same sound signal, x(t), but with varying amplitudes represented by g_i for each channel i, where N is the total number of loudspeakers.

Mathematically, the signal sent to loudspeaker i is expressed as

$$x_i(t) = g_i x(t) \tag{1}$$

In the paper, [13] the authors explore the localization accuracy of virtual sound sources using amplitude panning techniques in stereophonic audio systems. They introduce novel algorithms tailored for stereophonic panning, aiming to improve spatial distribution and realism. Through theoretical discussions and subjective listening tests, the authors demonstrate the effectiveness of their methods in achieving precise sound localization and enhancing the immersive experience in stereophonic reproduction.

When these signals reach the listener's ears, they are summed at the ear canals, resulting in the perception of localization. This phenomenon is termed "summing localization." Thus, amplitude panning involves adjusting the amplitude of a sound signal sent to each loudspeaker to create the illusion of spatial distribution, allowing for the perception of sound sources originating from different directions in the listening environment.

2.1 Vector Base Amplitude Planning - VBAP

Vector Base Amplitude Panning (VBAP) is a method for positioning virtual sources to multiple loudspeakers. This approach was first introduced by [15] and further developed by Ville Pulkki [10] which is a robust and generic algorithm that activates the smallest possible number of loudspeakers to achieve directionally robust auditory event localization for virtual sound sources. This innovative technique of VBAP used for achieving precise virtual sound source localization in spatial audio environments [16].

VBAP, introduced as an alternative approach, relies on the principles of amplitude panning and vector-based mathematics to position virtual sound sources within a defined sound field. Unlike conventional methods that use amplitude differences between channels to simulate directionality, VBAP computes the amplitude gains for each channel based on the desired sound source position and the geometrical arrangement of loudspeakers. The number of loudspeakers can be varying and they can be placed in an arbitrary 2-D or 3-D positioning. It follows an existing panning law, tangent law, and enables use of any number of loudspeakers which can be positioned anywhere. The other loudspeakers can be used to produce first reflections and diffuse sound field. By calculating the weightings for each channel in real-time, VBAP ensures precise localization of sound sources, regardless of their position within the listening space [10].

The two-dimensional VBAP approach can be generalized as a three-dimensional VBAP method. In the standard two-channel stereophonic setup, a third speaker is introduced at the same distance from the listener as the existing ones, but it is not situated within the two-dimensional plane defined by the listener and the other two speakers.

This approach is referred as the three-dimensional amplitude panning, it involves positioning the virtual sound source within this triangle by driving the three speakers with synchronized electrical signals of varying amplitudes. Consequently, the virtual source can be situated on the surface of a three-dimensional sphere, where the sphere's radius is determined by the distance between the listener and the speakers. The specific area on the sphere's surface where the virtual source can be placed is known as the active triangle. This setup enables the virtual sound source to be perceived within a triangle formed by the speakers when observed from the listener's position and it is illustrated in the Fig. 1.

The relation of three gain factors defines the virtual source direction perceived by the listener. Eq. (2) can be generalized into a three-dimensional form as

$$ {g_1}^2 + {g_2}^2 + {g_3}^2 = C \tag{2} $$

For vector base formulation, the loudspeakers are placed on the surface of a unit sphere in three-dimensional space, equidistant from the listener. The three-dimensional unit vector $\mathbf{l_1} = [I_{11} l_{12} l_{13}]^T$ originating from the center of the sphere, specifies the positioning to the direction of loudspeaker 1. The unit vectors $\mathbf{l_1}, \mathbf{l_2}$, and $\mathbf{l_3}$ then define the directions of loudspeakers 1,2,and 3, respectively. The direction of the virtual sound source is defined as a three-dimensional unit vector $\mathbf{p} = [p_1 p_2 p_3]^T$.

The virtual source vector $\mathbf{p}$ is expressed as a linear combination of three loudspeaker vectors $\mathbf{l_1}, \mathbf{l_2}$, and $\mathbf{l_3}$, analogically to the two-dimensional case, and express it in matrix form,

$$ \mathbf{p} = g_1 \mathbf{l_1} + g_2 \mathbf{l_2} + g_3 \mathbf{l_3} \tag{3} $$

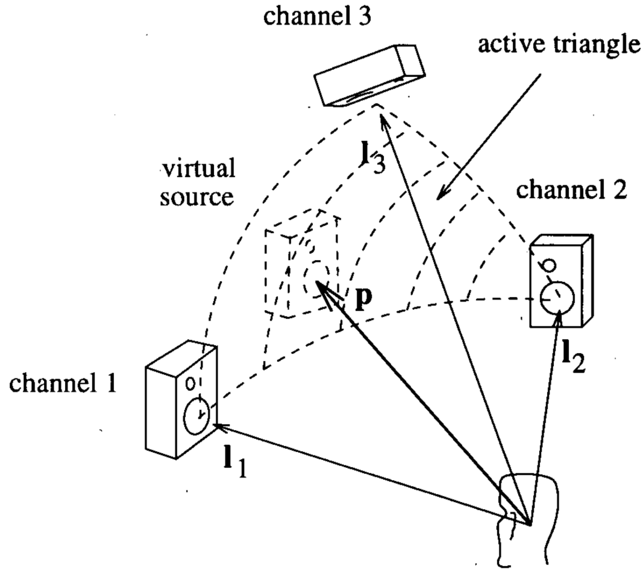

Figure 1: *Configuration for three-dimensional amplitude panning. Loudspeakers form a triangle into which the virtual source can be placed.(Taken from [10])*

$$\mathbf{p^T} = \mathbf{gL_{123}} \tag{4}$$

Here g_1, g_2, and g_3 are gain factors, $\mathbf{g} = [g_1 g_2 g_3]$, and $\mathbf{L_{123}} = [l_1 l_2 l_3]^T$. Vector g can be solved,

$$\mathbf{g} = \mathbf{p^T L_{123}^{-1}} = [p_1 p_2 p_3] \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \tag{5}$$

In VBAP, the virtual source cannot extend beyond the active arc, regardless of the listener's position. Localization errors increase with the size of the active region. To improve localization accuracy across a large listening area, more speakers are placed, especially around and behind screens in theaters. By reducing triangle sizes, varying loudspeaker-listener distances are accommodated. Since only three speakers receive signals simultaneously, only differences in their distances impact direction perception. These differences are corrected through time shifting and gain adjustments, enabling adaptable loudspeaker arrangements.

VBAP has three key properties: First, if the virtual source aligns with any loudspeaker direction, the signal emanates solely from that speaker, maximizing source sharpness. Second, if the virtual source lies on a line connecting two loudspeakers, only that pair receives the sound, following the tangent law. Third, if the virtual source is at the center of the active triangle, the loudspeaker gain factors are equal.

These properties ensure that VBAP generates virtual sound sources with maximum sharpness given current loudspeaker configurations.

VBAP addresses the limitations of conventional methods for sound localization in multichannel audio systems, highlighting issues such as spatial aliasing and the difficulty of accurately reproducing sound sources beyond the stereo field. The traditional methods often fail to provide convincing spatial cues, hindering the immersive experience for listeners. VBAP revolutionizes audio spatialization by enabling the use of numerous speakers arranged around the listener in any configuration. It requires equidistant speaker placement and low reverberation in the listening environment. VBAP simplifies amplitude panning equations through vector-based reformulation, enhancing computational efficiency.

2.2 Multiple Direction Amplitude Panning - MDAP

Multiple Direction Amplitude Panning (MDAP) is a modification proposed by Ville Pulkki that increases the number of activated loudspeakers to enhance direction-independence for auditory event localization of virtual sound sources [17]. This method aims to improve the perceived source width and reduce localization accuracy by activating more loudspeakers compared to traditional methods like Vector-Base Amplitude Panning (VBAP) and Ambisonics.

In the paper [17], the author addresses a critical aspect of spatial audio reproduction. In many conventional systems, virtual sound sources may not be evenly distributed across the listening area, leading to perceived imbalances or inconsistencies in the spatial image. The proposed methods is to achieve a uniform spreading of virtual sources using amplitude panning techniques. The amplitude ratios between loudspeakers are carefully manipulated to create a more consistent and immersive spatial audio experience. This research contributes to enhancing the realism and fidelity of spatial audio systems, particularly in applications such as virtual reality, gaming, and immersive multimedia. This work provides valuable insights and techniques for achieving more uniform spatial distribution in amplitude-panned virtual sources, improving the overall quality of spatial audio reproduction.

Here, gain factors are computed using multiple panning directions close to each other. These factors for each loudspeaker are combined and normalized to form unified gain values. Despite this, the listener still perceives a single virtual source. By adjusting the number and arrangement of panning directions in MDAP, the perceived spread of the virtual source can be controlled. When a sufficient range of panning directions is employed, the spread and coloration become independent of the panning direction, resulting in consistent spread and coloration in all directions. This method allows for the independent manipulation of the spread of virtual sources, and it can also simulate the spread of virtual sources in matrixing systems such as Ambisonics [18]. In 3-D loudspeaker setups, if two panning directions are

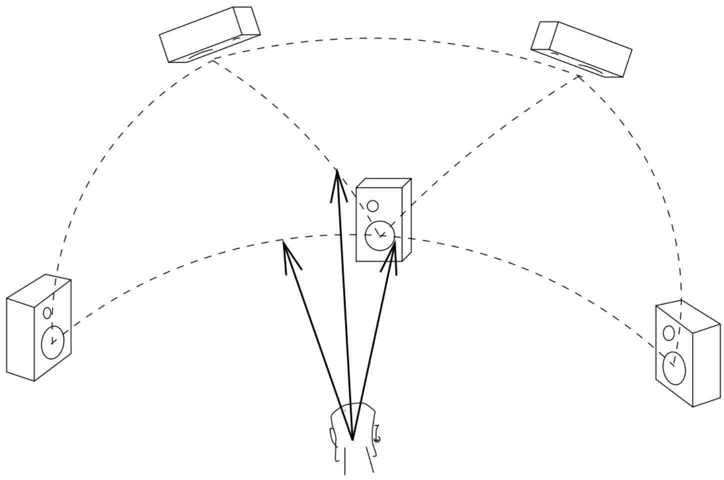

Figure 2: *Spreading the virtual source in 3-D loudspeaker setups using three panning directions. (Taken from [17])*

used in MDAP, the sound signal is applied at least to two loudspeakers at time. If the sound is desired to be applied at least to three loudspeakers at time, three panning directions are needed as is illustrated in Fig. 2.

In this configuration, three loudspeakers are activated for most panning directions, including those situated directly on a loudspeaker. Only the two adjacent loudspeakers are activated when the panning direction is precisely midway between two loudspeakers. In such special cases, MDAP produces identical loudspeaker gains as VBAP [19]. MDAP has shown to yield a reduced tendency towards the loudspeakers and provide a closer match to the ideal panning curve, enhancing the overall spatial perception of sound sources. The experimental results indicate that MDAP offers improved direction-independence and localization accuracy compared to other amplitude panning methods, making it a valuable technique for creating immersive and accurate spatial audio experiences. However, it comes at the cost of an increased perceived source width and reduced localization accuracy at off-center positions [16].

2.3 Distance Based Amplitude Panning - DBAP

DBAP was introduced by Lossius et. al [20] to enable flexibility of loudspeaker placement in artistic and scientific contexts. The algorithm allows for arbitrary loudspeaker locations in a 2D plane so that a virtual sound source may navigate the 2D space. Unlike traditional amplitude panning methods that solely rely on the angle between loudspeakers and the listener, DBAP introduces the consideration of distance between the virtual sound source and the listener. By taking into account both the angle and distance, DBAP aims to enhance the realism and accuracy of sound source localization in a three-dimensional space. This approach allows for more precise control over the perceived position of virtual sound sources, resulting in improved spatial fidelity and immersion in audio environments.

Consider a source placed in a Cartesian coordinate system at position P, where there are N loudspeakers placed at positions speaker S1, S2,... SN, each at a distance D1, D2,... DN from the intended source position [21]. Similar to stereo panning and VBAP, we normalize the gains on the speakers to ensure a balanced system with consistent energy.

$$\sqrt{\sum_{n=1}^{N} {g_n}^2} = 1 \tag{6}$$

The gain for each speaker is determined by assuming that it is inversely related to the distance between the speaker and the source position

$$g_n = c/D_n \tag{7}$$

Here, c may be dependent on the exact nature of the inverse distance law for sound propagation. But this is unimportant, since (6) and (7) may be combined to give,

$$g_n = \frac{1}{d_n\sqrt{\sum_{i=1}^{N} 1/{d_i}^2}} \tag{8}$$

This method allows for an unrestricted number of speakers, which can be arranged in any configuration. Similar to VBAP, where gain factors are determined based on speaker positions, in DBAP, we calculate gains using the distance alone instead of directional components of vectors. Unlike VBAP, the gains for each speaker remain unaffected by the listener's position; only the distances to the virtual sound source matter. However, when the listener's position is identifiable, additional enhancements become feasible. To ensure that sound from each speaker reaches the listener simultaneously, an appropriate delay needs to be incorporated into each speaker's output. Equation (9) can be utilized to compute the delay, represented as dn in samples, added to the output of speaker n.

$$d_n = (max(D_{L,1}, D_{L,2}, ...D_{L,N})f_s/v_s \tag{9}$$

where $D_{L,N}$ is the distance from speaker n to the listener's position, v_s is the speed of sound and f_s is the sampling frequency.

DBAP has several advantages over traditional methods of spatial audio reproduction [22]. It does not require the listener to be in a specific position, and can be used in a variety of speaker arrangements. It also allows for more flexibility in the placement of sound sources, and can be used to create more complex spatial sound fields. The idea of DBAP is universal and the method can be used both in 2D and 3D and makes it suitable for home movie or game systems. However, DBAP is not without its limitations. It requires accurate distance information between the sound source and each speaker, which can be difficult to obtain in some situations. It also requires careful calibration of the speaker gains to ensure that the sound field is reproduced accurately.

2.4 Volumetric Amplitude Panning

Volumetric Amplitude Panning extends traditional amplitude panning methods by incorporating the notion of volume into the spatialization process. The paper [23, 24, 25] introduces a novel algorithm for Volumetric Amplitude Panning (VAP) that addresses limitations in current 3D panning techniques such as VBAP and DBAP. These limitations include the reliance on a "sweet spot" listening position and specific loudspeaker layouts, which can degrade spatial accuracy and user experience. Unlike existing methods, the proposed VAP algorithm does not assume a sweet spot and can work with any symmetric or asymmetric loudspeaker arrangement. The suggested method provides precise panning capabilities even for irregular loudspeaker arrangements. To demonstrate the volumetric-based panning method, we examine two specific scenarios: panning across the wall and panning within the wall.

When panning a sound object across the wall within a 3D array of loudspeakers, it's crucial to preserve both the source direction and the overall loudness. This concept can be likened to envisioning the speakers positioned along the circumference of a circle (in 2D) or a sphere (in 3D), with a radius equal to the distance between each speaker and the listener. Seamless transitions between corners are essential to prevent any audible discontinuities. By effectively managing these transitions and boundary conditions, a normalized VBAP or MDAP approach can be employed to pan sound across any point on the walls. This is illustrated in the Fig. 3.

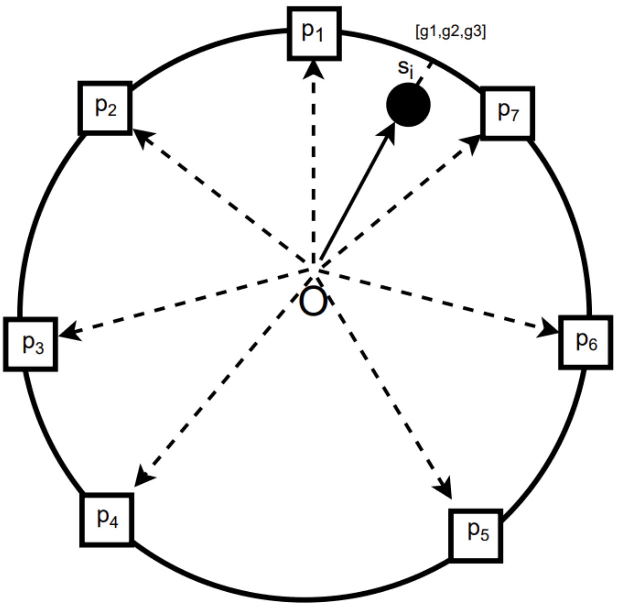

Figure 3: *Illustrating On the Wall and Inside the Wall panning for a circular/-spherical layout.(Taken from [24])*

Assuming g1, g2, and g3 represent the triplet weights calculated using VBAP/M-DAP, the scaled weights can be determined as sc * g1, sc * g2, and sc * g3, where sc can be computed as:

$$sc = \frac{1}{\sqrt{g1^2 + g2^2 + g3^2}} \tag{10}$$

These weights are then applied to distribute the total input signal energy among the speakers corresponding to g1, g2, and g3.

As the panning position shifts inward towards the listener from the circle or sphere encompassing the speakers, traditional VBAP becomes less perceptually accurate. To achieve accurate panning for a sound source within the walls while preserving overall loudness and source direction, it's necessary to maintain a constant total energy available for distribution. In most panning scenarios, this energy is assumed to be the energy of the mono/stereo input source signal, which needs to be distributed across the speaker grid.

The contribution of speakers for a source within the walls can be divided into two parts. Firstly, the VBAP contribution is calculated as if the source position were on the walls without considering distance. This VBAP contribution helps maintain the direction of the panned source within the walls. Secondly, the volumetric factor accounts for the distance-based effects. Assuming $E_{max}$is the sum of the maximum VBAP weights applicable to a speaker, the volumetric factor for a specific source position si is determined

$$E_{vol} = E_{max} - E_{VBAP} \tag{11}$$

where

$$E_{VBAP} = \sqrt{g1^2 + g2^2 + g3^2} \tag{12}$$

The weights g1, g2, and g3 are calculated for the source position si. Lastly,

$$E_{VBAP} = \sqrt{g1_{max}^2 + g2_{max}^2 + g3_{max}^2} \tag{13}$$

The weights $g1_{max}$, $g2_{max}$, and $g3_{max}$ are the maximum VBAP weights that can be applied to a source position. In most cases, this position will coincide with one of the speaker locations.

To consider the distance-based effects, the relative Euclidean distances from each speaker to the source are computed to determine each speaker's contribution to the total energy. The energy is then distributed to each loudspeaker based on the source's distance from them. The Euclidean distance between the source and each loudspeaker is calculated accordingly.

$$d(s_i, p_i) = \sqrt{(O - p_i)^2} \tag{14}$$

The Euclidean distance between the Center and any loudspeaker is given by :

$$d(O, p_i) = \sqrt{(s_i - p_i)^2} \tag{15}$$

The volumetric contribution of any loudspeaker for a source inside the walls is then given by :

$$E_{vol}p_i = E_{vol} * (\frac{d(O,p_i)}{D(O,p_i)}) \tag{16}$$

where $D(O, p_i)$ is:

$$D(O,p_i) = \sqrt{\sum_{i=1}^{N}(O-p_i)^2} \tag{17}$$

The total contribution of any loudspeaker for a source inside the walls is thus the sum of the volumetric ($E_{vol}pi$)and VBAP (E_{VBAP}) weights.

The approach of volumetric amplitude panning address the issue of abrupt transitions when transitioning from the speaker grid wall towards the listener position, while ensuring a consistent overall panning energy. Two methods of acoustic diffusion, one based on geometry and the other on distance, are introduced to minimize spectral fluctuations and improve spatial width. These techniques allow users to define and regulate the desired total diffusion energy for a given application, leading to more seamless volumetric amplitude panning across various directions and distances.

3 Conclusion

In conclusion, the exploration of various panning techniques and spatial audio reproduction methods presented in this article sheds light on the diverse strategies available for creating immersive auditory experiences. Panning techniques play a crucial role in positioning sound sources within stereo or multichannel audio configurations, allowing for precise control over the spatial placement of sound within the listening environment. The exploration of various spatial audio techniques such as Vector Base Amplitude Panning (VBAP), Multiple Direction Amplitude Panning (MDAP), and Distance Based Amplitude Panning (DBAP) offers valuable insights into the intricacies of sound localization and spatialization in immersive audio environments.

Amplitude panning techniques, including VBAP and MDAP, allow for precise control over the spatial distribution of sound sources, enhancing the realism and immersion of spatial audio reproduction. VBAP, in particular, leverages vector-based mathematics to accurately position virtual sound sources within a defined sound field, while MDAP increases the number of activated loudspeakers to improve direction-independence for auditory event localization. DBAP introduces the consideration of distance between the virtual sound source and the listener, enhancing the realism and accuracy of sound source localization in three-dimensional space. Volumetric

Amplitude Panning extends traditional amplitude panning methods by incorporating the notion of volume into the spatialization process, enabling more seamless transitions and improved spatial width.

In summary, these techniques provide valuable tools for creating immersive and accurate spatial audio experiences across various applications, including virtual reality, gaming, and multimedia. By addressing issues such as abrupt transitions, spectral fluctuations, and spatial width, these methods contribute to enhancing the realism, fidelity, and overall quality of spatial audio reproduction.

4 References

1. Zhang, W., Samarasinghe, P. N., Chen, H., Abhayapala, T. D. (2017). Surround by sound: A review of spatial audio recording and reproduction. Applied Sciences, 7(5), 532.

2. Pulkki, V. (2001). Spatial sound generation and perception by amplitude panning techniques. Helsinki University of Technology.

3. Blauert, J. (1997). Spatial hearing: the psychophysics of human sound localization. MIT press.

4. Steinke, G. (1996). Surround Sound-The New Phase An Overview. In 1996 AES Copenhagen workshop.

5. Condamines, R. (1972). LE RELIEF SONORE TETRAPHONIQUE.

6. Gerzon, M. A. (1973). Periphony: With-height sound reproduction. Journal of the audio engineering society, 21(1), 2-10.

7. Møller, H. (1992). Fundamentals of binaural technology. Applied acoustics, 36(3-4), 171-218.

8. Blumlein, A. D. (1986). UK Patent 394,325, 1931. Reprinted in Stereophonic Techniques, Audio Eng. Soc., NY.

9. Chowning, J. M. (1971). The simulation of moving sound sources. Journal of the audio engineering society, 19(1), 2-6.

10. Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. Journal of the audio engineering society, 45(6), 456-466.

11. Duda, R. O. (1997). Elevation dependence of the interaural transfer function. Binaural and spatial hearing in real and virtual environments, 49-75.

12. Wightman, F. L., Kistler, D. J. (1997). Factors affecting the relative salience of sound localization cues. Binaural and spatial hearing in real and virtual environments, 1, 1-23.

13. Pulkki, V., Karjalainen, M. (2001). Localization of amplitude-panned virtual sources I: stereophonic panning. Journal of the Audio Engineering Society, 49(9), 739-752.

14. Pulkki, V., Karjalainen, M., Huopaniemi, J. (1999). Analyzing virtual sound source attributes using a binaural auditory model. Journal of the Audio Engineering Society, 47(4), 203-217.

15. Pulkki, V., Huopaniemi, J., Huotilainen, T., Karjalainen, M. (1996). DSP approach to multichannel audio mixing. In International Computer Music Conference, ICMC'96, Hong Kong, August 19-24, 1996 (pp. 93-96).

16. Zotter, F., Frank, M., Zotter, F., Frank, M. (2019). Amplitude Panning Using Vector Bases. Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality, 41-52.

17. Pulkki, V. (1999, October). Uniform spreading of amplitude panned virtual sources. In Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452) (pp.187-190). IEEE.

18. Pulkki, V. (2000, August). Generic panning tools for MAX/MSP. In ICMC.

19. Frank, M. (2014). Localization using different amplitude-panning methods in the frontal horizontal plane.

20. Lossius, T., Baltazar, P., de la Hogue, T. (2009, August). DBAP–distancebased amplitude panning. In ICMC.

21. Kostadinov, D., Reiss, J. D., Mladenov, V. M. (2010, March). Evaluation of Distance Based Amplitude panning for spatial audio. In ICASSP (pp. 285-288).

22. Sundstrom, J. (2021, September). Speaker Placement Agnosticism: Improving the Distance-based Amplitude Panning Algorithm. In 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA) (pp. 1-6). IEEE.

23. Kareer, S., Sunder, K. (2022, October). Volumetric Amplitude Panning and Diffusion for Spatial Audio Production. In Audio Engineering Society Convention 153. Audio Engineering Society.

24. Kareer, S., Sunder, K. (2023, August). A Novel Algorithm for Volumetric Amplitude Panning and Diffusion in Spatial Audio Production. In Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio. Audio Engineering Society.

25. https://embody.co/blogs/technology/volumetric-amplitude-panning-and-diffusionfor-spatial-audio-production?currency=