Ville Pulkki and
Matti Karjalainen

Multichannel Audio Rendering Using Amplitude Panning

Spatial audio is a field that investigates techniques to reproduce spatial attributes of sound (e.g., direction, distance, width of sound sources, and room envelopment) to the listener. Such attributes cannot be reproduced accurately with one loudspeaker, therefore two-channel stereophony was introduced, later extended to different surround formats having five to eight loudspeakers. An even more accurate reproduction of spatial attributes can be achieved with loudspeaker setups often found in theaters and in some public venues, which contain a large number of loudspeakers around listeners, possibly also above and/or below them.

A basic question in spatial audio is how to position a sound source in a predefined direction in the virtual auditory space. An established technique, referred to as "amplitude panning," applies a sound signal with different amplitudes to different loudspeakers. In what follows, amplitude panning is first described in its traditional form, which is bound to two-dimensional (2-D) loudspeaker layouts. Next, a recent extension of amplitude panning to arbitrary three-dimensional (3-D) multichannel loudspeaker layouts is discussed.

SOUND LOCALIZATION

Some of the most interesting perceptual attributes of sound that are relevant to spatial audio are timbre and localization of sound objects [1].

Timbre is the attribute of sound that is formally defined as the difference in two sounds sharing the same loudness and pitch. In practice, the sound spectrum and its evolvement with time are the most prominent characteristics of timbre.

The sound localization mechanisms of humans have been studied using psychoacoustic tests, which found that humans decode four different cues from the ear canal signals, from which the sound source direction is obtained. In what follows, the median plane is the plane of symmetry separating the left and right sides of the listener. The interaural axis is the line passing through both ears. The cues are the following:

1) the interaural time difference (ITD) cue is sensitive to phase shifts in the frequency components of signals below 1.6 kHz and also to shifts in the envelope of signals at higher frequencies
2) the interaural level difference (ILD) cue is sensitive to level differences between ear canal signals; as a sound source moves from side to side, the ITD cue changes due to variations in travel paths most prominently at frequencies below 1.6 kHz, and the ILD changes due to shadowing by the head at higher frequencies
3) the monaural spectral cues
4) the effect of head rotations on the previous cues.

> A BASIC QUESTION IN SPATIAL AUDIO IS HOW TO POSITION A SOUND SOURCE IN A PREDEFINED DIRECTION IN THE VIRTUAL AUDITORY SPACE.

Because of the symmetry of the head, sounds originating from many different directions share the same ITD and ILD. The locus of all sound source locations that share the same ITD and ILD is called the "cone of confusion." Within a cone of confusion, the final estimation of sound source direction is based on monaural spectral cues and on the effects of head rotation on the cues.

In optimal conditions, humans can resolve the cone of confusion in which the sound source lies to an accuracy of within a few degrees. The accuracy is best for sound sources near the median plane and worst for sources near the interaural axis. The directional resolution within a cone of confusion is of the order of 20°. However, in the field of vision, the resolution is better. In suboptimal conditions, when reverberation or other sound sources exist, the directional accuracy in general is poorer.

SPATIAL REPRODUCTION

To be able to develop spatial audio technologies, the reproduction error has to be evaluated objectively or subjectively. The objective evaluation defines a target sound field and an error function, with the latter computed as the difference between the target and reproduced fields. The subjective evaluation quantifies human perception in both the target and the reproduced cases and defines the error as the difference between these.

Objective evaluation is beneficial in that it allows the mathematical analysis of the error function. Unfortunately, the physical difference between the target and reproduced fields does not necessarily tell much about perceptual difference, and often theoretically derived reproduction methods demand an excessively large number of microphones or loudspeakers. Also, in theoretical formulations of the systems,

Digital Object Identifier 10.1109/MSP.2008.918025

1053-5888/08/$25.00©2008IEEE

some characteristics of transducers, such as directivity and impulse response, are typically assumed to be ideal, whereas in reality the characteristics are nonideal.

Subjective evaluation is beneficial in the sense that the measured error directly quantifies the perceptual difference, which is of interest as the techniques are designed for human use. Unfortunately, the perceptual evaluation does not allow an easy development of reproduction methods. The development is then based on the trial-and-error approaches, which can be adequate in some cases.

Several different methods have been proposed for the reproduction of spatial sound over loudspeakers. Some of the typical error sources in the methods to be considered include coloration, the size of the listening area, and the directional error.

Coloration refers to the change of the timbre of the sound by a reproduction method. Typically, coloration appears when the method changes the magnitude spectra in the ear canal signals. Avoiding coloration is important for spatial sound reproduction methods, because strong colorations degrade the perceived overall quality of sound, no matter how accurately spatial aspects are reproduced.

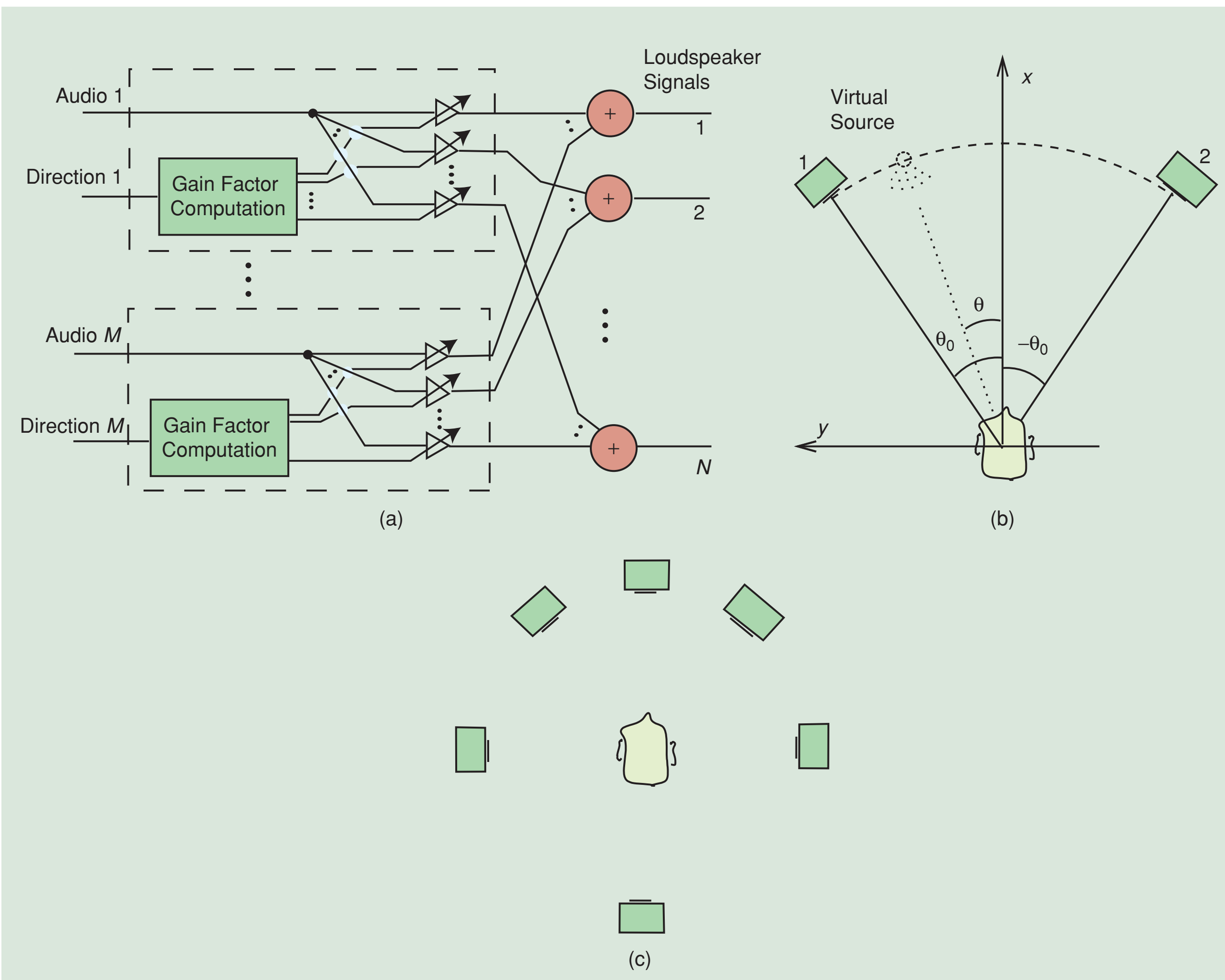
The size of the listening area where a desired spatial effect is perceived is another important factor. If the best listening area is very small and if the perceived overall quality of sound degrades significantly outside the area, the system will never be in wide use.

The directional error in a reproduction system is defined as the deviation between the intended direction and the actually perceived direction of the sound. In practice, directional error can be fairly large before the listeners rate it as annoying. Typically, an error of tens of degrees can be tolerable. However, when all of the virtual sources of sound are perceived as originating from the near-most loudspeaker when sitting outside the best listening area, the directional error can be extremely objectionable.

TRADITIONAL AMPLITUDE PANNING

SIGNAL FLOW

Amplitude panning refers to techniques in which a monophonic audio channel



[FIG1] (a) Signal flow in amplitude panning of M audio channels to N loudspeakers. (b) Standard stereophonic listening layout, where $\theta_0 = 30^\circ$. (c) Example of a multichannel horizontal loudspeaker layout.

is applied to all or a subset of loudspeakers with different gains, as shown in Figure 1(a). Depending on gain relationships the listener will perceive the virtual source, also known as phantom source, in a direction that does not necessarily match with the direction of any of the loudspeakers. It is very likely that this method was originally developed by a trial-and-error approach and validated by subjective listening, as it is clear that the created sound field does not match the sound field created by a single sound source, although the listeners perceive it like that.

The fusion of two physical sources into one perceived source is a slightly surprising fact when compared with vision. If two spatially separated light sources with the same color are presented, the sources do not generally merge together perceptually. The merging in the auditory case results from the summation of sound signals in the ear canals. The hearing system cannot separate the coherent signals from different directions after this summation.

AMPLITUDE PANNING IN STEREOPHONY

The most common audio format is the two-channel stereophonic audio format. The loudspeaker setup shown in Figure 1(b) is meant to be used when reproducing stereophonic audio content. Two equidistant loudspeakers are placed in front of the listener. An angular separation of 60° has been selected, as larger separations make virtual sources unstable. In most of the practical cases, the listening configuration may be different, since typically the listener is not placed in the best listening position. Therefore, it is important to obtain good sound quality outside the best listening position.

Amplitude panning is widely used for positioning virtual sources in the stereophonic setup, as it is included in all mixing consoles with stereophonic output. The success of amplitude panning lies partly on the simplicity of its electronic implementation in early mixers. Other reasons for its success are discussed next.

In the best listening position, the directional quality is good, since the direction of a virtual source is controllable in the sector between the directions of the loudspeakers and also the width of the perceived source is relatively narrow. The best listening area is fairly small, extending maximally tens of centimeters to the left and right from the line which has all points equidistant from the loudspeakers. Outside the best listening area, amplitude-panned virtual sources are localized to the closest loudspeaker if the signal level in the farthermost one is not much higher. However, prominent coloration does not exist at any listening position, although using two loudspeakers emanating the same signal produces a slightly audible comb filter effect.

[THE MOST STRAIGHTFORWARD APPLICATION OF VBAP IS IN THE ARTISTIC CREATION OF SOUNDSCAPES AROUND THE LISTENER IN MULTILOUDSPEAKER SETUPS.]

The relationship between gain factors and perceived direction is discussed next. In audio post-production, sound engineers manually adjust the ratio of gains until they perceive the virtual source in the desired direction, and a priori knowledge of virtual source direction is not required. However, in some applications, automated virtual source positioning is needed. The perceived direction θ of a virtual source can be quantified by analytically computing the sound pressure in the ear canals in the best listening position. If the paths of propagation from the left and right loudspeakers to both ear canals are assumed to be curved lines bending over a spherical head to the ear, it can be shown that θ is dependent on gains g_1 and g_2 of the loudspeakers in following way:

$$\frac{\tan\theta}{\tan\theta_0} = \frac{g_1 - g_2}{g_1 + g_2}, \tag{1}$$

where θ_0 is the loudspeaker base angle, as shown in Figure 1(b) [2]. The gain factors cannot be resolved from the given θ from this equation, as only their ratio can be computed. To solve the gain factors, the gain factors can be normalized by

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

where N is the number of loudspeakers.

In listening tests it has been found that the perceived direction can be estimated relatively accurately with these equations. However, with band-pass noise stimuli, the perceived direction deviates from the estimated direction near 1.6 kHz. To explain these phenomena, the ITD and ILD cues have been monitored with computational auditory modeling [3]. The *level* difference between loudspeakers has been shown to be turned in (a bit counterintuitively) to a *time* difference between the ears at frequencies below about 1,000 Hz. At these frequencies the head does not shadow the sound, and the sounds from *both* loudspeakers arrive at each ear canal, where they are superposed, or summed together. The sound from the left loudspeaker arrives at the left ear 0.6 ms earlier than the sound from the right loudspeaker at the same ear. This tiny difference in propagation time changes the phase of the summed signal as a function of the level difference between the loudspeakers. This produces the ITD of the virtual source and is called summing localization.

At high frequencies, the head shadows the signal coming from the other side of the median plane, and summing localization does not occur. The level difference between loudspeakers is there turned into a level difference between ear canals. Near 1,600 Hz, the ITD cue is unstable and ILD gets anomalous values, thus the interaural cues produced by amplitude panning do not match perfectly with natural cues produced with a real sound source. This explains why amplitude-panned virtual sources are perceived somewhat broader than the real ones.

HORIZONTAL LOUDSPEAKER LAYOUTS

When more than two loudspeakers are used in horizontal positioning, as shown in Figure 1(c), pairwise panning can be used to position virtual sources. The loudspeaker setup is treated as a set of stereo pairs, and the virtual source signal is applied to two adjacent loudspeakers at a time, as proposed in [4]. From the fact that amplitude panning works fine with a frontal loudspeaker pair, it thus has been guessed that this holds also with loudspeaker pairs in other directions. This has been verified by subjective listening, where it has been found that the virtual sources are perceived between the loudspeakers, although the virtual sources are wider than with a frontal loudspeaker pair. The directional quality of a virtual source degrades especially if the interaural axis lies between the pair, but the directional accuracy of humans is, anyway, low in such directions. When automated virtual source positioning is needed, the tangent law, or some other similar law, can be simply used for pairwise panning. However, this may not be always valid, and typically the perceived virtual source direction is to some extent biased towards the median plane from the desired direction [3].

A nice feature of pairwise panning is that, in any listening position, the virtual source cannot be perceived outside the sector defined by the loudspeaker pair and the listener. Thus, the maximal directional error is of the same order as is the angular separation between the loudspeakers, which implies that the directional error can be decreased by adding loudspeakers around the listener. In practice it has been found that already eight loudspeakers provide an acceptable directional quality over a large listening area.

VECTOR-BASE AMPLITUDE PANNING

DESCRIPTION FOR 3-D SETUPS

When loudspeakers also are placed above or below the level of listeners' ears, a natural approach to extend pairwise panning has been to use loudspeaker triangles to reproduce the virtual sources, also known as triplet-wise panning. Such loudspeaker setups as presented in Figure 2(b) are rare in domestic use. However, in public venues such as theaters, concert halls, and in some cinemas they are common. This kind of 3-D loudspeaker setups are problematic for automated panning, since the tangent law presented in (1) cannot be generalized to spherical coordinates in geometrical form. A method called vector base amplitude panning (VBAP) was developed by the first author to overcome this problem [3], whereby panning is defined with vector bases, which appears to be a generic reformulation of the tangent law.

A Cartesian unit vector $\mathbf{l}_n = [l_{n1}\ l_{n2}\ l_{n3}]^T$ points to the direction of loudspeaker n from the listening position. A loudspeaker triplet is defined by a vector base with unit vectors $\mathbf{l}_n, \mathbf{l}_m$, and $\mathbf{l}_k$. The panning direction of a virtual source is defined as a 3-D unit vector $\mathbf{p} = [p_1\ p_2\ p_3]^T$. A sample configuration is presented in Figure 2(a).

The panning direction vector $\mathbf{p}$ is expressed as a linear combination of three loudspeaker vectors, and in matrix form

$$\mathbf{p} = g_n\mathbf{l}_n + g_m\mathbf{l}_m + g_k\mathbf{l}_k\,, \tag{3}$$

$$\mathbf{p}^T = \mathbf{g}\mathbf{L}_{nmk}\,. \tag{4}$$

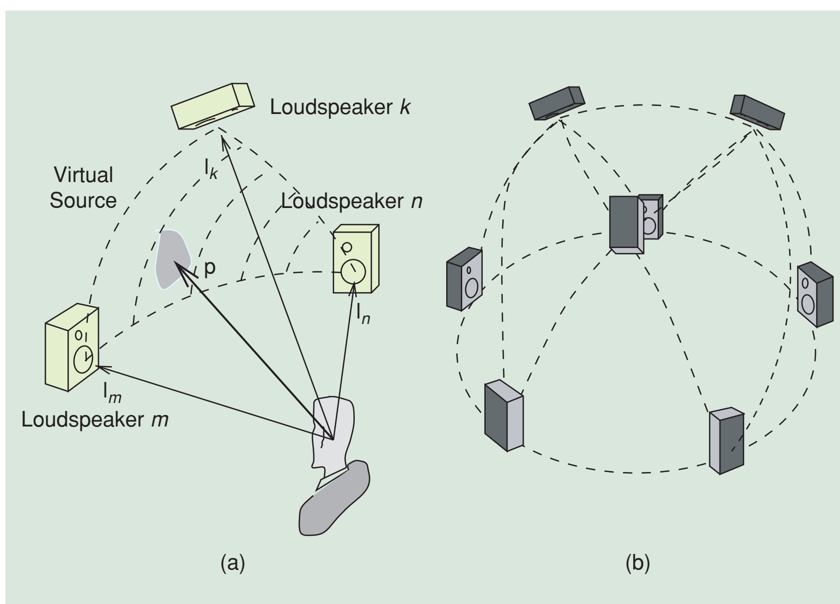
Here g_n, g_m, and g_k are gain factors, $\mathbf{g} = [g_n\ g_m\ g_k]$, and $\mathbf{L}_{nmk} = [\mathbf{l}_n\ \mathbf{l}_m\ \mathbf{l}_k]^T$. Vector $\mathbf{g}$ can be solved as

$$\mathbf{g} = \mathbf{p}^T\mathbf{L}_{nmk}^{-1} \tag{5}$$

if $\mathbf{L}_{nmk}^{-1}$ exists, that is true if the vector base defined by $\mathbf{L}_{nmk}$ spans a 3-D space. Using (5) one calculates barycentric coordinates of vector $\mathbf{p}$ in a vector base defined by $\mathbf{L}_{nmk}$. The components of vector $\mathbf{g}$ can be used as gain factors after scaling them with (2).

When there are more than three loudspeakers in 3-D positioning, the loudspeaker triangle in which the panning direction lies is selected, and gain factors are computed for it. VBAP can also be easily implemented for pairwise panning, where each loudspeaker base composes naturally of two loudspeaker vectors.

In perceptual tests with panning inside loudspeaker triangles, it has been found that virtual sources are perceived relatively consistently with the panning direction. Panning direction and perceived direction lie typically in the same cone of confusion, which means that the



[FIG2] **(a) Vector base formulation of triplet-wise panning. (b) Triangulated 3-D loudspeaker setup.**

error in the left/right direction is low. However, the perceived direction may vary within the cone of confusion individually. Typically this means that the virtual source is localized to a different elevation than intended. However, the virtual source is perceived in most cases inside the loudspeaker triangle. This means that directional error can be made smaller by adding loudspeakers, which decreases the sizes of the triangles [3].

APPLICATIONS

The most straightforward application of VBAP is in the artistic creation of soundscapes around the listener in multiloudspeaker setups, as used in theaters and installations. Virtual sources composed of single-track audio are positioned with VBAP around, above, or below the listener, and the listeners perceives the direction of sound similarly, although somewhat depending on the listening position. There exist multiple implementations of VBAP in sound processing software where the positioning of virtual sources is performed abstractly using directional angles, and the loudspeaker layout is defined as data which can be changed if the setup is changed. Thus the same control mechanism can be used with different loudspeaker layouts, which is beneficial when the same piece is presented in different venues, often having different 2-D or 3-D setups.

This processing does not add room effect to sound or control the distance of a virtual source. For artistic applications, where the room effect does not need to be accurately matched with predefined acoustical conditions, room effect can be controlled by processing each virtual source signal with a DSP technique that produces reverberation. Optimally, several reverberated signals are created from each input signal or from the sum of input signals, which are then reproduced over discrete loudspeakers in different directions.

In addition to the artistic use, VBAP has also been applied to convey some information to the listener, such as in navigation, orientation, or data sonification tasks. In such cases the panning directions have to be adjusted automatically. These tasks are often conducted in 3-D audiovisual virtual environments, which can be used for human training purposes, to visualize complex structures, or to conduct multimodal psychophysical experiments. For example, in molecule visualization, the 3-D structures can be so complicated that the viewer cannot see all atoms at once. Attaching continuous sounds to some atoms, which are automatically panned to the correct directions, helps keeping orientation when monitoring the molecule inside the structure.

An ambitious task for such audiovisual environments is to reproduce the acoustics of a virtual world accurately for added naturalness or for uses in design of acoustics. If the acoustics of a virtual space is modeled with ray-based methods, the direct sound path and each reflection can be auralized by applying the corresponding sound to a virtual source positioned with VBAP.

In the future, the authors foresee growing demand for such loudspeaker-layout independent techniques as VBAP. For example, the audio industry nowadays is selling different audio records for different loudspeaker formats, although the content originates from the same recording. A goal is to develop a generic audio format that would produce high-quality audio with different listening setups. Some techniques aiming to such a goal have already been proposed in which few channels of audio are transmitted together with metadata containing time-dependent information of spatial attributes of sound [5], [6]. This information is then used in reproduction over different loudspeaker layouts.

CONCLUSIONS

Pairwise and triplet-wise amplitude panning methods provide a robust directional quality to a large listening area without strong coloration or other artifacts, whereby monophonic sound is applied to a subset of loudspeakers with different gain factors. VBAP is a method to compute gain factors for arbitrary panning directions in arbitrary loudspeaker setups. The method is mathematically simple and computationally efficient. The perceived direction of a virtual source matches relatively well with the panning direction, and the error within a large listening area compares to the angular separation between adjacent loudspeakers viewed from the listening point.

AUTHORS

Ville Pulkki (ville.pulkki@tkk.fi) is an Academy research fellow with the Laboratory of Acoustics and Audio Signal Processing at the Helsinki University of Technology. His research interests include recording, reproduction, and perception of spatial sound.

Matti Karjalainen (matti.karjalainen@tkk.fi) is a professor in acoustics at the Laboratory of Acoustics and Audio Signal Processing at Helsinki University of Technology. His interests cover many fields of acoustics and audio signal processing.

REFERENCES

More resources are available at: http://www.acoustics.hut.fi/research/

[1] B.C.J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. San Diego, CA: Academic, 2003.

[2] J.C. Bennett, K. Barker, and F.O. Edeko, "A new approach to the assessment of stereophonic sound system performance," *J. Audio Eng. Soc.*, vol. 33, no. 5, pp. 314–321, May 1985.

[3] V. Pulkki, *Spatial Sound Generation and Perception by Amplitude Panning Techniques*, D.S.c. dissertation, Helsinki Univ. Technol., Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 2001 [Online]. Available: http://www.hut.fi/Yksikot/Kirjasto/Diss/2001/isbn9512255324

[4] J. Chowning, "The simulation of moving sound sources," *J. Audio Eng. Soc.*, vol. 19, no. 1, pp. 2–6, 1971.

[5] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, 2007.

[6] M.M. Goodwin and J.-M. Jot, "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, 2007, pp. I-9–I-12.

SP