Creating generic soundscapes in multichannel loudspeaker systems using vector base amplitude panning in Csound synthesis software

VILLE PULKKI

Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, PO Box 3000, FIN 02105, Espoo, Finland, and Sibelius Academy Computer Music Studio, PO Box 86, FIN 00251, Helsinki, Finland
E-mail: Ville. Pulkki@hut.fi
URL: http://acoustics.hut.fi/

Vector base amplitude panning (VBAP) is a new method for soundscape creation. It generalises the pair-wise panning paradigm to the triplet-wise panning paradigm, which can be used in three-dimensional loudspeaker setups. Similar soundscapes can be created with differing numbers of loudspeakers in different positions. The VBAP implementation in Csound synthesis software presented in this paper is a powerful tool for producing generic soundscapes. Virtual sources are positioned by defining the virtual source directions as (azimuth, elevation) pairs; the loudspeaker setup is defined in the initial settings. Virtual sources can be moved in three different ways: they can be moved along a polyline between specified directions, angular velocities can be applied to them, or their directions can be controlled arbitrarily.

1. INTRODUCTION

Audio research has throughout its history been largely directed to immersive virtual environments. The first audio reproduction devices were monophonic, which enabled the production of point-like soundscapes. Two-channel stereophony (Blumlein 1931) was a big step forward; it enabled the soundscape to be enlarged between the loudspeakers and is still the most used format.

There have been many attempts to enlarge the produced soundscape. One method is to increase the number of loudspeakers and place them around the listener. If the number of loudspeakers exceeds four, the panning of virtual sound sources with a traditional mixing console is rather cumbersome. There exist, however, some modern mixers that are able to pan sounds to certain multi-loudspeaker configurations.

Panning methods for matrixing systems, such as Ambisonics (Gerzon 1972) or Dolby Surround (Dolby 1996) exist. In matrixing systems, the soundscape is encoded analogically to a few audio channels and then decoded to more-or-less defined loudspeaker configurations. However, the analogical encoding–decoding suffers a variety of problems. The perceived direction may vary greatly in the listening room, and the methods are often designed only for certain loudspeaker configurations.

The other way to create immersive soundscapes is to control accurately the sound signal entering the listener's ears. In this way it is theoretically possible to create illusions of sound images from all directions using only two sound sources (Møller, Sørensen, Hammershøi, and Jensen 1995). The method relies on simulating the filtering effects of the listener's pinna, head and torso by digital filters. The filtering effects are often presented as head-related transfer functions (HRTFs). Different techniques exist for loudspeaker production (transaural stereo) and for headphone production (binaural stereo). There are, however, strict boundary conditions. In transaural stereo, the illusion can be perceived only in a small listening area, the listener's head must be immobilised and the listening room cannot be reverberant. In binaural stereo, individual measurements of the HRTFs are needed and the movements of the listener have to be tracked.

In computer music it is a common problem that the loudspeakers are in different positions in different concert halls and studios. The ideal is a system that would be able to produce identical soundscapes using different loudspeaker configurations. The panning methods presented earlier do not provide a generic solution to this problem.

Vector base amplitude panning (VBAP) (Pulkki 1997) is a new approach. It is an amplitude panning method which can be used to position virtual sound sources using an arbitrary loudspeaker configuration. VBAP allows control of the virtual source directions without knowledge of the locations and number of loudspeakers.

In this paper some features of the spatial hearing and amplitude panning are reviewed. The theory of VBAP is discussed and its use in the Csound synthesis software is presented. Some produced soundscapes are discussed.

2. AMPLITUDE PANNING AND SPATIAL HEARING

The audio signals coming from a sound source and reaching a listener's left and right ears are delayed and filtered differently. Due to these differences, there

Organised Sound **3**(2): 129–34 © 1998 Cambridge University Press. Printed in the United Kingdom.

https://doi.org/10.1017/S1355771898002052 Published online by Cambridge University Press

is a frequency-dependent interaural time difference (ITD) and an interaural level difference (ILD) between the signals arriving at the auditory system (Blauert 1997). In the brain's source localisation process, the ILD and the ITD are the main cues. However, the information from the ILD and ITD does not resolve unequivocally the direction of the sound source. The localisation is focused with additional cues, such as head rotation, signal spectrum perception, and visual cues.

In the simple amplitude panning method, two loudspeakers radiate coherent signals which may have different amplitudes. The listener perceives an illusion of a single virtual sound source (phantom sound source), where the relationship of the amplitudes of the signals controls the direction of the virtual source.

When the sound signals arrive within one millisecond of each other at the listener's ears, summing localisation occurs. The sound signals are summed and they form a new signal that enters the spatial hearing mechanism. The phase difference between the signals entering the ears is dependent on the loudspeaker signal amplitudes (Bauer 1961). Therefore, the perceived ITD between the summed signals entering the auditory system is an interpolation of the ITDs that would occur if each loudspeaker was emanating the signal alone. At high frequencies the summing localisation is not valid anymore because of head shadowing.

A typical loudspeaker configuration is illustrated in figure 1. Two loudspeakers are positioned symmetrically with respect to the median plane. The amplitudes of the signals are controlled. The loudspeakers are typically positioned with a base angle of 60°. The virtual source can be placed on an arc between the loudspeakers, which is called the *active arc*. The localisation on the active arc can be described with the stereophonic law of sines (Blumlein 1931); if the listener turns his/her head following the virtual source, the tangent law is more correct (Bernfeld 1973).

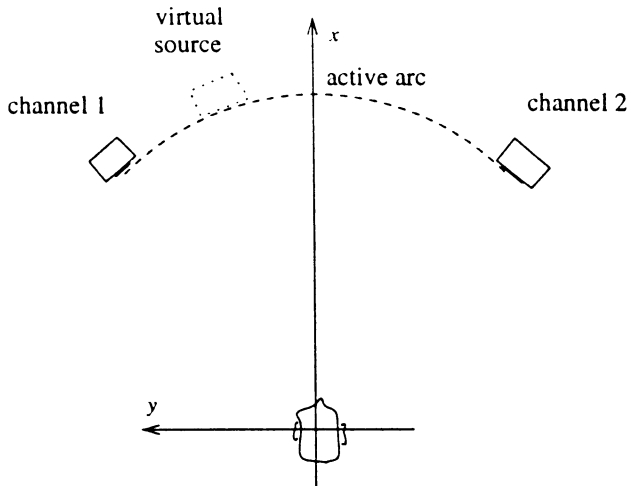


Figure 1. Typical two-channel stereophonic configuration.

The distance between the listener and the virtual source can also be controlled to some extent by altering the level of the signal and by adding reverberation. When this control is added, the virtual sources can be positioned in the region of the two-dimensional (2D) sector defined by the listener and the loudspeakers.

The virtual sources can be produced with loudspeakers in other directions than in the forward sector of the listener. However, there are some limitations on the loudspeaker placement. The side directions (left-most and right-most directions) should not be on an active arc of any loudspeaker pair. It this happens, there is an area around the side point where it is impossible to position virtual sound sources (Theile and Plenge 1977). This is avoided by placing loudspeakers to the right-most and left-most directions (±90° azimuth).

The angle between loudspeakers affects the localisation accuracy: the smaller the angle, the better the virtual source quality. When the angle becomes larger, the virtual source gets spectrally more spread and more colourations occur. The virtual source may also be perceived inside the listener's head. Also, the virtual source is localizable in a larger listening area the smaller the angle between the loudspeakers.

In modern surround loudspeaker systems there are typically five loudspeakers in a horizontal plane around the listener. When a single sound source is panned in such a system, this is in many cases carried out using a pair-wise panning paradigm (Chowning 1971) in which the sound is applied only to two adjacent loudspeakers. In this way, robust virtual sources can be created. This panning method is called '2D panning', and the loudspeakers arrangement is called '2D loudspeaker setup'.

Using the pair-wise panning paradigm in a 2D loudspeaker setup, it is possible to create 2D soundscapes. Using amplitude panning it is not possible to extend the soundscape to upper and lower directions. To accomplish this, we also have to place loudspeakers at high and low elevations and reformulate the panning laws for these situations.

VECTOR BASE AMPLITUDE PANNING

We may extend the stereophonic configuration with a third loudspeaker placed in an arbitrary position at the same distance from the listener as the other loudspeakers (figure 2). The virtual source can now appear within a triangle formed by the loudspeaker triplet, which is called the *active triangle*. The panning

https://doi.org/10.1017/S1355771898002052 Published online by Cambridge University Press

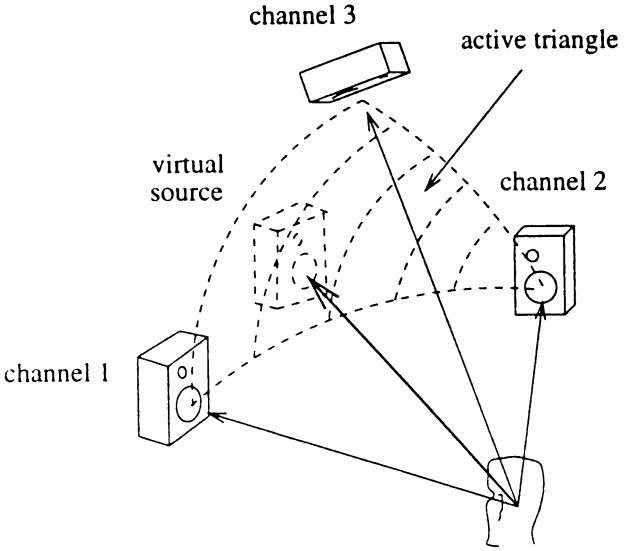


Figure 2. A sample configuration for 3D amplitude panning. The loudspeakers form a triangle into which the virtual source can be placed.

method for such a loudspeaker setup is called '3D panning'. Three-dimensional VBAP is a panning method which can be used to position virtual sources in the active region of a loudspeaker triplet with arbitrary loudspeaker positions.

In 3D VBAP, the loudspeaker positions and the direction(s) of virtual source(s) are defined using unit-length 3D vectors in Cartesian coordinates. Each loudspeaker triplet forms a 3D vector base. The direction vector of the virtual source is then expressed as a linear weighted combination of the vectors of a base. The gains are solved by simple matrix manipulation. If the virtual source is located outside of the active region, at least one of the gains is negative. If the gains are non-negative, they can be used as gain factors in amplitude panning after they have been normalised properly. The complete method is presented in Pulkki (1997).

As in the pair-wise panning paradigm, where the 2D loudspeaker setup is split into loudspeaker pairs, we may similarly split any 3D loudspeaker setup into loudspeaker triplets, as shown in figure 3. At any one time the virtual source is applied to one loudspeaker triplet.

The particular triangle on which the virtual source is located is selected to be the triangle to which the virtual source is panned. The selection can be performed by calculating the gain factors in all triplets and selecting the one which produces positive factors.

VBAP can also be used for any 2D loudspeaker setup. The panning is performed using the pair-wise paradigm. In this condition the dimensionality of the vector bases is naturally two. Two-dimensional VBAP is consistent with the tangent law (Pulkki 1997).

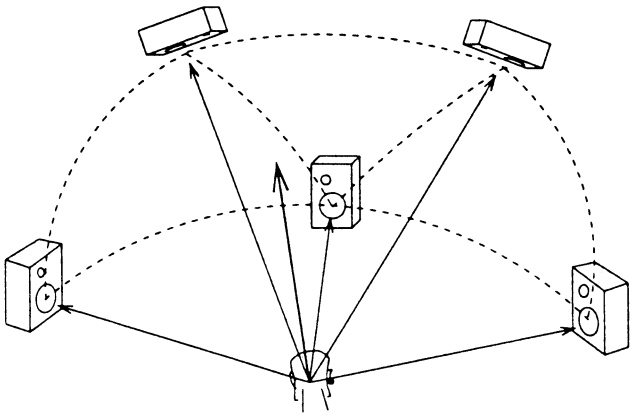


Figure 3. 3D VBAP with five loudspeakers. Vectors point to the loudspeakers and to the virtual source. Loudspeaker triplets are denoted with dashed lines.

3.1. Some features of VBAP

If the virtual source is panned in the same direction as any of the loudspeakers, the signal emanates only from that particular loudspeaker. If the virtual source is panned to a line connecting two loudspeakers, the sound is applied only to that pair, following the tangent law. If the virtual source is located at the centre of the active triangle, the gain factors of the loudspeakers are equal. These properties imply that VBAP produces virtual sound sources that are as sharp as possible with any loudspeaker configuration and amplitude panning. Also the direction errors and blur decrease when the number of loudspeakers is increased.

The above properties also imply that when moving virtual sources are applied, the movement over the borders of the active regions will not produce disturbing effects, since the loudspeaker(s) that are not active after transition are faded out before the borderline.

In amplitude panning it can be disturbing to find that the localisation blur and timbre of the virtual source vary slightly when the virtual source is moved. This happens because the quality of the virtual source is dependent on the number of loudspeakers producing it. This effect can be decreased by applying the same sound signal to two virtual sources located close to each other. The listener perceives still a single virtual source, but the colouration is less dependent on direction because the virtual source is never emanating from only one loudspeaker. Thus the virtual source would be blurred in all directions and coloured to some degree. When both virtual sources are in the same loudspeaker triplet, the situation equates to having a single virtual source between the two virtual sources.

Computationally, VBAP is very efficient: only three multiplications are required when a signal

https://doi.org/10.1017/S1355771898002052 Published online by Cambridge University Press

sample is panned. To calculate gain factors for one virtual source, less than 1,000 floating operations are needed. If a non-moving virtual source is applied, the gain factor calculation must be performed only once, during the initialisation. If a moving virtual source is applied, new gain factors must be calculated, e.g. ten times a second.

4. SOUND SPATIALISATION IN CSOUND USING VBAP

Csound is a free sound synthesis and composing software (Csound 1997). It uses two files, an orchestra file and a score file. In the orchestra file, the attributes of the instruments are defined. Correspondingly, the time-dependent information of the music is defined in the score file. The Csound application takes as input these two files and produces as output the audio file.

The VBAP has been implemented in Csound. The implementation lets the user define the positioning or movements of the virtual sources in the score file. The loudspeaker configuration is defined as initialisation data in the orchestra file. This enables the creation of abstract soundscapes. The same score file can be used to create similar soundscapes with different loudspeaker configurations.

A loudspeaker configuration is initiated with unit generator `vbaplsinit`, which takes as arguments the dimensionality of the loudspeaker system and the amount and the angular directions of the loudspeakers (figure 4). The function is performed during the initialisation of Csound: it selects the loudspeaker sets (pairs or triplets) and calculates the corresponding initial data. The data is stored in an internal array in Csound. This is the only place where the number and placement of loudspeakers is needed in the Csound files. Later the positions of the virtual sources are controlled with direction angles.

Static panning can be performed with unit generator `vbapl`*n*. It takes as arguments the audio signal and the direction in which the sound is panned. Direction is expressed in azimuth and elevation angles in degrees. The instrument 1 in figure 4 uses

```
;;;;;;;;;;;;;;;;  VBAP test orchestra ;;;;;;;;;;;;;;;;;

sr = 44100      ; sample rate
kr = 441        ; control rate
ksmps = 100     ; sample rate / control rate
nchnls = 4      ; number of csound output streams

; 3-D loudspeaker initalization
;          dim  loudsp amnt  directions azi, ele, azi, ele,...
vbaplsinit 3,   8,           -30,0, 30,0, -90,0, 90,0, 180,0, -30,40, 30,40, 180,50


; An example of 2-D loudspeaker initalization
;           dim  loudsp amnt   directions azi,azi,...
;vbaplsinit 2,   6,            -30, 30, -150, 150, -90, 90

instr 1  ; Static panning to direction p4 = azimuth, p5 = elevation
        asig oscil 10000, 80, 1
 a1, a2, a3, a4, a5, a6, a7, a8     vbapl8  asig, p4, p5
        outq      a1, a2, a3, a4
;       outq      a5, a6, a7, a8
endin

instr 2  ;Virtual source moves point to point
        asig oscil 10000, 80, 1
 a1, a2, a3, a4, a5, a6, a7, a8   vbapl8move   asig, 8, 0,0, 170,30, -90,80, -70,0
        outq      a1, a2, a3, a4
;       outq      a5, a6, a7, a8
endin

instr 3   ;Angular velocity is applied to virtual source
        asig oscil 10000, 80, 1
 a1, a2, a3, a4, a5, a6, a7, a8   vbapl8move   asig, -8, -30,0, -65,40, -50,-450, -170,0
        outq      a1, a2, a3, a4
;       outq      a5, a6, a7, a8
endin
```

Figure 4. Csound orchestra file using VBAP implementation.

https://doi.org/10.1017/S1355771898002052 Published online by Cambridge University Press

this unit generator. The direction in which the sound is panned is received as the parameters p4 and p5 of the instrument statement.

There are three ways to create moving sounds in the current Csound implementation of VBAP. The unit generator `vbapl`*n*`move` can move the virtual source on a path defined by given directions, or it can apply given angular velocities to the virtual source. The third way is to use the static panning function `vbapl`*n* and define the direction parameters as global variables that are altered depending on time.

When the virtual source is moved on a path with the `vbapl`*n*`move`, the directions are defined in 2D loudspeaker systems as azimuth angles and in 3D loudspeaker systems as (azimuth, elevation) pairs. The virtual source is moved along a polyline that connects all the specified directions. The function also takes as a parameter the number of fields in the function call. Instrument 2 in figure 4 pans the sound signal using the direction-to-direction path; the path is illustrated in figure 5.

If the number of fields is positive, direction-to-direction panning is applied, and if it is negative, angular velocities are applied. In the latter case, the starting direction and velocities of the virtual source are defined, and the angular velocities are applied to the virtual source during equal time intervals. Instrument 3 in figure 4 applies the angular velocities, and the path is illustrated in figure 6. The use of angular velocities is selected in a function call by setting the number of fields to a negative value.

In the unit generator names, the *n* denotes the number of audio outputs, which is currently four or eight. The number has to be the same or greater than the number of loudspeakers. If the number of output streams is greater, the supplementary streams have null output. In this implementation, the number can

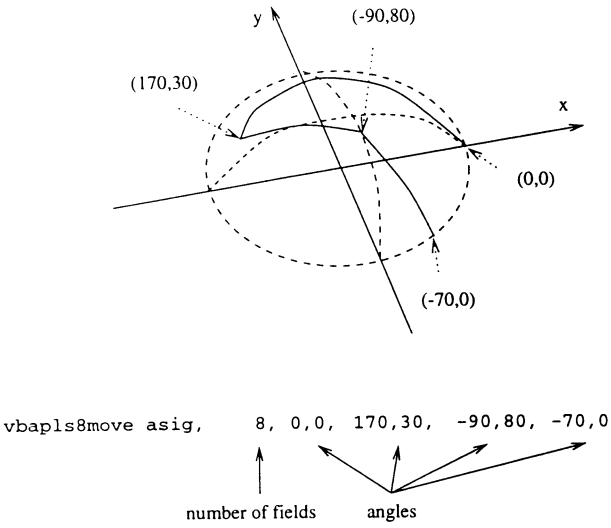


Figure 5. Virtual source moving along a polyline.

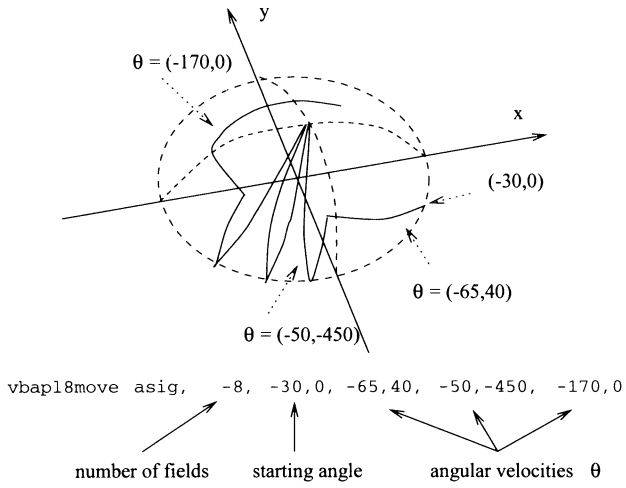


Figure 6. A time-variable angular velocity applied to a virtual source.

be eight at maximum, which can be raised easily if needed.

The audio output of Csound is stored to a file or is produced in real time. Its maximum number of output channels is currently limited to four. This limitation can be avoided by running Csound many times and by writing the output in four-channel blocks to disk. The blocks are then combined.

4.1. Test runs

The Csound implementation can be used to create very many different types of soundscapes. Some experiments have been run with the system. The capability to specify each sound's static direction can be effective in some situations. The chirp of a bird was panned to random directions and time intervals between chirps. This created a strong feeling that the listener was in a jungle. The feeling was completed with some yells of apes, lions and Tarzan from different directions.

The direction-to-direction panning capability was very useful when simulating moving such items as trains, cars or aeroplanes. The angular velocity panning was applied in a demonstration in which a soul music band was rotated around the listener. At the beginning of the demonstration the rhythm section and the lead vocalist were panned in front and backing vocals and brass were panned to the back of the listener. An equal angular velocity was then applied to all instrument azimuths. The band started to rotate around the listener and the velocity was accelerated constantly. When the velocity of the band was 60° s^{-1}, a small angular velocity was also applied to the elevation, and the listener perceived the band rising slowly in a spiral until reaching the zenith and starting to descend back to the horizontal plane. At the

https://doi.org/10.1017/S1355771898002052 Published online by Cambridge University Press

end of the song the angular velocity of the band was 900° s^{-1}.

The number of loudspeakers was limited to eight in all tests. Some tests were carried out with six loudspeakers in the horizontal plane and two in the upper directions. It was found that the horizontal panning was performed successfully, but two loudspeakers above the listener were not enough. A compromise was found in a system in which five loudspeakers were assembled in the horizontal plane in directions 30°, 90°, 180°, –90° and –30°. Three loudspeakers were mounted evenly in the upper directions of elevation 40° or 50°. This enabled a fairly good localisation accuracy in the whole upper hemisphere.

The VBAP implementation can also be used to create arbitrary virtual acoustic realities. The direct sound and reflected sounds can be panned to the directions where they occur. The effects of air absorption and wall reflections can be introduced to the signals with other Csound unit generators, as can diffuse reverberation.

The source codes and manual pages of VBAP unit generators can be downloaded from a Web site (VBAP 1998).

REFERENCES

Bauer, B. B. 1961. Phasor analysis of some stereophonic phenomena. *Journal of the Acoustical Society of America* **33**(11): 1,536–9.

Bernfeld, B. 1973. Attempts for better understanding of the directional stereophonic listening mechanism. *44th Convention of the Audio Engineering Society*. Rotterdam, The Netherlands.

Blauert, J. 1997. *Spatial Hearing*. Revised edn. Cambridge, MA: MIT Press.

Blumlein, A. D. 1931. UK Patent 394,325. Reprinted in *Stereophonic Techniques*. New York: Audio Engineering Society, 1986.

Chowning, J. 1971. The simulation of moving sound sources. *Journal of the Audio Engineering Society* **19**(1): 2–6.

Csound. 1997. The Leeds Csound Front Page. URL: http://www.leeds.ac.uk/music/Man/c_front.html

Dolby. 1996. URL: http://www.dolby.com/

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

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. 1995. Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society* **43**(5): 300–21.

Pulkki, V. 1997. Virtual source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society* **45**(6): 456–66.

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

VBAP. 1998. Vbap source codes. URL: http://acoustics.hut.fi/~ville/software/CSound_VBAP/

https://doi.org/10.1017/S1355771898002052 Published online by Cambridge University Press