3-D Sound Spatialization using Ambisonic Techniques

Author(s): David G. Malham and Anthony Myatt

Source: *Computer Music Journal*, Winter, 1995, Vol. 19, No. 4 (Winter, 1995), pp. 58-70
Published by: The MIT Press
Stable URL: https://www.jstor.org/stable/3680991

David G. Malham and Anthony Myatt
Music Technology Group
Department of Music
University of York
York YO1 5DD, UK
{dgm2, am12}@unix.york.ac.uk
http://www.york.ac.uk/insts/mustech/welcome.htm
http://www.york.ac.uk/depts/music/welcome.htm

3-D Sound Spatialization using Ambisonic Techniques

The exploration of sound spatialization is a preoccupation of many composers and performers of electroacoustic music. Two-channel stereo techniques are widely used in the genre, but more sophisticated forms of sound spatialization are often restricted to those with access to significant technical resources. Recent technological developments have increased access to high-quality, multichannel audio systems and computer music workstations, and new applications for sound spatialization have been identified: virtual reality, multimedia computing, films, videos, and computer games may all employ surround-effects of some type. In addition, increased computing power has led to the production of workstations that can support the data bandwidth required for multiple audio channels. All of these factors have led to a renewed interest in sound spatialization techniques, providing enormous potential for accelerated development of the field.

Ambisonics is a powerful technique for sound spatialization. It can allow recording, manipulation, and composition with naturally and artificially constructed three-dimensional *sound fields*. In this article we will present the results of experimentation and research into the application of ambisonics to the composition and performance of electroacoustic music. We also present Csound implementations of ambisonic encoding and decoding techniques that can be used on any computing platform supporting four or more independent audio output channels. Our hope is that these techniques will be a useful resource for those wishing to experiment with sound spatialization, and that their presentation here will encourage further developments.

Computer Music Journal, 19:4, pp. 58–70, Winter 1995

This article examines a variety of spatialization techniques within a historical and technical context, identifying some of the advantages and disadvantages of specific methods. This provides the basis for a review of ambisonic theory, which additionally describes several advanced spatial-processing methods that are possible with this technique. We go on to discuss experimental results and compositional considerations related to the use of ambisonic techniques for electroacoustic music performance.

Historical Context

The first recorded use of multiple audio channels to give a spatial effect occurred more than a century ago. Clément Ader's use of multiple sets of telephone transmitters and receivers to relay the sound from remote events created enormous public enthusiasm at the 1881 Paris Exhibition of Electricity (Askew 1981). However, no further work of any significance was done on multi-channel audio until the 1920s, when a binaural/headphone-based system was developed by Harvey Fletcher and his team at Bell Telephone Laboratories (Sanal 1976). Binaural signals are not suitable for presentation over loudspeakers, and further work was needed to produce methods appropriate for more than one listener.

Bell Telephone Laboratories (Fox 1982) developed a system that used a curtain of microphones hung in front of the sound source and fed a similar curtain of loudspeakers. The intention was to recreate the original sound wavefront. As we now know, accomplishing this requires a very large number of channels, running into many tens or even hundreds of thousands (Gerzon 1974). Fortunately for Bell, research revealed acceptable results when three spaced microphones were set left, center, and right, and they fed similarly placed speakers. They were much helped by the support of Leopold Stokowski in this work, and, indeed, excellent stereo recordings of his orchestra are available from as early as 1932.

Around the same time, at EMI in Britain, Alan Blumlein was working on a different approach, based on creating a convincing illusion of the original sound field, rather than the actual sound field itself. His simple, crossed pair of directional microphones that captured directional information on just two channels of audio is now widely regarded as one of the best ways to capture an audio event to be presented over two loudspeakers. The ideas inherent in this approach are also fundamental to the ambisonic 3-D surround sound system.

These stereo techniques were little used over the next few years, mostly as a result of the technological limitations of the available recording systems (despite the advent of multi-track film-based recorders). The next major development was featured in Walt Disney's animated film *Fantasia*, produced in 1939, again with Leopold Stokowski's involvement (Hope 1979). A team of engineers from RCA and Disney developed a special nine-track recorder, based on nine separate synchronized optical recorders. The orchestra was recorded with 33 microphones that were mixed by (orchestral) section onto six of the tracks. The seventh track contained a mono mix of the first six, the eighth track recorded the sound from a distant microphone, and the ninth track held a metronome or click track.

For the cinema presentation, a three-channel version of the eight-channel original was played back from a four-channel optical recorder that was synchronized with the film. The intention was that this would be played back over 90 loudspeakers spread behind the screen and around the auditorium, although financial pressures meant that this was rarely done in practice. Notches on the edge of the soundtrack film operated a switching mechanism that sent selected sounds to the side and rear speakers. Adrian Hope indicates that church bells were presented at the rear, and that the *Ave Maria* chorus progressed from the rear to link with the solo voice at the front. We believe that this may have been the first musical use of electronic sound spatialization.

In Europe during the early 1950s, composers involved in the new electronically based forms of music (and in this we are including both the *musique concrète* and *elektronische musik* schools) became interested in the possibility of using sound positioning and movement compositionally (Schaeffer 1951; Stockhausen 1956). For *Gesang Der Jünglinge*, Karlheinz Stockhausen employed five loudspeakers (or groups of loudspeakers). He intended four to be positioned around the audience and one above. The loudspeakers around the audience were fed from a four-track tape, and the fifth (vertical) signal was provided by a mono tape that was manually synchronized to the four-track tape. Even at the world premiere in the large broadcast studio of the West Deutscher Rundfunk (WDR), Cologne, it was impossible to use this configuration; the fifth loudspeaker had to be placed at center front. Since the premiere, only a four-track mix has been used. Stockhausen says of this piece, ". . . in this composition, for the first time, the *direction and movement of the sounds in space* [emphasis in the original] is shaped by a musician, opening up a new dimension in musical experience" (Stockhausen 1956).

Five years earlier, Jacques Poullin was working in Paris on the *potentiomètre d'espace* system for distributing sound among four loudspeakers, typically two in front of the audience, another above them, and one at the rear. This allowed a performer to position a sound by simply moving a small, hand-held transmitter coil toward or away from four large receiver coils arranged around the performer in a manner that reflected the loudspeaker positions. The sound controlled by the performer came from one track of a special five-track tape recorder; the other four tracks each supplied one loudspeaker to provide a mixture of live and preset sound positioning. It is interesting to note that this "position tracking system" is similar to those now used in virtual reality head-tracking systems. Pierre Schaeffer speaks of the experiments on the spatial projection of sound: ". . . le nouveau procédé est une dialectique du sons dans l'espace et je pense que le terme de *musique spatiale* [emphasis in original] lui conviendrait mieux que celui de stéréphonie," or, "the new process is a dialectic of sound in space and I think that the term *spatial music* could fit it better than stereophony" (Schaeffer 1951).

One of the most significant examples of spatial music was composed during the 1950s: Edgar Varèse's *Poème Electronique*. Featured in the Philips' pavilion at the Brussels World Fair of 1958, it was experienced by up to two million visitors. The technology used was complex and cumbersome, involving 15 tape recorders and 400 loudspeakers (Stimson 1991), and with the might of the Philips Corporation and their engineering expertise, it was highly successful. However, as with all work of this period (and indeed up until the 1970s), it lacked both a simple control system and the support of a comprehensive theory of sound localization.

Several attempts were made to design control systems for the movement of sound during this early period, including systems such as that developed at the West Deutscher Rundfunk in Cologne and used for Karlheinz Stockhausen's *Kontakte* (1960). This used a rotating, highly directional loudspeaker to distribute sounds between microphones. The outputs of the microphones were then recorded and played back over fixed loudspeakers. As computer-based systems became available, investigations started into various aspects of sound spatialization that were not previously accessible. In particular, work was done on artificial reverberation (Schroeder 1962; Moorer 1979) and Doppler shifts (Chowning 1971) which made a significant contribution to the understanding of distance and movement cues in artificially spatialized sounds. In 1983, F. Richard Moore published a description of a generalized model for spatial sound processing (Moore 1983), incorporated within the *cmusic unit generator* module, *space*.

In recent years, a number of systems offering spatial manipulations of sounds have appeared, and the application of computers has been of great benefit. There are both commercial systems, such as the various forms of Dolby Surround, ambisonics, Roland RSS, and Q-Sound, and also more specialized diffusion arrays like the UK-based Birmingham Electro-Acoustic Sound Theatre (BEAST) and Groupe de Musique Expérimentale de Bourges' GMEBaphone.

Approaches to Spatial Reproduction

Electronic systems that allow sounds to be positioned artificially in space can adopt one of two approaches. They can either attempt to mimic the complex changes in timbre, delay, and amplitude that occur directly at our ears as a sound moves from one position to another, or they can generate *phantom images* (Thiele and Plenge 1977) by amplitude profiling of feeds to multiple speakers. The first approach, based on the *Head Related Transfer Function* or HRTF, (and used in commercial systems such as QSound, Roland's RSS, and the Convolvotron (Begualt 1994), normally requires the use of headphones. Some variants, such as RSS, use *interaural cross-talk cancellation* (Cooper and Bauck 1989) to enable the images to be presented over a pair of loudspeakers. Systems using interaural cross-talk cancellation only work correctly in an extremely small "sweet spot" within the listening area. Presentation over headphones can be very good, approaching the listener's own capabilities for localization. This assumes, first, that personalized HRTFs are used, and second, that head movements are tracked to allow the system to modify the sound presented over the headphones, so that sound positions remain constant with respect to the external space, not the listeners head. Without the head tracking, *front-back reversals* are likely to occur: sounds intended to be due forward appear to come from behind, and vice-versa. This is a common problem with binaural recordings, i.e., stereo recordings made with microphones place in the "ears" of a dummy head. These front-back reversals are known to occur even within natural listening conditions, but there they are much less common. We use head rotation as one of the main ways to distinguish front from rear sound sources. Rotation should result in the sounds moving in one direction if they are in front of the head, and in the opposite direction if they are at the back. However, there appears to be no way at present to apply such rotations to naturally produced binaural recordings. Of course, some sounds are very difficult to localize under any circumstances (such as electronic telephone "bells"), so strategies such as head rotation do not necessarily work even within a natural sound field.

Whatever the technical merits of HRTF techniques, they are unlikely to be used in the concert situation until such time as someone can afford to equip every seat in an auditorium with its own headphones, head tracker, and digital signal processor unit providing HRTF processing. It will also be necessary to persuade concert goers to wear headphones. Such an approach may, however, be more acceptable within the context of concerts broadcast over the World-Wide Web. Artificial creation of sound fields using HRTFs also requires significant processing power, typically requiring many hundreds of multiple-accumulate instructions per sample, for each individual sound source.

The two-channel amplitude panned stereo, which we are accustomed to, is the most common example of a system that generates phantom images by amplitude profiling of feeds to multiple speakers. An 18-dB difference in the levels of a signal presented via two loudspeakers, spaced to subtend an angle of 60 degrees at the listener's position, will place a phantom image in the loudest loudspeaker. Progressively smaller differences will cause the image to move toward the halfway line between the speakers, then out toward the other speaker as that in turn becomes louder. The listening area where this works well is small, but larger than that for interaural cross-talk canceled HRTF. The image, especially for central sounds, becomes increasingly unstable as the angle between the loudspeakers (as seen by the listener) goes beyond the optimum 60 degrees. Once it reaches 90 degrees (as in four-loudspeaker *quadraphonic* systems!), stable central images cannot be formed, especially at the sides (Thiele and Plenge 1977). Image widths also tend to vary with the spectral content of the sounds used. Such amplitude-panned systems, whether they are two channel as above, four-channel quadraphonic systems, or some versions of the Dolby Surround family (including systems with higher numbers of channels, such as those proposed for HDTV systems), suffer from two major problems. In all systems of this type, the content of an audio channel is intended to provide the signal for a particular loudspeaker, implying that the loudspeaker layouts in the studio and playback venue must match exactly for the positioning of sounds to remain consistent. This was clearly recognized by those working at the Institute of Sonology at Utrecht State University, "There are four loudspeakers in this studio so that the user can form an impression of the sound field as it would be in the concert hall. The ideal dimensions of a multi-track studio are those of a medium-sized concert hall" (Weiland 1975).

It is, of course, rarely possible to match the size or shape of the final performance space when working in a studio. Moreover, a system of this type that has fewer than six channels directly feeding the same number of speakers cannot meet the criterion of 60-degree angles between adjacent speakers, so must inevitably suffer from unstable images in intermediate positions. While this can be tolerated for film or video work, where the listener's attention is locked to the relatively small area where the picture is presented, it is completely unsatisfactory for the composer of electroacoustic music. The only systems that overcome these difficulties to any real extent are the QMX system based on the work of Cooper and Shiga (Cooper and Shiga 1972) and the ambisonic system, based on the work of, among others, Michael Gerzon (Gerzon 1972). These two systems are fundamentally equivalent.

Control over sound positions and movements can either be accomplished in the studio or in the performance venue —or, ideally, both. For the composer in the studio to have a reasonably good idea of how sound objects will behave in the performance space, a system that allows for essentially seamlessly scalable loudspeaker configurations is necessary. This removes the changes produced as a result of differing numbers and positions of loudspeakers in the studio and performance spaces, leaving only the difference in actual acoustics to disrupt the composer's intentions. In purely technical terms, this rules out quadraphonics and related approaches, as noted above. However, in no way should this be taken to imply that it is invalid to treat loudspeakers and the acoustic space within which they reside as instruments to be played by skilled performers. In striving to present sounds as heard in the studio, we must be careful not to ignore the enormous potential of the speaker acoustics as an instrument, a potential that can be confirmed by anyone who has attended a BEAST or a GMEBaphone concert. In recent years, the York research group has been working toward hybrid systems, where both approaches can be combined.

Ambisonics

In our research program, we have concentrated on ambisonics, the only currently available system that comes close to achieving true scalability. This system has been fairly widely reported in the past (see, for instance, Malham and Orton 1991), but we still find a considerable lack of awareness of both its potential and principles.

The ambisonic surround-sound system is essentially a two-part technological solution to the problems of encoding sound directions (and amplitudes) and reproducing them over practical loudspeaker systems so that listeners can perceive sounds located in three-dimensional space. This can occur over a 360-degree, horizontal-only soundstage (*pantophonic* systems), or over the full sphere (*periphonic* systems). The system encodes signals in *B-format* which contains three channels for pantophonic systems and a further channel for periphonic, i.e., *with-height* reproduction. These signals convey directionally encoded information with a resolution equal to first-order microphones (cardioid, figure-eight, etc.). Reproduction requires four or more loudspeakers depending upon the required reproduction (pantophonic or periphonic) and on the size of the performance area. Practical minima are four if the sounds are limited to the horizontal plane, and eight if height is required. It is important to note that it is not necessary to consider the actual details of the reproduction system during the original recording or synthesis of a sound field. The only exception to this is that the vertical dimension is essential if a with-height replay system is required. If the B-format specifications are followed, assuming suitable loudspeaker/decoder systems are used, then operation in different venues will be as similar as local acoustics allow. In all other respects the two parts of the system, encoding and decoding, are completely separate.

Encoding Equations

Sounds positioned in ambisonic B-format are conceptually placed either on the surface of or within a "unit" sphere. If the maximum radius of the sound field is 1, sounds moved outside this sphere, i.e., with a radius greater than 1, will not be decoded correctly and they will tend to pull to the nearest speaker. This means that the sound source coordinates must obey the following rule:

$$(x^2 + y^2 + z^2) <= 1$$

where x is the distance along the X or front-back axis, y is the distance along the Y or left-right axis, and z is the distance along the Z or up-down axis.

When a monophonic signal is positioned on the surface of the sphere, its coordinates referenced to the center front position by the horizontal (A) and vertical (B) (counter-clockwise) angles, subtended at the listening position, are,

$$x = \cos A \times \cos B,$$
$$y = \sin A \times \cos B, \text{ and}$$
$$z = \sin B.$$

These coordinates are used as multipliers to produce the B-format output signals (X, Y, Z, and W) thus,

$$X = \text{input signal} \times \cos A \times \cos B,$$
$$Y = \text{input signal} \times \sin A \times \cos B,$$
$$Z = \text{input signal} \times \sin B, \text{ and}$$
$$W = \text{input signal} \times 0.707.$$

The 0.707 multiplier on W gives a more even distribution of signal levels within the four channels, and is the result of engineering considerations. This is particularly relevant to sound fields recorded with a sound field microphone or with synthesized sound fields containing many sources. These multiplying coefficients can be used to position monophonic sounds anywhere on the surface of the sound field.

It is possible to manipulate whole sound fields that contain many different sound sources in different positions, including naturally recorded ones. We have developed the following standard definitions about the way sounds move to new positions to keep the equations coherent and to minimize confusion between movement types.

positive angle of rotation — a counter-clockwise, or, by convention, leftward rotation.
rotation — a circular movement about Z-axis: the same as a counter-clockwise movement in the horizontal plane.
tilt — a rotation about the X-axis: the same as a counter-clockwise movement in the vertical left-right plane.
tumble — a rotation about the Y-axis. This is the same as a counter-clockwise movement in the vertical front-back plane. Note that a tumble is the same as a tilt rotated 90 degrees about the Z-axis.

Using these definitions it is possible, for example, to rotate the whole sound field around the Z-axis. For simplicity, consider the case of a sound field consisting of a single sound source with amplitude r positioned on the horizontal plane at an angle C from the center-front position. Given the B-format signals for the untransformed position,

$$X = r \times \cos C, \text{ and } Y = r \times \sin C.$$

If D is the angle which the sound field is rotated from its untransformed position, we have,

$$X' = r \times \cos (C + D) \text{ and}$$
$$Y' = r \times \sin (C + D),$$

where X' and Y' are the transformed B-format signals.

Simplifying and substituting X and Y,

$$X' = X \times \cos D - Y \times \sin D, \text{ and}$$
$$Y' = X \times \sin D + Y \times \cos D.$$

Since the rotation is about the Z-axis, W and Z remain unchanged. If the same procedure is applied to the tilt and tumble equations we have the following.

Tilt by Angle E

$$X' = X,$$
$$W' = W,$$
$$Y' = Y \times \cos E - Z \times \sin E, \text{ and}$$
$$Z' = Y \times \sin E + Z \times \cos E.$$

Tilt by Angle F

$$X' = X \times \cos F - Z \times \sin F,$$
$$W' = W,$$
$$Y' = Y, \text{ and}$$
$$Z' = X \times \sin F + Z \times \cos F.$$

These equations can be combined to produce complex transformations such as rotate-tilt.

Rotate-Tilt

$$X' = X \times \cos D - Y \times \sin D,$$
$$W' = W,$$
$$Y' = X \times \sin D \times \cos E + Y \times \cos D \times \cos E - Z \times \sin E, \text{ and}$$
$$Z' = X \times \sin D \times \sin E + Y \times \cos D \times \sin E + Z \times \cos E.$$

Many other combinations of movements are possible.

Ambisonics and Stereo Compatibility

The ambisonic B-format signals are not directly compatible with stereo, although it is possible to generate a response exactly equivalent to that of a crossed pair of microphones (Farrah 1979). However, there is a British two-channel system (known as UHJ) that allows most of the horizontal information from the B format W, X, and Y signals to be matrix encoded to form a standard two-channel stereo signal. The mono and stereo compatibility of UHJ recordings is very good; well-recorded UHJ provides good stereo presentation. A UHJ decoder and four or more horizontally placed loudspeakers can reproduce virtually all of the horizontal positional information contained in a full B-format signal. This involves designing wide-band, 90-degree phase shifters for both encoding and decoding (Gerzon 1977a, 1977b).

Decoding Ambisonics

Decoding ambisonically encoded signals can appear complicated. The complexity appears in the optimization of decoders for British systems (with a limited number of loudspeakers and a small listening area) that use psychoacoustic techniques, but these are not productive in systems used to cover large areas (Malham 1992). Designing decoders for large areas, such as concert halls, requires a design strategy aimed at achieving even power distribution. The distribution of loudspeakers should be as even as possible; recognized loudspeaker configurations include squares and regular hexagons for horizontal-only work, and a cube as the practical minimum for with-height work. Each individual speaker is then fed a combination of the B-format signals corresponding to its position with respect to the center of the array. For instance, for a square (horizontal only) array of four speakers, arranged left front (LF), right front (RF), left back (LB), and right back (RB), the signals are:

$$\begin{aligned}
\mathrm{LF} &= W + 0.707(X + Y),\\
\mathrm{RF} &= W + 0.707(X - Y),\\
\mathrm{LB} &= W + 0.707(-X + Y), \text{ and}\\
\mathrm{RB} &= W + 0.707(-X - Y).
\end{aligned}$$

For a cubic array, the signals for the four planar corners in the "up" (U) and "down" (D) planes are:

$$\begin{aligned}
\mathrm{LFU} &= W + 0.707(X + Y + Z),\\
\mathrm{RFU} &= W + 0.707(X - Y + Z),\\
\mathrm{LBU} &= W + 0.707(-X + Y + Z),\\
\mathrm{RBU} &= W + 0.707(-X - Y + Z),\\
\mathrm{LFD} &= W + 0.707(X + Y - Z),\\
\mathrm{RFD} &= W + 0.707(X - Y - Z),\\
\mathrm{LBD} &= W + 0.707(-X + Y - Z), \text{ and}\\
\mathrm{RBD} &= W + 0.707(-X - Y - Z).
\end{aligned}$$

The *directivity factor* of 0.707 in the equations above results in a cardioid *source directional response* for each loudspeaker. This is optimum for listening positions close to the loudspeakers or outside the loudspeaker array (Malham 1992). Where the intended listening area is significantly smaller than the speaker array, a more hypercardioid shape can be employed by increasing the directivity factor, which results in improved imaging for centrally located listeners.

Decoders based on these equations can easily be built with simple operational-amplifier circuitry, and it is possible to implement the cubic design by setting up an eight-in, eight-out mixing desk to produce suitable decoding. Loudspeakers and amplifiers used in the array should all be similar in size, response, and output.

Compositional Applications

In the past, the availability and expense of the hardware required to realize spatial compositions has often forced composers to position sound in one dimension only, and hope that further possibilities might present themselves when the piece is performed. Many composers use the careful control of reverberation, phase, and amplitude to introduce additional spatial cues within the stereo field. Some compositional expertise exists in creating stereo tapes that can recreate these effects in concert performances (Wishart 1986), but inaccuracies often occur where the acoustic conditions of the playback venue introduce additional spatial cues (Smalley 1992). This type of problem, coupled with the slow introduction of spatialization technology in domestic sound-reproduction systems, has restricted the use of space as a compositional parameter, and often discouraged composers from considering multidimensional sound localization. Electroacoustic composers recognize that sounds can contain or imply movement (Smalley 1986), but few have been able to fully explore the compositional implications and developments of this, specifically, where movement or the position of sounds is inherent in a piece and all its performances.

Our hope is that ambisonic technology will be a first step in widening composers' access to spatialization. We present below a summary of techniques appropriate to recordings encoded using this technique, along with a Csound implementation of B-format encoding and horizontal plane decoding. These can be used on computer platforms that support the output of four-channel sound files. We believe that ambisonics represents a potential standard for positional encoding techniques that will enable compositions with spatial information to be performed on a range of simple loudspeaker configurations without specialized hardware.

The compositional possibilities described here are based on the definitions of ambisonic theory (Gerzon 1972) and experiments by various composers at the University of York Department of Music. Work in ambisonics is also being conducted at the Australian Center for the Arts and Technology (Vennonen 1994).

The ambisonic formats that are currently implemented allow full three-, two- and one-dimensional reproduction, using B-format, horizontal decoding of B-format, and UHJ, respectively. The UHJ encoding can also reproduce two-dimensional soundscapes. The full three-dimensional encoding system, B-format, can be used for all implementations, as it can be converted to the lower formats.

Sound field microphones (Farrah 1979) can be used to provide three-dimensionally encoded sound-source material for composition. Mimetic source material (Emmerson 1986) and its spatial content can be captured in this way for compositional use, and it can also be combined with sound material with artificially encoded spatial information. Additionally, the X, Y, Z, and W signals can be manipulated to post-process environmental recordings (effectively changing the original microphone position); zoom into sounds within the captured landscape; or subject the entire sound image to rotation, tumbling, or other time-dependent motions.

Any source material without spatial encoding, such as synthesized sound or mono recordings, can be positioned or moved within an ambisonic sound field when subjected to the type of encoding process illustrated in the Csound examples accompanying this article. The number of sound sources reproduced within the sound field can exceed the number of loudspeakers used to reproduce it, and may be greater than the number of recorded tracks used to represent it. Each sound within a synthesized sound field can be encoded with movement. One of the great benefits of ambisonic encoding is that B-format signals can be mixed together to produce a resultant sound field that retains the positional information of all its components. Composers can manipulate the spatial path of individual sound sources, and mix them with further encoded sources to combine a series of different motions.

When sounds have been ambisonically encoded, a decoding process will reproduce the position and movement of sounds within the sound field if suitable decoder and loudspeaker configurations are used. A sound that has been encoded to move around the listener will do so in all recognized loudspeaker configurations, from a small, single-listener environment to a large concert-hall performance system. The relative position of sounds and the scale of the movement will depend on the distance between loudspeakers, i.e., the absolute sound position is not necessarily encoded, and a sound encoded to move across the front of the image will always do so despite the distance between the front speakers. Playing back a B-format signal, with an appropriate loudspeaker configuration, will automatically reproduce all encoded position and movement information.

Composers who have access to playback systems using four loudspeakers can only monitor a two-dimensional sound field plane. This type of system is capable of B-format playback, but not full with-height reproduction. It is possible, however, to monitor the horizontal and vertical planes individually for a B-format signal designed to include height encoding.

Considerations

Ambisonic systems use decoding methods that are based on physical and psychoacoustic positional information to reproduce sound fields, but there are further matters of perception, sound localization, and movement that should be considered when generating synthesized sound fields.

Our experiments with ambisonic playback systems show that the spatial perception of a sound is highly frequency dependent. Some localization of low-frequency sound is possible, but the strong positional cues are provided by higher spectral components. Also, sounds that have a widely distributed spectral energy can be localized more easily than can narrow band signals (consequently, sounds with sharp attack characteristics are usually easy to locate). There are certain conventions and expectations for the localization of high- and low-frequency sounds based on our experience of sound in the environment (Begault 1994): high-frequency sound normally occurs above us, and low-frequency below us (or at ground level). Inverting these relationships often results in poor localization, unless other cognitive sound cues exist. The acoustic response of the playback venue can produce spurious positional information that can also make localization difficult. Additionally, locating moving sounds is easier than locating stationary sounds, as with all playback systems that use phantom images.

Doppler shifts can be very important to the perception of movement (Dodge 1985). If a sound source which does not contain inherent clues to its movement is moved within an artificially constructed sound field, its perception can be impeded. Doppler shifts are not necessarily required in all circumstances, but they can assist in highlighting the movement of some sounds.

The most difficult problem in synthesizing or manipulating sound fields is the dominance of visual perception (Begault 1994). This acousmatic problem can be acute in ambisonic loudspeaker configurations, because the phantom image technique allows large angular distances to be subtended between loudspeakers at the listening position. The absence of a visual sound source is difficult for some listeners, but reducing visual dominance by lowering light levels has been shown to increase the perception of aural localization. This problem rarely occurs if the angular distance between loudspeakers is small.

Visualizing the movement of sound may also be a problem. Compositional processes that use graphic notations of sound movement can assume that visual and aural acuities are equal. Complex sound trajectories that can be visualized (and notated) are often impossible to perceive. Further assumptions about aural perception, influenced by a familiarity with multiple microphone (and multitrack) recording techniques and stereo reproduction, also produce certain expectations related to sound localization. These methods present sounds as point sources, usually within a stereo field, which is rarely a true representation of the spatial characteristics of the source. Three-dimensional recording techniques represent sound sources and their spatial positions, including diffuse sound sources.

Several experiments have shown that ambisonic sound fields played back over recognized loudspeaker configurations can be perceived from outside of the array (Malham 1992). Here the listener can "look in" on sound positions and movement, rather than being surrounded by the sound field. This has several possible implications for the integration of ambisonic and traditional sound-diffusion systems. Further work is being carried out to determine the possibilities of such systems and the potential sound field distortion effects that would be introduced by the use of independent loudspeakers. Several ambisonic speaker arrays have been tested for use as traditional diffusion systems using two-track tapes and hardware devices designed to position the two channels of a stereo source at different points within the sound field (Malham 1992). This has some advantages over placing the stereo image in specific loudspeakers, because smooth spatial transitions between loudspeakers can be achieved even if they are positioned at angles greater than 60 degrees to the listener. This method is not a parallel to the performance practice of sound diffusion artists, and does not reproduce the accuracy of sound positioning achieved by using very large arrays of loudspeakers.

Do It Yourself

The Csound orchestra and score files presented in Figures 1, 2, 3, and 5 enable simple ambisonic encoding and decoding with computer systems that support the playback of four-channel audio files. The files produce signals that can be fed directly to four amplifiers and loudspeakers to reproduce ambisonically encoded, two-dimensional horizontal sound fields. The encoding example is based on the unit sphere, and enables sound to be positioned and moved on its surface. It does not include the algorithms required to move sounds through the center of the sound field.

The Csound Examples

In the Csound ambisonic encoding orchestra file of Figure 1, the position of a mono sound source within the sound field is described by the horizontal and vertical angles (in radians) it subtends at the listening position. These angles are represented by parameters specified by the user in the score file of Figure 3. Any score-file event can be assigned either a static position or a linear movement within the sound field. Both the horizontal and vertical angles can be assigned "start" and "end" positions (parameters `p6` and `p8`, and parameters `p7` and `p9`, respectively). These parameters are passed from the score file to the orchestra file (Figure 1) during a Csound performance. The encoding orchestra uses a linear function (line) to interpolate between both start and end angles over the duration of each score event, producing two time-varying angles: `kone` and `ktwo`. If the `start` and `end` angles in the score file are equal, the sound will remain stationary.

The orchestra calculates the sine and cosine of the time-varying angles, which are then applied to the ambisonic B-format encoding equations, producing the audio-rate variables `ax`, `ay`, `az`, and `aw`. The sound source for these examples is created by a Csound generator routine which produces a composite waveform consisting of weighted sums of sinusoids (additive synthesis). This routine is called by the first line of the score file, and generates a wavetable that is read by the interpolating oscillator in the orchestra file. The output of the oscillator, variable `a5`, is passed to the Ambisonic encoding equations, where it is multiplied by the (varying) coordinates for the intended sound position or movement. The four audio signals produced by the encoding equations are written out to a sound file by the `outq` routine. The output from Figure 1 is an ambisonic B-format signal. If the B-format file is to be used in conjunction with the

Figure 1. A Csound orchestra definition for encoding B-format ambisonic data.

```
;*************************************
;* Ambisonic Encoding Orchestra       *
;*                                    *
;*  p4 = pitch; p5 = amplitude;       *
;*  p6 = start angle from center front *
;*  p7 = end angle from center front  *
;*  p8 = start angle from horizontal  *
;*  p9 = end angle from horizontal    *
;*                                    *
;*(all angles are expressed in radians)*
;*************************************

; Csound orchestra header
sr = 44100
kr = 441
ksmps = 100
nchnls = 4

instr 1
   ; linear transformation
   ; between start and end angles
  kone line p6, p3, p7
  ktwo line p8, p3, p9

   ; envelope for sound source
  kenv linen p5, 0.008, p3, 0.02
  a5 oscili kenv, cpspch(p4), 1

   ; calculate cos and sin of
   ; time-varying angles
  kca = cos(kone)
  ksa = sin(kone)
  kcb = cos(ktwo)
  ksb = sin(ktwo)

   ; B-format encoding equations
  ax = a5 * kca * kcb
  ay = a5 * ksa * kcb
  az = a5 * ksb
  aw = a5 * 0.707

   ; B-format output
  outq ax, ay, az, aw

endin
```

Figure 2. A Csound orchestra definition for decoding sound files encoded with B-format ambisonic data.

```
;*************************************
;*                                   *
;*   Ambisonic Decoding Orchestra    *
;*                                   *
;*************************************


sr = 44100
kr = 441
ksmps = 100
nchnls = 4

instr 2
   ; Read sound in from file
  ax, ay, az, aw    soundin 1

   ; Decode equations producing 4
   ; speaker feed signals for
   ; horizontal-only playback
  a1 = aw + (ax*0.707) + (ay*0.707)
  a2 = aw + (ax*0.707) - (ay*0.707)
  a3 = aw - (ax*0.707) - (ay*0.707)
  a4 = aw - (ax*0.707) + (ay*0.707)

   ; output speaker signals
  outq a1, a2, a3, a4

endin
```

decoding orchestra of Figure 2, the output sound file name should be `soundin.1`. Any mono audio signal can replace the synthesized sound source of Figure 1 (`a5`).

The score file of Figure 3 gives three examples of a circular movement (counter-clockwise) around the sound field: A slow rotation from center front, over 30 sec; a quicker rotation lasting for 5 sec; and the rotation of a repeating short sound. These examples are intended to illustrate movement in the horizontal plane, so no angles above or below the horizontal plane are given. The vertical angles can be used to produce height-encoded B-format, but this information will not be decoded by the orchestra file of Figure 2. For this simple illustration of movement, the pitch of the sound source remains constant. The score file shows how parameters can be passed to interpolating functions in the orchestra file, and the use of the score based linear "ramping" function to produce intermediate sound positions.

Figure 3. A Csound score file to demonstrate the encoding instrument of Figure 1.

```
;*************************************
;*                                   *
;*   Ambisonic Encoding Score        *
;*                                   *
;* p4 = pitch; p5 = amplitude;       *
;* p6 = start angle from center front *
;* p7 = end angle from center front  *
;* p8 = start angle from horizontal  *
;* p9 = end angle from horizontal    *
;*                                   *
;*(all angles are expressed in radians)*
;*                                   *
;*************************************


; Function table
f1 0 1024 10 1 .7 .7 .7 .7 .7 .7 .7 .7 .7
       .4 .4 .4.4 .3 .3 .3 .3 .2 .2 .2 .2

; Note commands
;p1 p2  p3   p4   p5    p6     p7     p8 p9
i1 0.0 30.0 5.07 15000 0      6.2832 0  0
i1 +   5.0  .    .     .      .      .  .
i1 +   0.16 .    .     0      0      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     >      >      .  .
i1 +   .    .    .     6.2832 6.2832 .  .
; End
e
```

Figure 2 is a Csound orchestra that decodes a four-channel B-format signal to produce speaker-feed signals for a horizontal loudspeaker configuration. A B-format sound file is taken into the orchestra as a `soundin` file. Each of the four channels is applied to the ambisonic decoding equations and

Figure 4. Loudspeaker arrangement for playing back four-channel ambisonic sounds.

Figure 5. A Csound score file to demonstrate the decoding instrument of Figure 2.

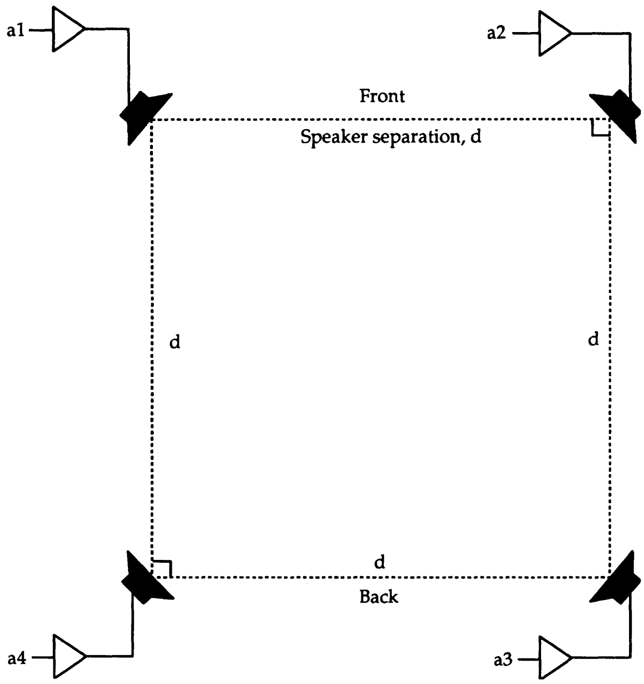

Figure 4

```
;***************************************
;*                                     *
;*      Ambisonic Decode Score.        *
;*                                     *
;***************************************

; Play a test note on instrument 2
i2    0    38

; End
e
```

Figure 5

scaled by the directivity factor of 0.707 to correctly position the sounds. The four speaker signals produced, `a1`, `a2`, `a3`, and `a4`, are written to a sound file that can then be played over the speaker configuration of Figure 4. The score file (Figure 5) that accompanies this orchestra acts as a simple switch to trigger the decoding instrument. Care should be taken to ensure that the duration in this score, parameter 3, is equivalent to (or greater than) the input sound file length.

The examples presented here demonstrate simple sound positioning. More complex interpolating functions could be applied to describe sound paths within the sound field. The encoding functions can be added to any Csound instrument if ambisonic B-format spatialization is required. The decoding orchestra file can be applied to any B-format signal (including recordings made with the sound field microphone) to produce horizontal loudspeaker feeds. No additional hardware is required.

Two stereo reproduction systems, with approximately equal gain, can be used for the loudspeaker configuration of Figure 4 (it is useful to produce a four-channel sound file to test the amplitude and position of each loudspeaker). The loudspeakers should be placed accurately at the four corners of a square.

Full three-dimensional sound field decoding from B-format signals is possible using computer systems that support eight audio channels, or by employing a hardware decoder.

Conclusions

Our experimental work has been carried out using both software and hardware ambisonic implementations, including a purpose-built programmable periphonic decoder that produces full three-dimensional surround sound over 16 loudspeakers. The Csound examples are presented here to enable other users to experiment with ambisonic encoding and decoding, without using specialized hardware devices. The simplicity with which ambisonic encoding can represent sound sources in three-dimensional space is a great advantage, and some of the possibilities for sound field manipulation and generation are almost unique to this technique. There is also the potential to convert ambisonic signals to other formats, e.g., binaural, for other applications (Malham 1993).

The techniques described here only refer to first-order ambisonic encoding. The original work by Gerzon (Gerzon 1972) includes descriptions of higher-order systems that can increase the spatial accuracy of both recording and playback.

There are many musical and psychoacoustic issues that require further investigation. We hope that the information presented here will encourage progress in these areas.

Further information can be obtained on the World-Wide Web ambisonic home page, via the URL "http://www.york.ac.uk/insts/mustech/3d_sound/ambison.htm."

References

Askew, A. 1981. "The Amazing Clément Ader." *Studio Sound* 23:9–11.

Begault, Durand R. 1994. *3-D Sound for Virtual Reality and Multimedia*. Boston: Academic Press. pp. 65–66, 84, and 191–245.

Chowning, J. 1971. "The Simulation of Moving Sound Sources." *Journal of the Audio Engineering Society* 19(1):2–6. Reprinted in *Computer Music Journal* 1(3):48–52.

Cooper, D. H., and J. L. Bauck. 1989. "Prospects for Transaural Recording." *Journal of the Audio Engineering Society* 37(1/2):3–19.

Cooper, D. H., and T. Shiga. 1972. "Discrete Matrix Multi-Channel Stereo." *Journal of the Audio Engineering Society* 20(5):346–360.

Dodge, C., and T. A. Jerse. 1985. *Computer Music*. New York: Schirmer Books. pp. 245–247.

Emmerson, S. 1986 "The Relation of Language to Materials. In *The Language of Electroacoustic Music*. London: MacMillan, pp. 17–39.

Farrah, K. 1979 "The SoundField Microphone." *Wireless World*. November: 99–103.

Fox, B. 1982. "Early Stereo Recording." *Studio Sound* 24(5):36–42.

Gerzon, M. A. 1972. "Periphony: With-Height Sound Reproduction." *Journal of the Audio Engineering Society* 21(1):2–10.

Gerzon, M. A. 1974 "Surround-Sound Psychoacoustics." *Wireless World*. December: 484.

Gerzon, M. A. 1977a. "Design of Ambisonic Decoders for Multi Speaker Surround Sound." Paper presented at the 58th Audio Engineering Society Convention, 4 November, New York.

Gerzon, M. A. 1977b. "Surround Sound Decoders" (7 parts). *Wireless World*. January to August issues, 1977.

Hope, Adrian. 1979. "Fantasia-Multitracked." *Studio Sound* 21(8):29–30.

Malham, D. G. 1992. "Experience with Large Area 3-D Ambisonic Sound Systems." In *Proceedings of the Institute of Acoustics* 14(5):209–216. St. Albans: Institute of Acoustics.

Malham, D. G. 1993. "3-D Sound for Virtual Reality using Ambisonic Techniques." In *Proceedings of the 3rd Annual Conference on Virtual Reality* (addendum). Westport: Meckler.

Malham, D. G., and R. O. Orton. 1991. "Progress in the Application of 3-Dimensional Ambisonic Sound Systems to Computer Music." In Proceedings of the International Computer Music Conference (ICMC). Montreal: ICMC, pp. 467–470.

Moore, F. R. 1983. "A General Model for Spatial Processing of Sounds." *Computer Music Journal* 7(3):6–15.

Moorer, J. A. 1979 "About this Reverberation Business." *Computer Music Journal* 3(2):13–28.

Sanal, A. J. 1976. "Looking Backward." *Journal of the Audio Engineering Society* 24(10):832.

Schaeffer, P. 1951. "Journal d'Orphée." In F. Bayle, ed., *Pierre Schaeffer l'œuvre musicale*. France 1990. Paris: INA/GRM and Librairie SEGUIER.

Schroeder, M. R. 1962. "Natural Sounding Artificial Reverberation." *Journal of the Audio Engineering Society* 10(3):219–223.

Smalley, D. 1986 "Spectro-Morphology." In S. Emmerson, ed., *The Language of Electroacoustic Music*. London: MacMillan, pp. 73–80.

Smalley, D. 1992 "The Listening Imagination." In H. O. Paynter, et al., *Companion to Contemporary Musical Thought*, vol 1. London: Routledge.

Stimson, Ann, 1991. "The Script for *Poème Électronique*; Traces from a Pioneer." Proceedings of the International Computer Music Conference. Montreal: ICMC, pp. 308–310.

Stockhausen, K. 1956. "Programme Notes for the 1956 World Premiere of *Gesang Der Jünglinge*." In liner notes for the 1992 CD *Stockhausen 3 Elektonische Musik 1952–1960* Stockhausen: Verlag.

Thiele, G., and G. Plenge. 1977. "Localization of Lateral Phantom Sources." *Journal of the Audio Engineering Society* 25(4):196–200.

Vennonen, K. 1994. "A Practical System for Three-Dimensional Sound Projection." In *Proceedings of the Symposium on Computer Animation and Computer Music*. Canberra, Australia: Australian Centre for the Arts and Technology.

Weiland, F. C. 1975. "Electronic Music—Musical Aspects of the Electronic Medium." Internal publication, Institute of Sonology, Utrecht State University.

Wishart, T. 1986 "Sound Symbols and Landscapes." In S. Emmerson, ed., *The Language of Electroacoustic Music*. London: MacMillan, p. 45.