The Spatial Soundscape

James L. Barbour
Swinburne University of Technology, Melbourne, Australia
jbarbour@swin.edu.au

Abstract
While many people have sought to capture and document sounds for posterity, particularly those considered in danger of extinction, the author is concerned with the process of capturing the spatial context along with the sound, archiving the sound with its spatial identity intact. This paper will consider characteristics of human auditory perception and our ability to perceive the location and distance of any sound source. Information will be provided about techniques available for the capture, archiving and reproduction of spatial audio, and will also consider how soundscape recreation or composition might produce an immersive, three-dimensional soundfield for any listening environment and how this soundfield may be delivered to a new listening audience.

Introduction
All sounds exist in a three dimensional acoustic space, and there is great pleasure derived by the author simply by immersion in the sonic environment, aware of the totality of the soundscape but revelling in the detail of each individual sound: its location, character, movement and evolution. For every sound within the soundscape, its spatial context is an integral part of its identity. Sound isolated from its location in the soundscape loses a significant part of its meaning, it loses its relationship to other sound events and it loses any influence that one sound may have purely because of its location. It is therefore important to consider how we might document the spatial context of every auditory event. This documentation may include an attempt to describe the sound’s location and relationship with other sounds or objects using words or diagrams. However, words and diagrams are poor tools to classify the ephemeral nature of sound itself, let alone its three dimensional location and evolution over time. Recording the sound for posterity is a valuable documentation, provided the spatial context is also captured, embedded in the recording. This paper will consider how we may document the spatial context of sounds, archive them for posterity and also how we might recreate for a listener the original sound with its spatial context intact.

Aural Perception
But first, let’s briefly review aural perception. We live in a three-dimensional immersive soundfield. We can hear sounds emanating from all around us, out to our acoustic horizon; an invisible acoustic space. The acoustic horizon is continuously fluid, changing its dimensions according to the loudness of all sounds at any moment. Close, loud sounds cause the horizon to contract to our immediate environment; in fact, our own speech causes the acoustic space to contract to our body! Our ears can detect sounds emanating from a multiplicity of directions simultaneously, and the totality of these directions defines this acoustic space. Our ability to perceive the identity, location and distance of a sound source can be aided by visual cues, but is built on binaural discrimination, the differences in perception between our two ears, and our experience of the variations in the loudness and timbral qualities of a sound source at different distances. One of the most important aids to localisation is the initial arrival of the sound pressure wave, known as the onset transient, which provides information for binaural time discrimination and contains temporal changes in timbre.

Binaural Discrimination
Blauert [3] described the head-related system of spherical co-ordinates using the horizontal plane, frontal plane and median plane, with angles of incidence for azimuth and elevation, Figure 1. Sound source localization in the horizontal plane and the frontal plane primarily relies on binaural discrimination due to inter-aural amplitude, inter-aural time and inter-aural spectral differences, with the exception of the median plane where inter-aural differences are zero for a normal listener. Research into localization in the median plane, [3], concluded that three requisites were necessary for an auditory stimulus to be accurately located in a vertical space: (a) the sound must be complex, (b) the complex sound must include frequencies above 7000 Hz, and (c) the pinna must be present. Our ability to localize sound sources on the median plane relies primarily on...
pinna effects. According to Blauert, the curves and ridges within the pinna reflect different frequencies depending on source elevation.

![Figure 1: Three-dimensional planes (after Blauert)](image)

**Acoustic Mapping**

Depending on the distances between the source, the listener and any reflective surfaces, we will perceive characteristics of the size, shape and material construction of the surfaces and the size and shape of the acoustic space itself. Complex patterns of early and late acoustic reflections will surround and envelop us in a natural soundfield. The aural envelopment we perceive is aided by our ability to localize the original sound source and our ability to perceive the aural spaciousness with no particular direction. We have learned to interpret the sounds we hear and the space in which we hear them through the interaction of all our senses, building auditory memories. Human perception of acoustic environments is learned through a lifetime of experience of moving through different spaces and acoustically mapping the auditory characteristics. In the *Journal of the Acoustical Society of America* in 1974, Plenge [6] identified our ability to perceive acoustic environments, noting that:

> Only a few seconds of listening was sufficient for calibration of a room's acoustic properties, which were stored for as long as the listener stayed in the room, and then cleared immediately upon leaving, so that the listener could recalibrate at once for a new acoustic environment. (Plenge, 1974, 44)

Two particular characteristics of a natural soundfield that we need to consider are coherent imaging and homogeneous directionality. Within a natural soundfield, as a listener changes position, the perceived location of the sound source relative to the listener will remain stable or coherent, with no significant discontinuity. The direction, distance and envelopment will change, but the change will be smooth and remain perceptually consistent with the motion of the listener or source. For a soundfield to be homogeneous, it places no preference on a particular direction from which sound may originate. We are able to perceive sounds coming from any direction around, above or below us. While we are very familiar with locating sounds around us on the horizontal plane, sounds above and below us are less readily identified.

**An Acoustic Audit of Height**

In our everyday environment, there are many sound sources with a significant angle of elevation. The author has had a particular interest in height perception and has been conducting detailed listening to establish an inventory of real sounds with elevation characteristics, based on Acoustic Ecology principles articulated by Schaffer [8], Truax [10] and others. When specifically considering elevated sound sources, there are natural
environments like open fields, forests, caves and seascapes, and constructed environments including city and suburban streets, under bridges, inside buildings and underground tunnels and trains. Within each environment, there are natural elevated sources like wind effects, flora and fauna, and human constructed sources like flying machines, tall ground-based objects, ceiling mounted loudspeakers and many more. Also important are the acoustic reflections from the upper reaches of halls, cathedrals and concert venues where aural events are experienced and recorded. An extensive listing of elevated sound sources is beyond the scope of this paper, but a few common examples in Table 1 will illustrate key characteristics.

### Elevated sound sources in the Open:

#### Aeroplanes:
- very high: low frequency spectrum drone, slow and small Doppler shift, sound well behind visual position, some localization blur
- low, barely above roofline, eg helicopter: high frequencies clear, sharp transients, rapid and wide Doppler shift, distance effects of amplitude and frequency spectrum very clear, precise localization

#### Birds:
- unusual to be high, eg crows: low frequency spectrum of calls, slow change in position and sound
- flying low: aurally appear and disappear very quickly, rapid Doppler shift, wing flapping heard when very close, precise localization
- Stationary in trees: bright spectrum of calls, precise localization

#### Inside buildings:
- footfall above and floor creaking: low frequency spectrum, poor localization
- fluids in utility pipes: low frequency spectrum, possible localization
- rain on the roof: dull, muffled sound, no specific location

#### Performance venues:
- direct sound from high galleries at the side, front or rear: clear full spectrum, precise localization
- acoustic reflections: amplitude and frequency spectrum highly variable, imprecise localization, distinct echoes occasionally heard

<table>
<thead>
<tr>
<th>Table 1: characteristics of elevated sound sources</th>
</tr>
</thead>
</table>

Key factors identified in locating elevated sources in everyday environments are fundamentally the same as for ground based or horizontal sources. These include binaural discrimination, (inter-aural time, inter-aural amplitude and inter-aural spectral differences along with torso influences), distance effects of amplitude, spectral changes and the direct to reverberation ratio, and motion effects of Doppler pitch shifts, amplitude and spectral changes. The principle additional factor utilized for elevation perception is pinna effects. As identified by the experimental tests conducted by the author, when an elevated source lies on the median plane and would present difficulties in accurate location, a listener would further interrogate the auditory scene for additional information to resolve the location, for example, head turning.

### Deconstruction and Recreation

While many acoustic ecologists have sought to capture natural sounds for posterity, particularly those sounds considered in danger of extinction, the author is raising the importance of capturing the spatial context along with the sound, archiving the sound with its spatial identity intact. The reproduction of the sounds in their true three dimensional perspective is an ancillary issue. For many listeners, the recreation of a realistic immersive soundscape is a goal, although the perfect recreation is still a technological impossibility. How might we document sounds in acoustic space in a way which will preserve their spatial relationships and allow for a virtual recreation?

### Three Dimensional Soundfield Recording and Reproduction
Binaural microphone recording techniques utilize two omni-directional microphones mounted at the entrance to the ear canal on both sides of a human head, or a dummy head with moulded pinnae. The sounds thus recorded incorporate realistic spatial characteristics but their spatial qualities are only audible with headphone listening. Recording natural sounds for loudspeaker reproduction requires very different techniques, involving the careful selection of microphones and their placement. There are numerous microphone techniques devised and tested for recording aural events, described in books, magazines, on websites and in conference proceedings, for example, Holman [4]. Depending on the skill and taste of the recording engineer, some incorporate coincident positioning of microphone capsules while others prefer spaced microphone arrangements. For multi-channel recording incorporating spatial information, additional microphones are required to capture spatial ambient information from the environment. Consequently, many multi-channel recordings require some microphones to be relatively close to the source for clarity in reproducing its sonic identity, while other microphones are positioned placed further from the source to record spatial information from the sonic environment. If a true periphonic reproduction is required, additional microphones should be employed to capture height information, for reproduction from overhead loudspeakers. The use of multiple microphones separately recorded and used in recreating an aural environment introduces many technical and artistic problems due to temporal, spectral, amplitude and phase distortions, which adversely affect the soundfield coherence and homogeneity. This is further exacerbated if the listener is able to move within the recreated soundfield, where source location may become blurred or incorrect and true spatial envelopment may not be preserved. The Ambisonic technique for recording and reproduction attempts to overcome many of these problems using a special microphone and loudspeaker arrangement.

**Ambisonic Recording and Reproduction**

The Ambisonic concept is an extension of the original Blumlein stereo pair technique, and involves measuring a soundfield at a single point in space as four component mono recordings. The W component represents an omni-directional recording at this point. The X component is a bi-directional recording with positive phase facing zero degrees forward, and negative phase at 180 degrees behind. This recording has minimum pick-up from left-right and above-below. The Y component is a bi-directional recording with positive phase facing left at 90 degrees, with minimum pick-up from front-back and above-below. The Z component is a bi-directional recording with positive phase directly up, with minimum pick-up from front-rear and left-right. The diagram below shows these four components.

![Ambisonic microphone arrangement](image)

Figure 2: Ambisonic microphone arrangement
Auditory Scene Analysis

Thus, any sound source at any location in space will have its acoustic output recorded by the W, X, Y and Z microphones in unique proportions, depending on its location relative to the point of recording. One version of the combined microphones is called the Soundfield Microphone™ [9], and the output is four analogue mono signals. Using auditory scene analysis, we can break up any sound environment into the total number of individual sound streams, each generated by a sound source. The Soundfield Microphone will simultaneously encode each sound stream into its WXYZ components and the total output from the microphone will be an encoded sound field, called B-format. To recreate the original soundfield, we specify the location of all loudspeakers used for reproduction as XYZ directions, and the decoder will then assign the correct proportions of the recorded B-format signals to each loudspeaker. For a more detailed explanation of what is a very complex process, please consult the Ambisonics website [1]. In this way, a soundfield can be recorded and reconstructed with great clarity, though the number of loudspeakers required is large, with a cube example of eight loudspeakers considered to be the minimum necessary. The loudspeaker array should be regular, with each loudspeaker matched by another in the opposite position measured through the listening position. While the Soundfield microphone does measure spatial information accurately, it is not well liked by many recording engineers because of the inherent compromise between close recording and spatial recording: if it is close enough to accurately record the direct signal, it is not in the ideal position to record the ambient field. Consequently, recording engineers are supplementing a Soundfield microphone with either close microphones to enhance the direct sound, or distant microphones to capture the ambient sound.

Ambisonic Production

A particular advantage of the Ambisonic approach to recording is the preservation of the spatial information in a format that can be adapted to any listening environment, present or future, with a relatively few number of recorded channels of information. This flexibility is ideal for the archiving of endangered sounds for posterity, allowing the spatial context to be preserved for future listeners. The recording technique is also very robust in allowing sounds recorded in close proximity with a single microphone to be mixed with an Ambisonically recorded scene, with the close sound panned Ambisonically to its original perspective. This approach using an Ambisonic recording along with individual close recordings is attracting much attention currently and is applicable to loudspeaker reproduction in homes, museums, zoos, theme parks, in fact, any indoor or outdoor listening environment. The Ambisonic technique captures height information accurately, allowing for loudspeakers to be used overhead to create a truly immersive, three-dimensional soundfield, and it is also successful in reconstructing motion perception for a moving sound source. One benefit of Ambisonic production is to effectively recreate soundscapes that may be difficult to capture, perhaps due to loud, close sounds or irregular intervals between sonic events. Individual sounds may be recorded Ambisonically, preserving their spatial context, then combined in post-production to create the desired soundscape. An example of a simulated soundscape is considered by Truax [10]:

Another type of simulated soundscape is found in the ‘Entry to the Harbour’ sequence in the ‘Vancouver Soundscape’ document. It comprises the various foghorns and other sounds one would encounter on a ship passing from the outer to the inner harbour in Vancouver, then docking. The motor sound that obscures an actual recording of the journey necessitated each sound being recorded separately and mixed with the appropriate spatial illusion of the sound approaching and receding. (Truax, 2001, p239)

The spatial illusion required for each separate sound desired could have been embedded in individual Ambisonic recordings and combined in the production process. Despite its apparent robust nature, Ambisonic recording and reproduction does have some limitations, particularly with the reproduction loudspeaker array. The ideal listening position, or sweet spot, is at the very centre of the three-dimensional array and is the point at which the constructed soundfield is most accurate. While small movements away from this position lead to a gradual breakdown of the coherency of the soundfield, even off to the sides there is still a perception of the soundfield existing out to, and even beyond the physical boundaries of the array. This is a major improvement over any other form of loudspeaker reproduction format. Another significant difficulty highlighted by some
users is a relatively loud self-noise, which may interfere with quiet sounds. However, using Ambisonic techniques we are able to accurately record sounds with their spatial context intact, to archive those recordings in a format which is future proof and to recreate a realistic copy of the original soundfield. Is it also possible to offer this recreation to listeners in their own home environment, and thereby extend the opportunity for the Acoustic Ecology community to offer our perspective on sonic environments to a broader audience?

Soundscapes at Home

An immersive soundfield could not be delivered into listener’s homes with any significant success until the digitisation of video and audio permitted the introduction of the Digital Versatile Disc (DVD) format in the mid 1990’s, using software formats including Dolby Digital (also referred to as Dolby AC-3) and Digital Theatre Systems (DTS), allowing surround sound reproduction for cinemas and home theatres. The technical specifications for DVD-video include the audio standard ITU-R BS.775 (ITU 1993), known as 5.1 surround sound. This utilises five full range loudspeakers distributed around a listener and a low frequency sub-woofer loudspeaker, as shown in the accompanying diagram.

ITU-R BS.775 5.1 Surround specification

The growth in sales of home theatre surround sound systems is evidence of the consumer acceptance of this new delivery and listening platform, which allows the delivery of high quality, multi-channel sound and music without images into any listening environment. However, audio researcher Francis Rumsey, author of the book *Spatial Audio* [7] and chairperson of the committee that developed the ITU 5.1 standard, identified a problem for the audio community:

> Although ‘purist’ sound engineers find it hard to accept that they must use a layout intended for movie reproduction, ….. most pragmatists realise that they are unlikely to succeed in getting a separate approach adopted for audio-only purposes and that they are best advised to compromise on what appears to be the best chance for a generation of enhancing the spatial listening experience for a large number of people. (Rumsey, 2001, 18)

This pragmatic approach is also adaptable to the needs of the Acoustic Ecology community, particularly with newer high resolution audio delivery formats.
A High Quality Immersive Soundfield
Two new delivery formats have been introduced in recent years, DVD-Audio and SACD, and both are capable of carrying six channels of high-resolution uncompressed audio. DVD-Audio is an extension of the DVD specification first made commercially successful with the DVD-Video application. While DVD-Video uses Dolby Digital AC-3 or DTS compression to deliver 5.1 audio, DVD-Audio uses virtually the entire bandwidth of the medium to deliver six channels of very high quality multi-channel audio. DVD-Audio may also carry images, song lyrics, band or recording information, but there is not sufficient bandwidth in the data transfer rates available for simultaneous video. SACD is a competing format to DVD-Audio, using a different data recording technology, and is also capable of delivering six channels of high-resolution audio. SACD has limited capacity for images and no video. While there is slow consumer adoption of both formats and few releases available locally, the potential problem of two competing formats is being addressed by the latest generation of multi-format or universal disc players, capable of playing DVD-Audio and DVD-Video, SACD, standard CD and mp3 CD. Inexpensive authoring software is available for DVD-Audio, so it is possible for soundfield recordists to use the DVD-Audio medium to deliver spatial environments to listeners using up to six channels of high quality audio. The author has created environments using Ambisonic techniques, mixed them into the 5.1 surround format and delivered them via a DTS encoded DVD video disc, suitable for reproduction in a standard home theatre listening room. A high quality, immersive soundfield created this way may be experienced and enjoyed by many listeners.

Conclusion
All sounds exist in a three dimensional acoustic space, with their spatial context an integral part of their identity. We have a well developed ability to perceive the location and distance of any sound source, and clear auditory memories of the acoustic character of internal and external spaces. Our hearing physiology has no preference for any particular direction, but our brain can accurately determine source location around, above and below us. Sound isolated from its location in the soundscape loses a significant part of its meaning, and any attempt to document sounds through recording must capture the spatial context along with the sound, archiving the sound with its spatial identity intact. Techniques for recording and reproducing a realistic three-dimensional soundscape include binaural and Ambisonic microphone techniques, both of which will preserve the spatial identity for future reproduction formats. Soundscape recreation or composition can produce an immersive, three-dimensional soundfield for any listening environment, and readily available home audio technology can deliver this soundfield to a new listening audience.

References