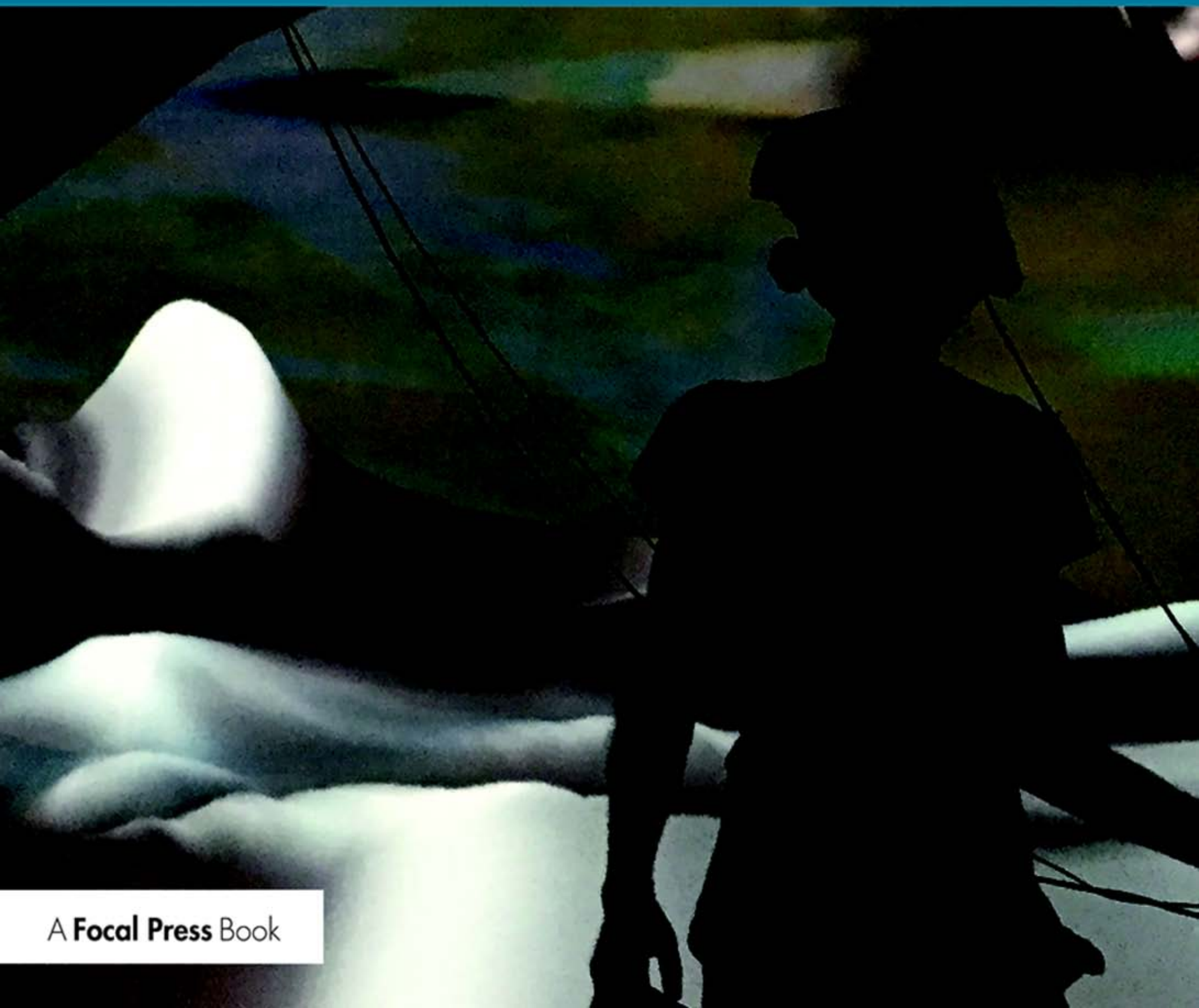
SOUND DESIGN

FOUNDATIONS IN SOUND DESIGN FOR INTERACTIVE MEDIA

A Multidisciplinary Approach

Edited by MICHAEL FILIMOWICZ



A Focal Press Book

7

Sound Spatialization

Enda Bates, Brian Bridges and Adam Melvin

7.1. Introduction: Sound as Spatial

> . . . you are *in* the sound, the space is actually the sound, and you're this slug with a few discs that's quivering a bit.
>
> (Brophy 2014)

In a recent panel discussion on the future of sound and music, composer/director Philip Brophy provided a humorous, yet eloquent, explanation of his own attraction to sound over visual stimuli. For Brophy, it is the "totality" of sound—that fact that it is "360 degree spherical in its diffusion, its apparition, its aura [and] its presence"—that enables sound to overwhelm, making it "impossibly social" and "impossibly engaging."

It is the parameter of space, and its fundamental importance to sound installation, with which this chapter is concerned. Sound, as a perceptual entity, has two dominant dimensions of expression: (1) the temporal development of actions, and (2) the *space* (physical and acoustic) those actions "inhabit." The spatial aspect be can implied, as in the virtual "space" of pitch relationships that music often articulates, or it may be physically present, as in acoustic spaces and environments and the differentiating positions of objects within them. As Labelle (2003) notes, "sound operates acoustically within space, gaining definition in relation to the space in which it is heard and, in turn, lending definition to spatial organization through sonic agitation."

7.1.1. Spatiality and Sound Installations

In sound-based installation practices, the parameter of space acquires further agency, both as a creative tool, and within a performative context: "the place of sound"—one might also add, the place-*ment* of sound—"becomes as much a part of auditory experience as the material of sound itself."

(Labelle 2006, p. 197). While they may have some element of temporal development, sound installations are predominantly spatial forms, through the combination of materials they incorporate and the framing of perspectives within a defined space.

In an installation context, sound harnesses this existing sense of spatiality, reinforcing the spatial impressions offered by the visual perception of the installation space, or (perhaps, more typically) subverting or expanding them. Beyond the choice and placement of speakers or other sounding mechanisms, the type of sound materials themselves (i.e. their frequency content, amplitude and how this evolves over time) may also change how audiences perceive the characteristics of a sonic space or the sound materials within its frame. In addition, the participant's own interaction, and the interposition of sounds from the wider environment, may also affect the auditory spatial perspective within an installation. This chapter will address some of these key attributes and parameters of sound and space in installation design.

7.1.2. Spatial Audio Cues: The Spatial Affordances of Sound

Spatial sound could be said to have two main dimensional attributes or cases: (1) the sense of spatial "presence" versus "absence" (or defined "location" versus indistinct "immersion") that derives from a sense of relative distance within a spatial field, and (2) its particular angular direction on a plane or within a 360-degree sphere surrounding a listener. Thus, we have localization *direction*, in terms of angles of *azimuth* and/or *elevation* (the horizontal and vertical planes), and a sense of *distance* (near/far), or perceptual clarity/distinctiveness standing as a proxy for this. The majority of what we can easily control about the sonic expression of space will relate to directional aspects of sound, at least in terms of a single source, moving with respect to a listener. However, when multiple sources are present, or when particular frequency ranges are emphasized/de-emphasized within materials, the second attribute may also become an accessible avenue for creative exploration.

Regarding direction, as a species, humans are able to compare the signals received at our two ears (*inter-aural cues*). Two aspects of the sound waves from an incoming auditory event are significant here: the differentiation between time of arrival (*interaural time difference*, or ITD), and level (*interaural level difference*, or ILD). The ITD is caused by the spacing between the two ears and a speed of sound that is relatively slow in comparison with other physical phenomena (often taken to be 340 meters per second at a given standard temperature at approximately 15 ^{o}C). ILD is caused by a *head shadowing* effect in sound, with higher frequencies (and, hence, shorter wavelengths). Any sound wave that contains frequency components which give rise to short wavelengths (and, hence, a small physical scale when propagating as waves) are reflected and diffracted by a bigger object such as the human head. As a result, a sound that comes from directly in front of the listener will arrive at each ear at the same time (ITD), and at the same level (ILD), whereas a sound coming from the right side will arrive earlier (ITD), and at a higher level in the right ear compared to the left (ILD); see Figure 7.1, below.

Certain listening conditions or certain types of materials may subvert these cues. The ITD for a given sound source will be same whether a sound source is in front or behind a listener. The *cone of confusion* is the region of a set of points that could produce the same interaural time differences (other cues are needed to resolve this confusion).

The frequency range of sound materials may also impact on the efficacy of the basic ITD and ILD cues themselves. The ILD head shadowing effect may be pronounced enough to lead to as much as a 10-dB reduction within ranges around 4 kHz (Hartmann 2013, p. 146). The process behind ITD cues will only operate up to approximately1.4 kHz for relatively steady state tones. Beyond this range, envelope or transient characteristics may facilitate the operation of this cue (ibid. p. 148); it may also be subverted through the presence of very low tones.

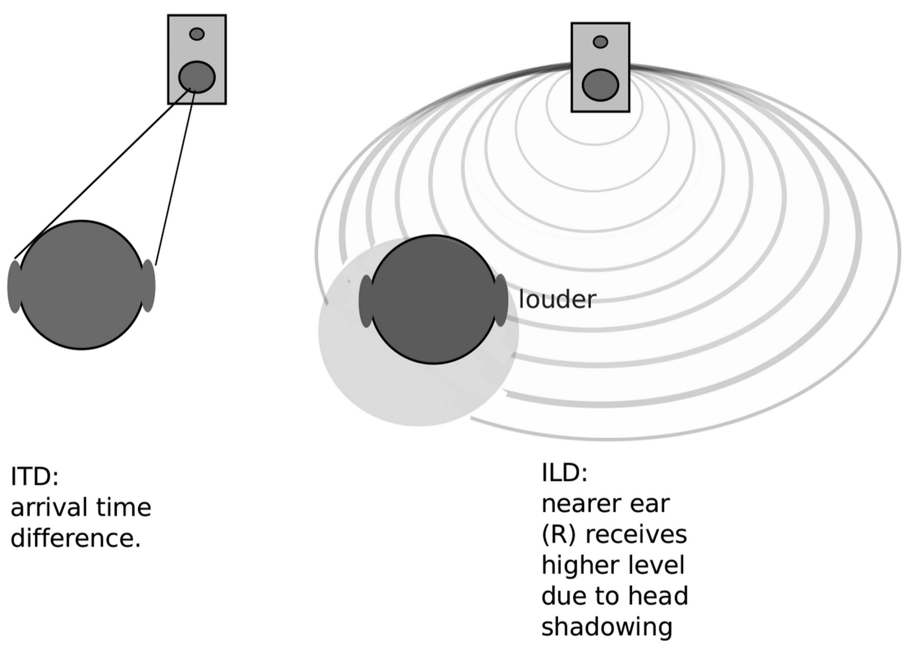


Figure 7.1 ITD: inter-aural time difference, and ILD: inter-aural level difference, compared.

Because of these limitations, sustained tones of between 1.4 and 4 kHz without significant harmonic content may be much more difficult to localize. Furthermore, sources whose frequencies coincide with the resonances of the ear canal (between approximately 2.5 and 4 kHz) may produce a perception of an extremely close, almost internal, source, evoking sonic intimacy or claustrophobia. To summarize, sound materials that contain a combination of significant transients or pronounced amplitude variation over the course of an envelope will facilitate ITD; ILD will be facilitated by the presence of higher frequency content; lower, sustained tones (e.g. drones) may be more difficult to localize.

The second component of spatial sound is the apparent distance of sources from the listener. The most obvious distance cue is amplitude, which gradually decreases as the distance to the listener increases. However, amplitude alone can only serve as a distance cue if the sound is highly familiar, such as a voice (e.g. we can easily distinguish between someone shouting far away, and someone whispering nearby, even if the amplitude of the two signals is quite similar). In interior locations, the relative levels of the direct signal, and the acoustic reverberation (commonly referred to as the *Direct-to-Reverberant ratio*, or DRR) also strongly contributes to the perception of auditory distance. In exterior locations, the high frequency content of very distant sounds will also be reduced due to air absorption.

7.1.3. Speakers, Spatial Audio and Installations

> One of the main things about my work is the physical aspect of sound. A lot of people think it's the narrative quality, but it's much more about how our bodies are affected by sound.
>
> (Cardiff, quoted in Tubridy 2007)

Broadly speaking, there are two distinct approaches to the use of audible space in installations. The first simply involves the allocation of a single loudspeaker for each sound source, positioned at every desired location in the performance space. In effect, each loudspeaker acts in isolation as a single, real source, which can be readily localized in terms of both direction and distance, regardless of the position or movement of the audience. Such an approach was adopted by Janet Cardiff's installation *Forty-Part Motet*, in which a recording of Thomas Tallis' 1575 choral work *Spem in Alium* is reproduced by 40 loudspeakers positioned in a circle, with one single singer reproduced by each individual loudspeaker (Tubridy 2007). Visitors to the installation therefore have the freedom to experience the entire spatial polyphony of the work from the center of the reproduction area or, alternatively, move closer to individual voices in the choir. While this approach is conceptually simple and offers some advantages in terms of robustness, it is also logistically demanding and may not be practical if large numbers of sounds, or the dynamic movement of sounds, is a requirement. In this instance, a second, more practical, solution may be the use of a spatialization technique, in which a smaller number of loudspeakers in a defined layout are used to virtually position sounds.

7.2. Spatialization Techniques Compared

A great many spatialization techniques have been developed over the past century; however, in general, the approach taken consists of one of the following:

1. The manipulation of level and/or time differences in pairs or multiple pairs of loudspeakers.
2. The reconstruction of a sound field over a listening area using a loudspeaker array.
3. The reconstruction of the ear signals using headphones.

The first approach, of manipulating either phase/time or, more usually, level differences between pairs of loudspeakers, is often referred to as *stereophony* of which two-channel stereophony (or just stereo) is by far the most common form. However, this exact same approach may also be extended to greater number of loudspeakers, as used in cinematic *surround sound*.

Ambisonics and *wave field synthesis* are two more esoteric techniques that attempt to reconstruct a sound field within a listening area using loudspeaker arrays. While these techniques are not quite as common as stereophony, their fundamentally different approach can be advantageous in certain contexts. For example, Ambisonics lends itself well to 360 video, and *virtual/augmented reality* (VR/AR) and is now widely used for the recording and production of spatial audio for such applications.

The third approach attempts to record or synthesize the two ear signals and directly reproduce the spatial auditory cues we perceive in normal hearing. This *binaural* approach (meaning having or relating to two ears) is highly applicable for a single listener, as it requires a strict separation of the two ear signals, such as when listening with headphones. While this technique also has a long history, much like Ambisonics, it was rarely used until the emergence of VR/AR in the twenty-first century, for which the delivery of spatial audio over headphones is highly common.

All of these different approaches generally feature both a recording technique, involving some type or arrangement of microphones, and a panning/encoding technique based in software or hardware. We will start however by looking at stereophony and see how these recording techniques and panners can be extended beyond simple two-channel stereo.

7.2.1. Stereophony

The word *stereophonic* derives from the Greek words *stereos*, meaning "firm, solid" and *phōnē*, meaning "sound, tone, voice." Two-channel stereo (or just stereo), which was first developed in the 1930s, has now become the standard audio format for most applications and is generally based on two loudspeakers positioned at a 60^0 angle in front of the listener. We can record stereo audio using two microphones that are routed to the left and right loudspeakers, or alternatively a monophonic audio file can be positioned between the loudspeakers using a panner. In the latter case, the position of the audio in the stereo field is controlled by simply adjusting the relative amplitude of the signal in each loudspeaker. This same principle can be extended to greater numbers of loudspeakers for sound spatialization and is the typical approach implemented in the spatial panners found in most digital audio workstations (DAWs), typically for common cinema surround sound formats such as 5.1 as shown in Figure 7.2.

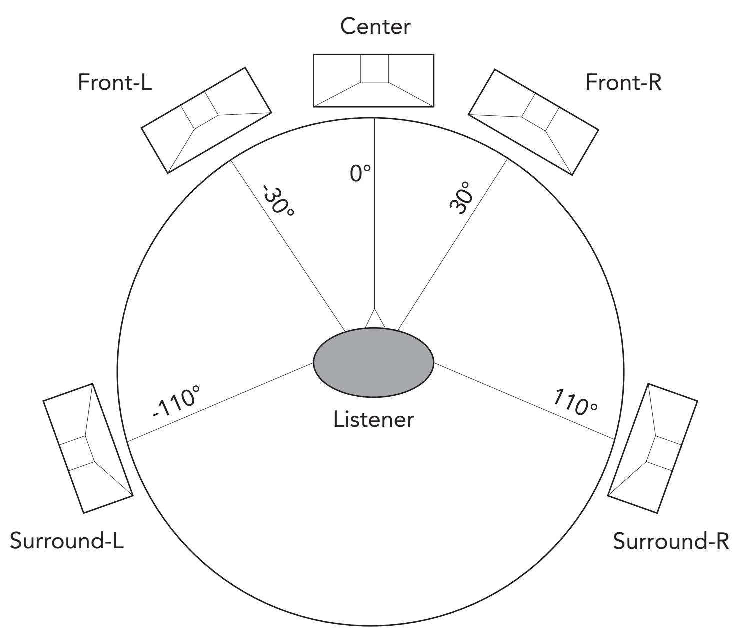


Figure 7.2 The 5.1 Loudspeaker Arrangement.

Source: Wikimedia commons, licensed under the Creative Commons Attribution-Share Alike 3.0 Unported licence

Likewise, different arrangements of multiple microphones can be used to record spatial audio for playback over such loudspeaker arrays (Theile 2000).

One of the challenges of multichannel stereophony is that these different microphone techniques and spatial panners are often closely tied to a specific, matching arrangement of loudspeakers (which is not the case with alternative techniques such as Ambisonics). So, while microphone arrangements such as the *Decca Tree* or the *Optimized Cardioid Triangl*e (OCT) are well matched to the left, center and right loudspeakers found in cinema surround sound formats such as 5.1, they are less suited to other arrangements of loudspeakers such as a regular octagon or loudspeaker dome. In the case of spatial panners, simpler implementations are similarly tied to a specific loudspeaker arrangement while more sophisticated tools may be able to adapt to different arrangements as required. The simulation of distance is sometimes also included in these panners, although this is often implemented separately using reverberation, filtering and amplitude changes (Chowning 1971). It is important to note that while many stereophonic panners may visually suggest that sounds can be positioned at any point inside the array, this is often simply achieved by routing the audio signal to greater numbers of loudspeakers. In the case of a 5.1 arrangement, a sound at the center of the array would be simulated by simply routing audio to all 5 loudspeakers. However, this effect can be rather fragile and may not be effective if the listener is seated at an off-center position, or in the case of very widely spaced loudspeakers. For this reason, a number of new cinema surround formats such as Auro 3D and Dolby Atmos emerged in the 2000s which included elevated and overhead loudspeakers that can more reliably position sounds at different points inside the listening area.

In a stereophonic system, if a signal is panned to the position of a loudspeaker, then only that loudspeaker will be used, but a signal panned to any other position will be created as a *phantom image* by a pair of loudspeakers. This results in a slight yet perceptible change in timbre as a signal is panned from a loudspeaker position to a point in between a pair of loudspeakers. As a result, when amplitude panning is used to dynamically move a sound around the array, the number of contributing loudspeakers is constantly changing, producing a small yet clearly perceptible timbral shift that tends to emphasize the loudspeaker positions and distort the perceived trajectory (Pulkki 1999). For this reason, many panners include a control to spread the signal and ensure that multiple loudspeakers are always used regardless of the panned angle, thereby eliminating this panning artifact for moving sounds.

Just like two-channel stereo, multichannel stereophonic recording techniques can be based on coincident, near-coincident or spaced microphone arrays. As with stereo techniques, spaced microphone arrangements will produce an increased sense of "spaciousness" due to the capture of timing differences in the recording and the resulting decorrelation of the loudspeaker signals. However, this may also reduce localization accuracy. Coincident techniques will often produce more accurate localization, but with a less spacious and "open" sound. The choice of approach is as much an aesthetic issue as a technical concern, but in general, spaced microphone techniques will produce a more robust, spacious reproduction, particularly for listeners positioned away from the center of the loudspeaker array (the so-called *sweet spot*). In contrast, coincident techniques can potentially produce more accurate directionality, particularly for listeners in the sweet spot, but may be less optimal at off-centre positions. In either case, there is a close connection between the type of microphone array used and the specific layout of loudspeakers required to reproduce the recording. One novel, and highly practical, approach to spaced microphone recordings using portable, low-cost sound recorders has been developed by artist Augustine Leudar (Leudar 2014) and is potentially useful for site-specific installations.

7.2.2. Ambisonics

Ambisonics is an alternative spatialization technique that was first developed in the 1970s by Michael Gerzon, among others (Gerzon 1974). Although originally based on analog hardware, Ambisonics is now a complete system for the recording, panning and reproduction of spatial audio using specific microphones and software panners which, unlike stereophony, can readily be adapted to a variety of different loudspeaker arrangements (although regular symmetrical arrangements are preferred). Unlike stereophony, the different audio channels in an Ambisonic signal are not simply loudspeaker feeds, but instead comprise a three-dimensional description of the entire sound field that must be decoded for a particular loudspeaker configuration (or headphones). Most commonly this comprises just four audio channels labeled W, X, Y and Z, but higher-order formats are available that provide increased localization accuracy but require a greater number of audio channels.

Ambisonics can be recorded using a specific type of microphone containing multiple, coincident microphone capsules in a single housing. Such recordings can be smoothly rotated and processed in different ways, which is highly beneficial for VR applications, and usefully can also be decoded for different loudspeaker configurations. Existing sounds can also be positioned using an encoder that differs from a stereophonic panner in that its output does not comprise individual loudspeaker feeds but is instead an Ambisonics signal that must be decoded for playback (see Figure 7.3). Although this approach is somewhat more complex

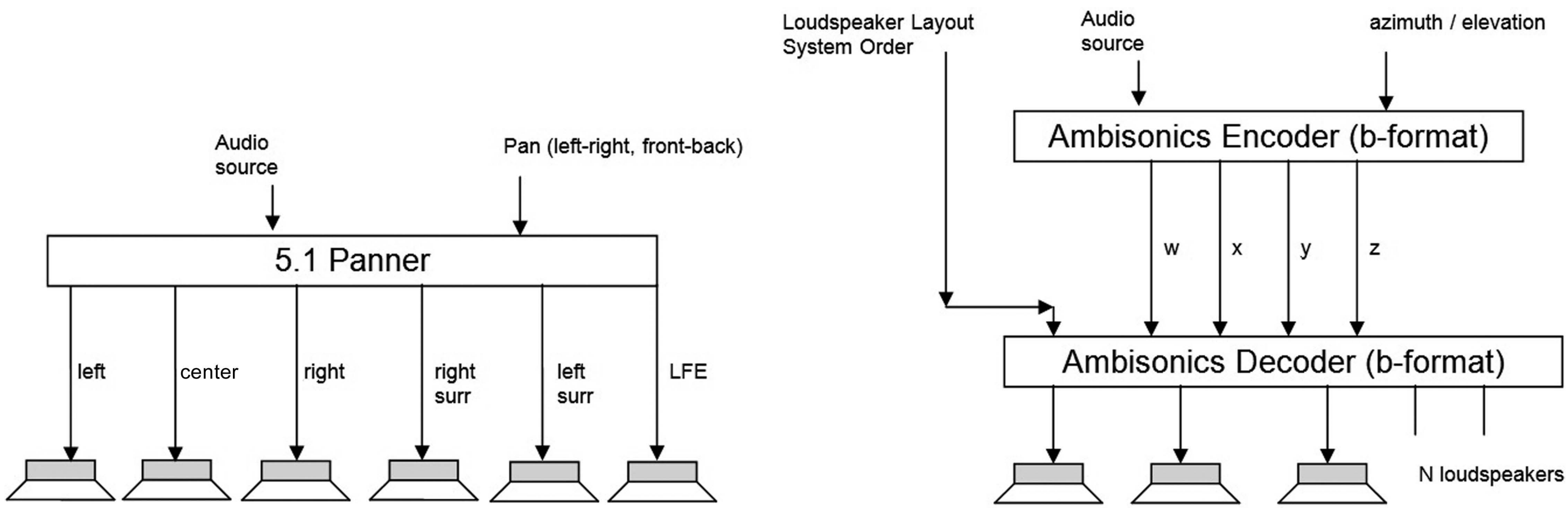


Figure 7.3 Stereophonic Panners Directly Output Loudspeaker Feeds while an Ambisonics Encoder's Output Must Be Decoded Before Playback.

than stereophony, the separation of encoding and decoding into separate stages allows for more flexibility in terms of the final reproduction system, whether that be a loudspeaker array or headphones. Ambisonics has been used in numerous art installations, with the work of composer and artist Natasha Barrett a prominent example (Barrett 2018). Much like coincident stereophonic recording techniques, Ambisonics performs optimally over a relatively small listening area, albeit with more flexibility in terms of the loudspeaker layout.

7.2.3. Binaural

Like stereo, *binaural recording* techniques were first developed in the 1930s and are generally intended for playback over headphones. Binaural techniques aim to recreate both binaural localization cues (ITD, ILD) and monaural cues due to the spectral filtering of the head pinna and torso, all of which are characterized by the *head-related transfer function* (HRTF). When these recordings are played back over headphones the effect is quite different from that of normal stereo. Instead of the highly unnatural (although by now quite familiar) effect of perceiving sounds as if inside the head, sounds are now externalized in much the same way as in normal hearing.

Binaural recordings can be made using either a dummy-head microphone or, alternatively, using a binaural headset in which two microphone capsules are inserted into the ears using an earbud-type arrangement. In the latter case, the recorded audio will be encoded with that particular individual's HRTF. As a result, the recording will closely mimic normal listening for that individual, but may be far less convincing for others. In the case of a dummy-head recording, a generic HRTF is used that is intended to work reasonably well for most listeners, but again the effectiveness of the spatial cues can vary considerably between individuals. In particular, reproducing sounds that appear to come from directly in front or behind or at different elevations can be difficult and can often be perceived as coming from inside the head. This occurs because the perception of these directions is largely determined by the filtering effect of HRTFs, which, much like fingerprints, are highly individualistic.

These recording techniques can also be used to capture a *binaural impulse response* (BIR), which can then be applied to an audio signal to artificially position that sound in the captured direction (much like convolution reverbs). Multiple BIRs can be used to dynamically position or move sounds in space, however, this comes with significant costs in terms of processing power and requires the capture of large numbers of BIR's for many different directions. An alternative approach is often used to render stereophonic or Ambisonic spatial audio over headphones by convolving each signal with a BIR for the angle corresponding to the position of the loudspeaker (generally referred to as virtual loudspeakers (McKeag and McGrath 1996)). Usefully, this only requires a limited number of BIRs (one for each loudspeaker feed) and is therefore a more efficient approach. The dynamic movement of sounds can then be created using standard panning tools to adjust the loudspeaker feeds prior to their convolution with BIRs.

This virtual loudspeaker approach is often used in combination with Ambisonics for sound spatialization in VR applications. When listening to stereo or binaural recordings on headphones, as we rotate our head, the entire sound field will also move, which is particularly problematic for VR. To avoid this, head-tracking can be used so that the perceived position of sounds does not change as the listener moves their head. This can potentially alleviate some of the problems with front-back reversals as the listener can rotate their head slightly to help resolve ambiguous source positions. To implement these systems, a head-tracker is attached to the headphones (or built into the VR headset) transmitting head rotation angles back to the reproduction system. This data is then used to smoothly rotate the entire sound field in the opposite direction (which is relatively straightforward with Ambisonics) so that sounds hold their position regardless of the orientation of the listener.

Binaural audio has been widely used in art installations, particularly when headphone reproduction is required. The highly personal nature of this recording and reproduction technique (if recordings are made using two small microphones placed in the ears) is highly suitable for the production of mobile works, such as some of the audio works of Janet Cardiff, or the binaural sound art of Dallas Simpson.

7.3. Acoustic Environments, Installation Practices and Spatial Aesthetics

Sound installations can generally be considered as spatial territories and containers for a variety of sonic materials. These defined spaces may partly be perceived within the context of their visual aspect (from galleries to more specific spaces and sites) in which dimensions the borders of the space may be quite defined, halting at walls, doors or windows. However, the sonic borders of such territories and spaces may be more porous or even nonexistent. Indeed, while many authorities have sought to enumerate the differences between sonic space and visual reference points (Sterne 2012, p. 9; Blesser and Salter 2006, p. 21), the use of sound itself as a medium creates the potential to shrink and expand the perceived audiovisual boundaries of a given space: “sound necessarily exceeds itself, washing over borders” (Labelle 2003).

Rooms and buildings do provide some aspects of hard boundaries for sound, and the distinction between sound and vision within such spaces is not always so stark when considered on the basis of a more scientific model of such phenomena. The key distinction may be seen as related to the speed of the sound waves; echoes of sonic events are often heard due to reflections from various (hard) surfaces producing later arrivals of "copies" of the original vibrational impression, a temporal distinctiveness that is not a factor in our everyday visual experience. This can give rise to sonic boundaries that are analogous to visual borders. While a doorway into a room may be a particularly lax border-guard for external sonic events, the sonic environment within is truly a different land; the faint echoes of reflected sound heard within signal both its interiority and the particular dimensions (and materiality) of its construction. Installation artists who utilize sound will need to bear in mind the perceptual principles of what Blesser and Salter (2006, p. 1) term *aural architecture*: how the intersection between sound waves, objects and surfaces may give rise to a sense of a location's dimensions and structure.

7.3.1. Aural Architecture and Acoustic Territories

A generic room is a rectangular box, and this is the kind of spatial archetype that many performance and gallery spaces tend towards. However, though elegant in its visual simplicity, this type of space poses significant problems for the installation designer. Such rooms will resonate in various different ways based on their dimensions, the types of surface materials and the types of sound materials. These factors may contribute *frequency-based effects*, that is, the subtle "coloration" acoustic structures impose on sound materials (Blesser and Salter 2006, p. 42), and *temporal effects* (discrete echoes and diffuse reverberation).

The geometry of smaller rooms will affect the audibility of frequency content. Certain frequencies—those with wavelengths that are close to multiples of the room dimensions—will be amplified through reflections off the various surfaces. These resonance effects are known as *room modes*: the room will have a more pronounced response to certain frequencies (in effect, amplifying them) and will have a less active response in relation to others. Such responses are heavily dependent on location and may be particularly obvious in relation to certain lower frequencies; if certain bass materials sound quite "boomy," it may make sense to move a speaker or other sounding device to another part of the room. (Higher frequency materials are likely to pose less of a problem in terms of this differential response: modes within an octave increase significantly with frequency, making them hard to differentiate (Rumsey and McCormick 2006, p. 23)). Such active room acoustic responses can be creatively exploited, as in Alvin Lucier's *I am Sitting in a Room* (1969) which re-recorded the products of a room's audible response iteratively in order to process a recording of speech. In this case, the significant room responses are to be found in higher frequency region, including *comb filtering*: the effect of (very) short delays (produced by reflected sound), combining and constructively and destructively interfering with the original sound source (or other delayed materials). Small, especially narrow, spaces may provide distinctly colored responses that may be creatively exploited in terms of their ability to reinforce or subvert an audience's sense of aural perspective (Brant 1967).

In larger rooms, the temporal attributes of acoustic responses predominate. A large concert hall, church or even school gymnasium or swimming pool will have *reverberant* acoustic responses; multiple/repeated echoes from the hard surfaces, dispersed over time, will become blended together into a diffuse "afterimage" of the original sound event, which is known as *reverb*. The decay time of this reverberant tail can be related to the dimensions of the room and various figures (coefficients) for the materials of the room's surface. Reverb time is standardized as RT_{60}, or the amount of time it takes for the reverberant response to decay by 60 dB. Installations that occur within such large spaces will need to make use of strategies to relate to their sound materials to the space. Any rhythmic or transient details (including speech sounds) may become indistinct if the reverberant response predominates (beyond a certain distance from the source, termed the *critical distance*). In addition, there may be some discrete echoes that are audible in a given space and location; although it might be expected that these would interfere with localization, the *precedence effect*—sometimes termed the *Haas effect* (Haas 1951)—holds that localization direction is judged on the basis of the "first arriving wavefront" rather than the direction of any reflections (within approximately 50 msecs). As such, we are able to localize with a reasonable degree of accuracy within typical rooms.

Installation designers therefore need to bear in mind the role of rooms as active sonic environments. Thus, it may be helpful to consider a number of exemplar cases:

1. Anechoic or outdoor spaces: spaces that have no significant reflected sound created through the presence of reflecting walls. Such spaces may be created through the deployment of absorbent and diffusing surfaces (an anechoic chamber) or may be found in certain outdoor environments.
2. Small, colored spaces: smaller rooms in which modal characteristics or comb filtering effects (through reflections from hard walls) may impose changes upon the frequency content of sound materials.

3. Reverberant spaces: as noted above, larger spaces have less pronounced modal characteristics and more pronounced temporal responses. What is an appropriate reverb time will depend, to some extent, on the type of sound materials to be presented. Content with significant transient detail may benefit from a shorter reverb time; significant harmonic or slowly developing textural detail may benefit from a longer reverb time (many orchestral halls have reverb times of around two seconds; see (Beranek 2003)).

Typically, gallery spaces will tend towards the second case (small, colored spaces), but larger-scale or multi-room/multi-space installations may entail encounters with more reverberant spaces.

7.3.2. Soundscapes and Spatial Perspectives

Apart from the temporal and frequency characteristics, the presence and type of background sound may affect the suitability of a space for sound installations. Ambient sound materials may impose themselves upon a sound installation, either through their audibility and the "salience" of this audibility (through distractingly recognizable or repetitive sources) or, if of significant amplitude, through *masking*, in which relatively high amplitude sources may perceptually "cover" sources within adjacent frequency ranges.

More generally, such background, ambient sounds may help to articulate a spatial structure or opposition against which a sound installation may function. Emmerson (2007, pp. 97–101) proposes a spatial-territorial typology of musical sound events based on their inhabiting different *space frames* that are typified by two primary functional associations: *local* (a foreground or region of clearly perceived events) and *field* (a more diffuse ambient background context for local/foregrounded events). These frames may deal with the interrelationships between various composed sound materials and the wider context of the soundscape, and between ideas of spatial "presence" versus "absence" or "dispersion"; see Figure 7.4.

Emmerson suggests that listeners track sonic-performative space within concentric circles arising from gestures within a focused area, framing localized events within the frame of a *stage*, which is encapsulated within both an *arena* (performance/installation environment) and a wider background *landscape*. An important consideration for the sound designer here is that of audience position. Electroacoustic concerts, like private listening, typically employ a *fixed*, forward-facing audience orientation (Smalley 2007, p. 52; my emphasis) within the delineated "stage" of a

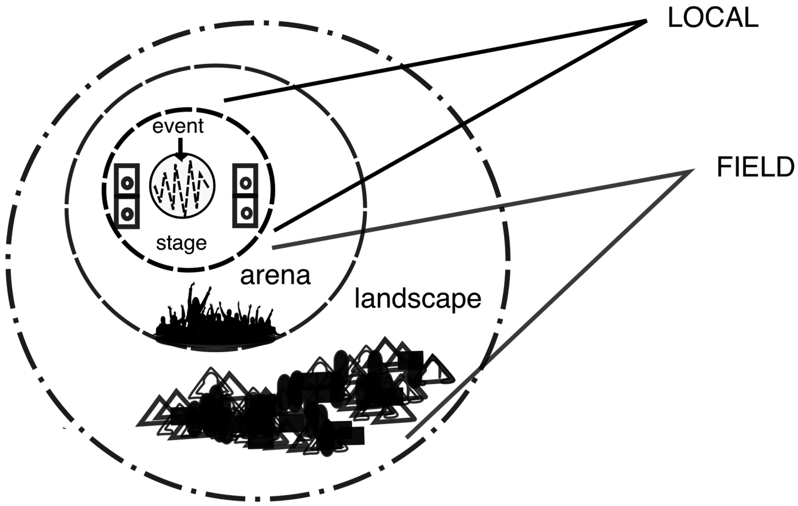


Figure 7.4 Local and Field Space Frames.

Source: Adapted from (Emmerson 2007)

stereo or circumspatial, multichannel speaker array; in turn, the spatial manipulation of sonic events serves to blur perceived boundaries between fields. Works may be mixed with an intended audio sweet spot (usually central) in mind, with the potential drawback that the optimal conditions for experiencing the work are confined to a relatively small zone. In contrast, *variable* listening positions may offer an alternative audio experience of a piece that, while arguably less ideal in a circumambient context, can nonetheless be considered equally favorable (Smalley 2007, p. 52). Arenas such as Stockholm's Audiorama embrace the contrasting experiences and, consequently, potential multiple readings of a sonic work that result from different listening perspectives by placing random seating positions facing in various directions within an extensive, circumambient speaker array.

The common practice of *sonic diffusion* furthers the potential for fluidifying the boundaries of sonic space frames. Here, a designated individual (often the composer) manually controls the amplitude and spatial distribution of an electroacoustic piece (Harrison 1998) to various speakers around the performance space; in its most basic form, this may involve mixing a stereo composition to a multichannel array. Sound is adjusted "live" in response to the particular speaker system as well as to the acoustic and spatial properties of the arena; the interface (mixer) and the room become "instruments" (Gibbs 2007, p. 132). (Some speakers may be positioned specifically to produce the impression of greater distance via the off-axis response of being turned away from the audience, producing the attenuation of high frequencies (Harrison 1998).) Diffusion practices are not uncontroversial; they may introduce *phase cancellations* through the sonic interactions of multiple copies of audio with different amounts of delay, based on the different positions of multiple speakers (Piché, cited in Harrison 1998). As Harrison responds (ibid.), the acoustics of a performance space are, in any case, likely to cause phase cancelations for audience members further away from any fixed "stage" position.

In the installation context, multiple "stages" may be provided via speakers or sounding mechanisms, across several acoustic arenas, with local or field assignment, depending on the relative distance or nature of sound materials. Works may allow and indeed encourage further movement from the listener(s) in order to facilitate their full experience. Multiple, conflicting, complementary, cumulative or interpenetrating audio zones may be articulated, within which the listener can discover their own favored listening position. Indeed, the speakers themselves may constitute their own spatial "corners," creating quasi-architectural half-box havens of immobility (Bachelard 1964, p. 137) that entice the listener. Bruce Nauman's *Raw Materials* (2005) depends on this very dynamic; amid the complex of samples emanating from 22 speaker channels placed within the vast turbine hall of London's Tate Modern, audience members move towards individual speakers to discern the piece's separate audio threads. Ultrasonic technology and the use of *audio spotlights*, extends this concept further, enabling the sound designer to isolate entirely egocentric zones of listening, whose material is otherwise excluded from the rest of a given audio space resulting in a more aggressive, almost *anti*-sweet spot. Elsewhere, multiple and mobile listening positions may be used as a means of prompting a greater sense of integration between audio content and its architectural vessel as well as interaction within mixed-media contexts, resulting in what Smalley terms *peripatetic* listening (2007, p. 52). Like the broader definition of installation art proposed by de Oliveira et al., here, audio scrutiny is quelled "in favour of a consideration of the relationships between a number of elements or of the interaction between things and their contexts" (1994, p. 8).

7.3.3. Immersion Versus Localization in Spatial Audio Aesthetics: Space, Place and Scale

In many ways, both Emmerson's space-frame model and Smalley's three vantage points of listening depend on an overall dynamic of *immersion versus localization*. Although it is a commonplace, almost trite term today, the notion of immersion does, at least, resonate with the ideas implied in this chapter's epigraph, namely, the acknowledgement of sound and sound installation as a physical, dimensional entity in which we are *immersed*, "a spacious thing that you can inhabit" (Brewster 1998). "Walking through [sound] in its resonant state provides an experience similar to perusing a landscape but from the inside with all of your body instead of from the outside with just your eyes" (ibid.). It is sonic immersion that arguably provides the conditions—the backdrop—for experiencing more localized convergences of sound or isolated sonic gestures as found in the Cardiff and Naumann examples mentioned earlier. This relationship between immersive sound *en masse* and its component strata—"a universe dotted with little stars of sound, moving in compact nebulae or isolated" (Xenakis 1992, p. 237)—forms the basis for the employment of space as a compositional parameter in the musical works of composers such as Xenakis and Stockhausen. For both composers, the spatial dispersal of instrumentalists around a concert space allows the musical material of a given piece to become "purer" (Xenakis in Varga 1996, p. 97), more "transparent" (Stockhausen in Cott 1974, p. 187) than if emanated from a single, fixed (ensemble) source, thanks to the increased dynamic of spatial mobility between gestures that the greater distance between players (and reduced effects of frequency-based masking) facilitates. Immersion is thus the somewhat disorderly state that enables the composer to "tame" space (Varga 1996, p. 97). (Presumably silence may also be considered immersive, since it functions in a similar manner.) Grimshaw interrogates the concept of immersion further, citing Slater's distinction between immersion and "presence." Here, "presence is the feeling of being within a space in which there is potential to act. It is subjective whereas immersion is objective and relates to the potential of the . . . environment to facilitate presence" (Grimshaw 2015, p. 280). A fundamental aspect of what both writers distinguish as presence is the sense of a "coherent 'place' that you are 'in'" (Slater 2003, p. 2; cited in Grimshaw 2015, p. 287). It may therefore be more beneficial to consider sonic installations as architectural constructs or environmental sites, rather than draw the more obvious comparisons that tend to be made with musical composition. Thierry du Duve's notion of *site*, as discussed by Mavash (2007, p. 55) and Mittosian (1997, p. 64), offers a useful point of reference in this respect. De Duve theorizes site in the context of architectural design as a harmonization of three fundamental concepts: (the cultural associations of) *place, space* (as a perceived grid of references) and (human/embodied measurements of) *scale*, noting the propensity for many examples of contemporary art, sculpture and architecture to spotlight two of these parameters while compromising the third. Unlike "fixed," visual entities, installed sound might be considered as encompassing a shifting dialogue between all three. As Labelle (2003) notes, when used as a creative medium, "sound is both subject *and* object." Sonic material, its spatial diffusion and guise as a spatial form in itself can be used to evoke *place*, map *space* and both articulate

and manipulate *scale* to achieve what is, for all intents and purposes, a variant of Du Duve's site, albeit in flux.

7.4. Conclusion

Sound installation practices entail a number of creative affordances through the presented audio's relationship with the acoustics of the space, its speaker system and the audience's perspective. The manipulation of the audio via various spatialization/diffusion techniques activates, reinforces and subverts the relationship with the presentation space. Whether through the more *ad hoc* approaches of point source, multiple audio-channel-per-speaker approach (as in Cardiff's work), diffusion of identical materials through multiple loudspeakers or more technically considered approaches such as Ambisonics or binaural techniques, contemporary sound artists and sound designers have a wide variety of approaches at their disposal through which to activate various types of auditory-spatial experience. Aesthetic dynamics of local and field frames, and diffusion as activating aspects of a presentational space or moderating a fixed auditory perspective, may challenge the audience's relationship with a given space. Installation art's concern for the redefinition of space finds in sound a means to reconfigure apparent dimensions and, indeed, relationships with neighboring spaces and soundscapes. The intimacy and presence of the localized sound event, as against the immersion entailed by diffusion, provides listener-participants in any sound installation with spatialized experiences ranging from the proximal (or even internal) to the expansive and enveloping. That such experiences can sometimes operate in tension with the audience's preconceptions of a given space via fairly simple technological means is one of the reasons why sound is a powerful modality in the hands of an installation artist. Sound's ripples activate and challenge our perceptions of, and egocentric perspective within, space and place.

References

Bachelard, G., 1964. *The Poetics of Space*. Translated from the French by Jolas, M. Boston: Beacon Press.

Barrett, N., 2018. Installations. Viewed June 28, 2018 www.natashabarrett.org/installations.html

Beranek, L., 2003. *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. New York: Springer.

Blesser, B. and Salter, L., 2006. *Spaces Speak: Are You Listening?* Cambridge, MA: The MIT Press.

Brant, H., 1967. Space as an Essential Aspect of Musical Composition. In: Schwarz, E. and Childs, B. (Eds.), *Contemporary Composers on Contemporary Music*. New York: Holt, Rinehart and Winston, pp. 221–242.

Brewster, M., 1998. Where, There or Here? An Essay About Sound as Sculpture. Viewed June 28, 2017 http://michaelbrewsterart.com/news/where-there-or-here-sound-as-sculpture-michael-brewster

Brophy, P., 2014. Melbourne Conversations: What's the Future of Sound and Music? (panel discussion—Fox. R. chair). Knowledge Melbourne. Viewed Aug 5, 2017 www.youtube.com/watch?v=w7eO5Aj2IGA

Chowning, J., 1971. The Simulation of Moving Sound Sources. *Journal of the Audio Engineering Society*, 19, pp. 2–6.

Cott, J., 1974. *Stockhausen: Conversations with the Composer*. London: Picador.

De Oliveira, N., Oxley, N. and Petry, M., 1994. *Installation Art*. London: Thames & Hudson.

Emmerson, S., 2007. *Living Electronic Music*. Aldershot: Ashgate.

Gerzon, M.A., 1974. Sound Reproduction Systems, US Patent No. 1494751.

Gibbs, T., 2007. *The Fundamentals of Sonic Art & Sound Design*. Lausanne, Switzerland: AVA.

Grimshaw, M., 2015. Presence Through Sound. In: Wöllner, C. (Ed.), *Body, Sound and Space in Music and Beyond: Multimodal Explorations*. Abingdon, Oxon and New York: Routledge, pp. 279–297.

Haas, H., 1951. On the Influence of a Single Echo on the Intelligibility of Speech. *Acustica*, 1, pp. 49–58.

Harrison, J., 1998. Sound, Space, Sculpture: Some Thoughts on the 'What', 'How'and 'Why'of Sound Diffusion. *Organised Sound*, 3, no. 2, pp. 117–127.

Hartmann, W., 2013 *Principles of Musical Acoustics*. New York: Springer.

LaBelle, B., 2003/2007. Short Circuit: Sound Art and the Museum. *Journal BOL*, no. 6, Insa Art Space, Korea, pp. 155–175.

LaBelle, B., 2006. *Background Noise: Perspectives on Sound Art*. New York and London: Continuum.

Leudar, A., 2014. An Alternative Approach to 3D Audio Recording and Reproduction. *Divergence Press*, Dec, 3, no. 1.

Mavash, K., 2007. Site + Sound: Space. In: Muecke, M.W. and Zach, M.S. (Eds.), *Resonance: Essays on the Intersection of Music and Architecture*. Ames, Indiana: Culcidae Architectural Press, pp. 53–75.

McKeag, A. and McGrath, D., 1996. Sound Field Format to Binaural Decoder with Head Tracking. *Preprint of the Audio Engineering Society for the 6th Australian Regional Convention*, no. 4302.

Mittosian, C., 1997. Public Art/Public Space. In: Ramesar, A. (Ed.), *Urban Regeneration: A Challenge for Public Art*. Barcelona: Universitat de Barcelona, pp. 61–70.

Pulkki, V., 1999. Uniform Spreading of Amplitude Panned Virtual Sources. In: *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. Mohonk Mountain House.

Rumsey, F. and McCormick, T., 2006. *Sound and Recording*. Oxford: Focal Press.

Smalley, D., 2007. Space—Form and the Acousmatic Image. *Organised Sound*, 12, no. 1, pp. 35–58.

Sterne, J., 2012. Sonic Imaginations. In: Sterne, J. (Ed.), *The Sound Studies Reader*. London: Routledge, pp. 1–18.

Theile, G., 2000. Multichannel Natural Recording Based on Psychoacoustic Principles. *Preprint of the Audio Engineering Society for the 108th Convention*, no. 5156.Tubridy, D., 2007. Sounding Spaces Aurality in Samuel Beckett, Janet Cardiff and Bruce Nauman. *Performance Research*, 12, no. 1, pp. 5–11. doi:10.1080/13528160701397984

Varga, B.A., 1996. *Conversations with Iannis Xenakis*. London: Faber & Faber.

Xenakis, I., 1992. *Formalized Music: Thought and Mathematics in Music*, Revised edition. Hillsdale, NY: Pendragon Press.