Sound diffusion and the Sonic Image

Robert J. Dow

The University of Edinburgh, School of Arts, Culture and Environment
Department of Music, 12 Nicolson Square, Edinburgh
Telephone: 0131 650 9352 E-Mail: R.Dow@music.ed.ac.uk
Web: http://www.music.ed.ac.uk/staff/directory/rdow/

Dr Robert Dow is a Research Fellow at the University of Edinburgh

Abstract

In recent years, the use of multi-loudspeaker systems in the performance of electroacoustic music has experienced a small but palpable renaissance. This has manifested itself not only in an renewed interest in stereo diffusion, but also in a re-emergence of multi-track work, the latter inspired by a wish, among other things, for more refined spatial control and whose realisation is made very accessible through advances in software and the decreasing cost of hardware.

The acousmatic composer working in stereo is (normally) very much concerned with the creation of various coherent sonic images between the two playback monitors. Here, the speakers are not simply treated as individual monophonic 'voices' rather a sense of space is synthesised between the speakers, creating malleable artificial soundscapes. The problem for the performer splitting this stereo soundfield into a large multi-loudspeaker array, is that although sound events are often easily rearticulated in such a scheme, the original sense of sonic image can often be broken. This tends to be even more problematic in the performance of multi-channel works, where there may not be any sense of sonic image in the first place, the work far more concerned with monophonic sounding circular movement of the sort often difficult to achieve in standard stereo diffusion.

In this paper, the concept and nature of the sonic image is revisited, and reassessed with respect to multi-channel work.

Many discussions concerning recorded sound often appear to rely on analogies in the visual realm. This is, perhaps, hardly surprising given the perceptual primacy that sight has in our (mostly sighted) society. As Michel Chion points out, sight is:

> '[...] the most highly structured sense. It takes on a remarkable variety of forms and disposes of a highly elaborated language, which dwarfs the vocabularies for phenomena of touch, smell, and even hearing. Sight is generally what we rely on for our orientation, because the naming and recognition of forms is vastly more subtle and precise in visual terms than with any other channel of perception.'[1]

For the normal-hearing person, the audio domain is up to two orders of magnitude poorer than the visual domain in terms of rendering spatial detail. Although we are able to hear in all directions, our reaction to auditory events of interest is to 'point' our head in their direction in order to allow for a more accurate spatial analysis by our visual system.[2]

It is worth noting at this stage, however, that our auditory system is, nonetheless, extremely sophisticated. At the lower limit of sound intensity, this system allows us to detect sounds just above the thermal limit of molecular Brownian motion. At the upper limit, it allows us to deal with sounds one thousand billion times higher than this: an amazing dynamic range.[3] The frequency range of human hearing is similarly impressive, normally spanning around ten octaves.[4] Furthermore, the longer wavelength of sound, compared to that of light, allows sound waves to bend round corners in a useful fashion, allowing us to be warned of even invisible dangers. Finally, our auditory system's ability at spectral analysis allows us to detect weak spectral signatures even in the presence of broadband noise.

When we consider the 'sonic image', particularly in the context of what could be referred to as 'representational' sound recordings, there would seem to be little *prima facie* affinity between the concepts of the sonic and the visual image. Indeed for some, what is heard when recordings of sound events are transmitted is indistinguishable from

---

[1] Chion M, *The Voice in Cinema* New York: CUP (1999)
[2] Grantham D W, 'Spatial hearing and Related Phenomena' in Moore B C J, *Hearing* San Diego:Academic Press (1995), 297
[3] Schroeder, M R, 'Listening with Two Ears' *Music Perception* 10:3 (1993), 255
[4] Watkinson J *The Art of Sound Reproduction* Oxford: Focal Press (1998), 75

the original pro-phonographic event, which leads to a clear distinction between the concepts of sound copy (or image) and sound reproduction (or recreation). As Béla Balázs has stated:

> 'What we hear [...] is not an image of the sound, but the sound itself which the sound camera has recorded and reproduced again ... there is no difference in dimension and reality between the original sound and the recorded and reproduced sound'[5]

This view expresses the recording/reproduction process as forming an essentially linear, dyadic relationship between input and output. The sounding objects propagates acoustical waves through some physical medium (gas, liquid or air) which are 'captured' and transformed into a recording. On replay, the information held on the recording is transformed back into acoustical waves identical to the original ones. Thus, sound is reproduced not only legibly but also literally, recorded from the perspective of what James Lastra has described as the 'invisible auditor',[6] the ideal recording process being one which creates a transparency between input and output. Even the common notion of the 'fidelity' of a sound recording/reproduction system is habitually perceived as being an assessment of this transparency; a measure of how 'faithful' the reproduction is to the original.

The more recent advances in sound technology would appear to align themselves with this reproductive theory of sound recording, particularly with the advent of digital sound and the nascent evolution of more refined periphonic spatial encoding techniques such as ambisonics where even our normal, unframed, omnidirectional experience of sound can be recreated. However, there are many problems with this view of recording. The perceived inherent automatism in at least the idealised process of recording would

[5] Balázs B, *Theory of the Film: Characterr and Growth of a New Art* New York: Dover, 1970
[6] Lastra J, *Sound Technology and the American Cinema* New York: CUP (2000), 143

seem to guarantee a causal association between pro-phonographic object and its phonographic representation, creating what C.S. Peirce has described as an indexical relationship between the two. However, even in the simplest recording situation, for example a single omni-directional microphone attached directly to recording apparatus, or a more spatially detailed one, using binaural, stereo or periphonic methods this relationship is broken, because of the technical, environmental and contextual issues which arise.

To begin with, it is impossible to reproduce the original sound of some particular event through recording (or probably any other means), or indeed even one perspective of it. Technically, sound recording and playback systems are not perfect, and will tend to alter a variety of parameters in a non-linear fashion. Indeed, these characteristics are often selected in order that recordings conform to current normative practices, such as the selection of valve technology to create a 'warmness' in recorded vocals, or the use of bass-heavy loudspeakers for playback to emphasise the beat in a dance venue. Furthermore, as Alan Williams has noted, during a sonic event, it is the 'entire volume of air that vibrates'[7] within the production space and it is this which creates our perception of sound. Therefore, sound is 'spatio-temporally *specific*, or in a broad sense, *historical*',[8] and thus the replay of recorded sound in a different space creates a different sound event. Indeed, the auditor's perception of the sound is dependent on his or her position in the reproduction space. Rather than the 'reality' of the original sound event, what is heard is its image, the original sampled and contextualised through the act of recording and recontenxtualised and almost framed through the act of reproduction.

---

[7] Williams A 'Is Sound Recording Like a Language?' *Yale French Studies* 60 (1980), 53

[8] Lastra J 'Reading, Writing, and Representing Sound' in Altman R (ed) *Sound Theory Sound Practice* New York: Routledge (1992), 67

However, sound recording and playback are not only mediated by their technologies and the environmental/contextual parameters of recording and realisation. The very deed of sound recording requires the subjective influence of human mediators whose incentives are 'the goals of the very act of representing'.[9] Roland Barthes, in his discussion of the photographic image, refers to this mediation process as 'connotation', the connoted image being 'an object that has been worked on, chosen, composed, constructed, [and] treated to aesthetic or ideological norms.'[10] Sound recording requires the use of artifice to create the impression of such things as naturalness and authenticity and to allow recordings to be intelligible within the framework of their representational intent.

For example, in recordings of instrumental concerti it is common to lift the sound of the soloist out of the general plane of the orchestral mix, which among other things is to improve legibility, to recreate our experience of the concert hall (where, for example, visual cues would allow us to 'focus in' on the soloist), and indeed to conform to expectations of current recording practice. Another example is in the recording of sound for cinema, where it is often the case that narrative speech takes precedence over both other sound events and the maintenance of sound realism, being recorded in 'close-up' even despite changes in the image scale, again to ensure intelligibility. According to Carl Dreher:

> 'Since the reproduction of sound is an artificial process, it is necessary to use artificial devices in order to obtain the most desirable effects. For example, it is normal procedure to reproduce dialog at a level higher than the original performance. This may entail a compromise between intelligibility and strict fidelity [...]'[11]

---

[9] Ibid. 75
[10] Barthes R *Image Music Text* London: Fontana Press (1977), 19
[11] Dreher C 'Recording, Re-recording, and Editing of Sound' *JSMPE* 16:6 (1931) 756

What we actually hear has been constructed, and what has been created is something divorced from our experience of listening to the 'live' sonic event. For instance, it is axiomatic to suggest that listening to a recording of Beethoven's 5th is not the same experience as hearing it in the concert hall. Although in sound recording '[t]he thing we're seeking is a method of packaging a human experience in such a fashion that we can unpackage it anywhere anytime and enjoy it',[12] what is finally presented to us is only analogous to this experience: an image of the pro-phonographic event from which the real experience is inferred, yet absent.

The absence of an original will be of no surprise to those who compose with recorded sound. In acousmatic music, it is only, as it were, the copy—what an auditor hears at the moment of performance—that is important. This will be the case in even those works, such as environmental soundscapes or works with a high proportion of recognisable, 'everyday' sounds, which may appear to seem documentary in nature on the surface. This is not to say that a possible 'original' (or number of 'originals') will not be imagined, their nature lying somewhere between the quite specific (the sound of the St Giles organ in Edinburgh) or the generic (a metallic sound). John Young has pointed out that:

> '[i]n normal listening experience, we are highly motivated toward correlating a sound with a source object or action, since it forms a part of the way we deduce and interpret the physical nature of our surroundings, assisting orientation and survival in our immediate environment.'[13]

Based on our familiarity with environmental and cultural indexical relationships between sound source and sound cause, our need for correlation between them is so strong, that it is hard to imagine any possibility of a completely abstract sound existing. Rather, we are

---

[12] Campbell J W, 'Hearing is Believing', in Hoopes Jr. R H (ed), *The High Fidelity Reader* New York: Hanover House (1955), 25

[13] Young J, 'Imagining the Source' in Norman k (ed) *Contemporary Music Review* 15:1-2 (1996), 75

more likely to have a sense of increased ambiguity the more the sound image appears to represent something whose concrete origin cannot be fully ascertained. Even within such an equivocal environment, however, certain sonic traits may still be apprehended, for example the perceived forces between sonic objects, and their actions and reactions. To cite John Young again:

> '[...] when the source-cause origin of sound is ambiguous we may still detect the interaction of motivating energies and the apparent responses of sources—such as the impact of metallic objects, or friction between surfaces—yet in a composite fashion that prevents us from making a definitive recognition of a specific identity.'[14]

So far, very little has been discussed regarding the sound image and its relationship to space. The sound image may be considered as comprising divers sound objects articulated within a perceived spatial setting. In the case of sound recording, the recorded sound carries with it the stamp of its recording circumstances: what Rick Altman refers to as its *spatial signature* which is:

> '[...] carried in the audible signs of each hearing's particularities. Even when those signs are contradictory or have been tampered with, [...] they still carry information that is narrative and spatial in nature'.[15]

This spatial information, as has been discussed previously for the sound image in general, is subject to the mediation of such things as technological non-linearity, postproduction and the circumstances surrounding subsequent replay. For instance, a sound event closely recorded in a very reverberantly dry location, will sound distant and wet is replayed to an audience situated at one end of a large cathedral from its opposite end. In the rock and pop industry, it is common practice to record players in a more or less spatially neutral setting, and then synthesis spatial information at the postproduction stage. In acousmatic

---

[14] Ibid. 79

[15] Altman R 'Material Heterogeneity of Recorded Sound' in Altman R (ed) *Sound Theory Sound Practice* New York: Routledge (1992), 24

music, the spatial context, to use Trevor Wishart's distinction, can be either 'real' or 'unreal',[16] that is to say, perceived to have more concrete or more ambiguous origins. This spatial environment need not remain static: it is quite possible for it to change organically throughout a work, and indeed change its nature between the seemingly real and the seemingly unreal. Therefore, there are really two types of movement within the sound image: movement of sound space and movement of sound object.

The disposition of sound objects within their changing spatial context becomes particularly significant in the performance of acousmatic music. It is important to remember that the act of sound diffusion is the act of realisation of the acousmatic work: the final compositional act. There is no absolute version of the work, no original to reproduce, only the multiple versions as perceived by individual members of an audience in a particular location, in a particular space who are listening to sound produced by a particular configuration of replay equipment. The aim of diffusion is to recompose the acousmatic work for a particular spatio-temporal situation.

For works that have been encoded in stereo, it is usual to create an illusion of space in the stereo soundfield rather than using the speakers as monophonic 'voices'. This stereo illusion is both able to produce a sense of spatial width and spatial depth, creating a 2-dimensional placeholder for the realisation of sound spaces and sound objects. Using a multi-stereo diffusion system, where stereo pairs of loudspeakers are places in different perspective planes to each other, the sonic image, comprising spatially contextualised sound objects, may be moved between planes, articulating the actual 3-dimensional space. The trick of stereo diffusion is to maintain a coherent sonic image while achieving its articulation in real space. Part of this trick to achieve the perceptual

[16] Wishart T 'Sound Symbols and Landscapes' in Emmerson S (ed) *The Language of Electroacoustic Music* London: MacMillan (1986), 48

separation of sound objects and their corresponding sound contexts, to allow their evolution in space to occur independently. Since there is no real sound data separation on the recording, this really is artifice. To achieve such counterpoint successfully, the composer has to be fully aware of the limitations of such diffusion techniques in order to 'compose in' such possibilities in his or her work. For example, the use of spectral and/or dynamic separation of sounds can help the realisation of spatial separation during diffusion.

Stereo diffusion is not without its many pitfalls. Insensitive diffusion of a coherent sonic image can easily make it become incoherent and unintelligible. What we get in a typical stereo diffusion system, is the production of multiple copies of the sonic image, distributed over the real reproduction space. These copies do not generally glue together to create some sort of 'hyper-image'—a larger, 3-dimensional but equally coherent version of the stereo image—but have a tendency to fight each other, creating a perceptual muddle. Contrapuntal separation of various sound objects and contexts is not always possible, depending on the material, particularly where the material is in the same spectral band or in the same dynamic range.

Some approaches to multi-channel diffusion hope to address a number of these concerns. By having individual control of the content which is fed to individual loudspeakers or groups of loudspeakers which are placed around the audience, there can be in theory more control over the spatial juxtaposition of the discrete sound events making up the sound image. Furthermore, although one approach would be to use the loudspeakers as point monophonic sources, it is equally possible to deploy them in pairs, allowing the potential for multiple yet discrete stereo images to surround the audience.

In practice, however, multi-channel sound diffusion is often less convincing that it would appear it should be: it often seems difficult to maintain a coherent sense of sonic image. This must be in part due to the difficulty in controlling the increased number of input channels with respect to fewer output channels: that is to say the relationship of the multiple channels tens to be fairly fixed. The more channels, the more this will tend to be the case. Moreover, in stereo works, the sound sources are often based on stereo recordings, giving a good illusion of space within the sonic image. In many multiple channel works, the sound is treated as multiple mono or multiple stereo sources. As with in bad stereo diffusion, we have the problem of gelling these multiple images together in a real space.

To conclude, the notion of the sound image is far more analogous to that of the cinematographic image than might have first been thought. During the replay of recorded sound, an original is not reproduced but represented, and thus the original is absent. In acousmatic music, which exists solely as recorded sound, the sound image is only fully realised during diffusion, which can almost be considered as the final stage of the compositional process. To be fully in control of the process of acousmatic composition, it is important to be aware of this final stage, and not assume that it is possible to reproduce an absolute version of the music: the illusive acousmatic original. As is usual in music, performance potential needs to inform the composition process.

In this multi-channel renaissance, it is tempting for the composer to believe that he or she has regained complete spatial control. However, during their ultimate realisation in performance, multi-channel works are in fact often difficult to control, and indeed the perception that multi-channel works are offering something definitive (an original), often gives composers the illusion that they need not even try to perform the work in any interactive manner: a ‘balance’ between the loudspeakers is set, and the

piece left to its own devices. This is not to say that multi-channel work does not offer many advantages, such as the true separation of spectro-morphologically similar material. Rather, just as in stereo works, performance still needs to be carefully considered during and after the progress of the composition, and ultimately this may necessitate new diffusion techniques to allow multiple channels to be manipulated successfully in real time.