Composing Images in Space: Schaeffer's Allure projected in Higher-Order Ambisonics

Natasha Barrett
Norwegian Academy of Music
natasha.barrett@nmh.no

ABSTRACT

In post-Schaefferian acousmatic music space has taken centre stage. Many composers would now regard space and spatialisation to be the unique quality that sets acousmatic music apart from other contemporary music genres. Yet as the popularity of ambisonics and other spatialisation techniques has increased, many of the tools available encourage a 'point in space' rather than 'image in space' way of working; an approach that inhibits the development of the sound object as a spatial element in the compositional process. This paper investigates the acousmatic object as a spatial musical phenomenon, drawing on Schaeffer's idea of 'allure' as a way to contextualise tangible images that occupy the same real space as the listener. The work focuses on applying higher-order ambisonics spatialisation in the quest for both compositional and technical solutions. The work addresses the integration of technology, musical theory and creative methods, and is underpinned by examples from my recent composition.

1. INTRODUCTION

This paper describes a method for composing images in space, developing Schaeffer's concept of 'allure' as a spatial phenomenon, and applying technical solutions in higher-order ambisonics (HOA). Allure is the unique character of the sound defined by the material and the way it behaves. It is about revealing the energetic agent's way of being and, very broadly, whether this agent is living or not [1]. The acousmatic object as a spatial image can therefore be characterised by the manner in which it moves and behaves. By projecting this information in HOA, I aim to create images that appear to occupy the same real space as the listener, or in other words address tangible sound. To explain these ideas, it is first necessary to present the connection between acousmatic abstraction and the real-world, which is an interesting paradox discussed in section 2. Section 3 considers important differences between space in reality and space in acousmatic composition. One obstacle to composing spatial images concerns ambisonics tools which encourage a 'point in space' approach, the results of which are here deemed too far from how we hear real-world images and in contradiction to the musical aims of this paper. Section 4.1 summarises the recent developments in HOA processing that are designed to spread the perceived width of a point source, or to create immersive environments. The results they offer have become popular amongst composers, yet do not, and indeed are not intended to, address details akin to real-world images. The final section describes an alternative that integrates HOA and allure as a compositional method.

2. BACKGROUND, TERMINOLOGY, AND A POINT OF DEPARTURE

'Acousmatic', 'tangible', 'allure' and higher-order ambisonics (HOA) are some of the terms used in this paper that warrant discussion, and an explanation as to how they connect to the spatial image.

Simply put, 'acousmatic' describes the process of listening and composing with sounds removed from their visual causation. This commonly used explanation is however insufficient for the needs of this paper in which acousmatic abstraction is positioned in the context of the real-world. To elaborate on the acousmatic discussion we can add the subtleties of Schaeffer's original explanation: in the process of listening and composing, "by deliberately ignoring any references to instrumental causes or pre-existing musical meanings, we seek to give ourselves over entirely to listening..." and, "The acousmatic situation first disconnects the audio-visual context, but above all it makes possible, but not compulsory, to explore the sound in itself… this type of interest does not follow automatically from simply being disconnected from the audio-visual complex but from a specific intention on the part of the listener." [2]. The choice of words - 'deliberately' and 'not compulsory' - denote that acousmatic listening is not an effortless experience. However, composers may make it easier for the listener by the way materials are framed. Here the spatial discourse can serve as both a frame and the subject, allowing listeners to explore connections to the real-world and characteristics with which they may be familiar, despite a high degree of abstraction in the sound. Furthermore, in doing so, we may empower listeners to dynamically drift through the different modes of listening [2] as the salient features of materials are spatially revealed.

In this paper the term 'tangible' is applied to acousmatic sound. We cannot touch sound, but we can experience the sensation of an invisible entity occupying our own space, maybe within physical reach, apprehended through tokens of spatial reality. The popular field of embodied cognition supports the experience of tangible acousmatic sound: sound connects to our motor skills, and we internalise a concrete image through our listening imagination [3]. The clearer the location, size, and dimension of the sound, the stronger is the sense of spatial tangibility [4].

HOA has been chosen as the spatialisation technology for the compositional approach described in this paper. HOA captures the 360-degree sound field as a set the spherical harmonics representing the sound pressure on a sphere [5]. Ambisonics is either synthesized or derived from spherical microphone-array recordings, and then decoded to loudspeakers. The technology is widely documented, and as tools become integrated into common software packages and virtual reality engines, it is amassing users. HOA was chosen instead of Wave Field Synthesis (WFS) due to both practical and technical matters, including considerations concerning the necessary numbers of loudspeakers, the projection of height information, the nature of the audience listening area, and the role of ambisonics blur, (which will become clear later). HOA offers a number of advantages over other panning methods such as VBAP. These include practical features, such as enabling composers to think about space independently from loudspeakers, and technical features, such as the spatial results appearing to be more ‘in space’ than hugging the loudspeakers, advantageous use of ambisonics blur, and more features which will become evident in the course of this paper. Relevant considerations, and the implications of specific HOA choices are discussed in section 4.1.

Clear source and causation identities serve to complicate, for better or worse, spatialisation; they tend to dictate where we think the sound *should* be based on how recognisable the sound is as they are often more immediate and salient than distance and movement cues. A conflict of cues can create interesting paradoxes [6], but also break the tangible spatial scene. If we avoid the clearest of anecdotes or extra-musical connections, we avoid this potential conflict of information. Yet how then can an abstract sound have a ‘character’, which by definition needs a real-world connection? Answers can be found in Schaeffer's ‘allure’. For Schaeffer, allure was broadly about the behaviour of a sound’s spectrum and typology, and lacking our spatialisation technologies, it would have been impossible for him to have investigated the spatial potential of allure.[1] Here we arrive at our point of departure: energy is movement in space.

3. IMAGES IN SPACE

3.1 A moment of reality

First let us consider the real-world: a source may propagate sound waves from a point (yet this is less likely), or propagate frequencies in different directions from surfaces, volumes and through cavities. For any object sounding outside anechoic conditions, the propagating sound waves will interact with the environment and with the listener (in distance and in orientation). In essence, most objects create complex directivity patterns. Noisternig et al. [7] investigated the role of directionality in the perception of the illusion of realism. In their work, they measured the directionality of some acoustic instruments in anechoic conditions, and then auralised the measured radiation patterns, calculated the interaction of the direct wave fronts within a simulated acoustic environment, and projected the information in HOA and WFS for perceptual investigation. Their work demonstrates a method as well as the importance of source directivity in image formation.

3.2 The reality of acousmatic composition

Now let us consider the acousmatic context, which for the purposes of our discussion can be divided in two: our perception of the music and our aesthetical intent. In terms of perception, how much of the real-world context are we in fact sensitive to? Our sensitivity to spatial difference can vary significantly depending on the angle and elevation of the sound in relation to our listening position [8]. Our understanding of distance is even less precise, determined by a complex connection between sound identity and acoustical cues. Without doubt, our sensitivity to spectral and temporal variations is far more refined. Rather than consider the precise values of these parameters, it is more interesting to draw attention to two observations: the first is that in the acousmatic context, when sufficient comparative references are absent, some subtleties of the real-world spatial scene are redundant to our perception, especially very small angular and distance changes. The second is that as our sensitivity to spectral and temporal information is more precise than that of spatial information, a correlation in the changing states of all aspects will serve to enhance our perception of spatial change.

Elaborating on Schaeffer’s typology and morphology of allures, which is a classification derived from the sounding ‘agent’ and its mode of ‘sustainment’ (and for practical purposes is here equivalent to variations in energy through time), Schaeffer places three landmarks: variations in the sound may be mechanical, living, or unpredictable and more akin to other sounds from nature. He proceeds to propose that we first perceive the general profile of the sound, then the nature of the allure, and then afterwards become aware of the presence of a microstructure termed ‘grain’[2]. Yet Schaeffer’s spatial discussion is ambiguous and somewhat different from modern understanding. For example, when discussing time and duration, Schaeffer states that, “sound objects, unlike visual objects, exist in duration, not space” [9], and again later insists on a separation of spatialisation and spatial music, where spatialisation is for achieving clarity, rather than to project or embody the sound’s character [10]. If we refer to Schaeffer’s own summary diagram of the theory of musical objects [11],

[1] In Schaeffer’s description of 'Symphonie pour un homme seul', space is highlighted as a performance, rather than compositional concern, "...what if the conductor... where in charge of the three-dimensionality, if his gestures sketched out in space the trajectory that the sounds would make in the hall?” [26]

[2] From here on, the term ‘grain’ will be used to describe the microstructure of the sound.

allure is placed in the context of all other parameters investigated in his thesis – parameters that also relate to how we now address sound images and their behaviour in space – yet in Schaeffer's summary the spatial aspect is missing.

Chion further proposes, in his reading of Schaeffer, that allure and grain together affect subtle details in the sustainment of sound, linking form and matter in time [12]. We can summarise that Schaeffer's types of allure, and the nesting of profile, allure and grain, can guide us in the development of spatial allure regardless of how abstract the material is in terms of anecdote.

4. APPLICATION IN COMPOSITION

Figure 1 illustrates some of the connections between compositional goals, the real-world, the acousmatic context, and technological opportunities. The top section simplifies the aesthetical investigation. The middle section considers spatialisation and the acousmatic context. The bottom section concerns practical application.

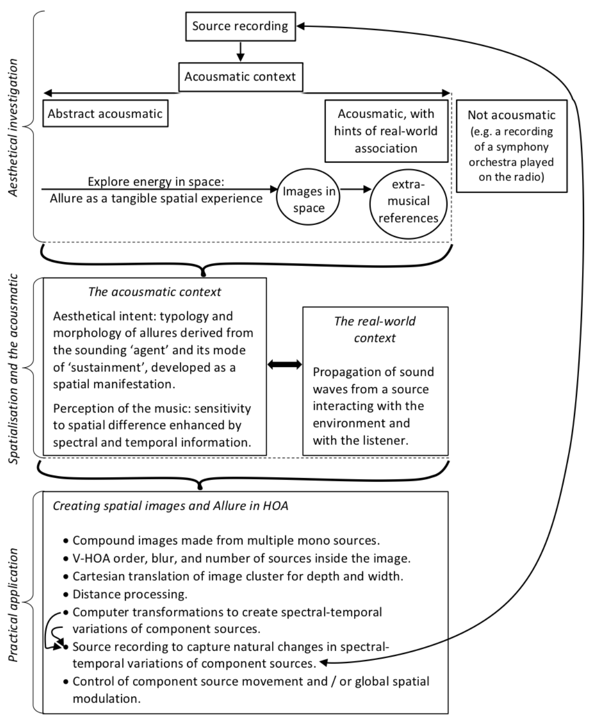

Figure 1. Connections between composition goals, the real-world, the acousmatic context, and technological opportunities.

4.1 Spatial images in general

In sound diffusion performance over a loudspeaker orchestra [13], images are normally created with mono or stereo sources. The image is created live as a real-time performance, where the performer adjusts the spatial results for each unique concert context, interacting with the mixed speakers and layouts of the acousmonium. The acousmonium, stimulating the room acoustics, adds a rich melt of delays, spectral colourations and dynamics to the result. Ambisonics technology instead synthesises the sound field by the user specifying the position of mono sources, whether static, fleeting, moving or spread in some way.

The work presented in this paper targets the use of very high order ambisonics (V-HOA), such as 7th order in 3D or 12th order in 2D. The reasons for working with V-HOA rather than lower orders are two-fold: to achieve greater spatial accuracy (higher orders result in less angular blur), and to provide a large audience listening area suitable for public concerts. However, the advantages of less blur likewise enhance the point-in-space rather than image-in-space experience. Solutions for creating images in V-HOA fall into two areas: technical solutions, where we apply new methods (which are becoming embedded in commercial tools and, for better or worse, determining how we work), and artistic solutions, where composers informed by technologies may explore how to realise musical goals. Technical solutions for creating a sound image from a mono source, or in other words manipulating the apparent source extent, so far gravitate towards spreading or blurring. A selection of options that operate in the ambisonics domain include: phantom source widening [14], which disperses the direction of arrival of the sound with frequency; ambisonic spatial blur [15], that takes advantage of lower order blur, where at the extreme, a 0th order correlated signal is sent to all speakers (with the problem of reducing the size of the sweet spot, as would be the result of encoding in low-order at the outset); reprocessing 1st order recordings with a new spatial encoding [16]; controlling the apparent source size by combing a number of methods, including ambisonic order reduction, rotating de-correlated copies of the original stream, and focusing on minimising the inter-aural cross-correlation coefficient [17]; W-panning [18] and other weighting functions to change source extent, and ad-hoc breaking of encoding-decoding guidelines, such as mismatches between encoder and decoder or combining lower and higher-order sources in the same composition. Granular techniques have also been applied to source distribution and then encoded in the ambisonics domain [19].

Although these methods control of the apparent source extent in V-HOA, we can argue that they less adequately address the features of the spatial image laid out in section 3.1. In particular, for the methods referenced above, the spectral-temporal details are relatively constant as the source changes in width, and there is little consideration of the change in spectral-temporal information with distance. In the context of our musical discussion, changes that reveal the allure, as well as the associated changes in the grain of the sound are lacking. Instead we can reserve these source-spreading methods for broader cases such as envelopment, and turn our attention to the artistic path where images are constructed from multiple mono sources arranged in space. Each source reflects one aspect or angle of the final image, preferably interacting correctly with environmental cues. The behaviour of directivity patterns of acoustic instruments is well known, and as [7] has demonstrated, it is possible to auralise this information in HOA.

The procedure is however less practical in composition, as it requires special measurement for each source, and that non-instrumental sources create chaotic or less predictable radiation patterns that change with the energy and nature of the excitation (which is also why they are fascinating sources for composers). Furthermore, a method that can be used for composing fictional sounds is needed.

In designing the image a number of questions are to be raised: how many component sources, how are the sources to be correlation, and how to build a practical method with existing tools? These questions are addressed in the following sections.

4.2 Image distance

Ambisonics encoders provide composers with a means to easily control the angle and elevation of as many sources as the computer CPU allows, but we soon encounter a problem: the encoding does not in itself address distance. Some encoders allow the user to modify the relative gains of the spherical harmonics, which may alter the impression of distance due to spreading effects, (assuming a reference sound is present for comparison), and indeed W-panning is commonly used for this purpose. If the sound is moving at the correct speed to trick our perception into believing it is approaching or departing, we may perceive a change in proximity[3]. This method proves less successful for slow moving or stationary sounds as well as resulting in a reduction in the size of the acceptable audience listening area when the W-channel is at a high gain ratio. Another solution is to pre-process the source with the most salient distance cues prior to encoding. These cues include the attenuation of amplitude and higher frequencies (mimicking air absorption) as the sound moves further away, and a change in the levels of the direct and reverberant fields, (where reverberation may also be part of a post-ambisonics processing routine using convolution). This type of processing is embedded in some spatialisation software but can also be easily added.

Equally important is the change in image width and depth, how spectral-temporal details and motion internal to the sound may be revealed or reduced, and how the directionality of the image interacts with the environment. Figure 2 visually represents a compound sound object placed at two different distances from the listener. A simple Cartesian translation serves to correctly change the perceived image width and depth with distance. We can also see that when the compound image is further away, ambisonics angular blur creates a spatial-spectral overlap obscuring detail and depth, while when the image is closer, the separation between the mono sources increases, revealing greater spectral-temporal variation and depth of geometry. Sound example[4] 1 [20], prepared from the development materials for the composition ‘Dusk’s Gait’ [21], illustrates this procedure. In this example reverberation as a distance cue is reduced, so that changes in distance are

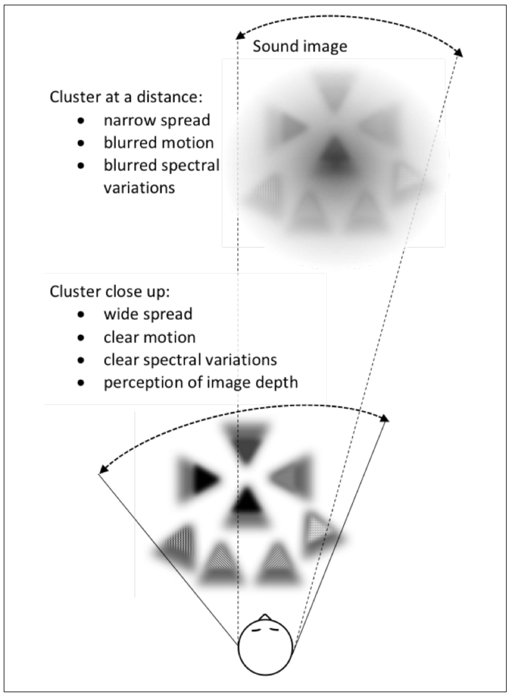

Figure 2. Compound sound object at different distances.

more clearly derived from changes in spectral variation and image dimension. The image is made from 12 mono sources arranged similarly to the image in figure 2. The image starts at a distance from the listener and angled to the right. At 0'03 the image moves towards the centre line, and from 0'03 to 0'16 approaches, at a varying speed, to settle close up directly in front. At 0'21 we move inside the image, and at 0'28 we emerge on the other side, with the image now behind and slightly towards the left. At 0'35 the image moves to the side left and stays in place to 0'45. From here we pass once more through the image, it then emerging on the right and retreating into the distant right. A more complex compositional example combining a number of images as well as environmental cues and can be heard at 4’11 in ‘Dusk’s Gait’.

With this method, the number of sources the image should consist of will be a factor combing ambisonics order (value of the angular blur) and how close we intend to allow the image to approach the listener. When close-up, with too few sources and the image will disassociate into mono points (which of course can itself be part of a compositional scheme, and in example 1, the use of 12 sources prevents this occurrence). It is also necessary to ‘handcraft’ each of the sources; ensuring some spectral-temporal coherence, but not so much as to lead to artefacts such as phase cancellation. This is the topic of the next section.

4.3 Allure

Having assessed the technical and practical considerations of image creation and control in spatial composition, we

[3] Near-field compensated HOA can in theory recreate the curved wave fronts of near-field sounds, but to date no perceptual studies in a compositional context, nor adequate real-time implementations, have been made.

[4] Sound examples are originally in 7th order SN3D ambisonics. Examples accompanying this paper are rendered in binaural from the 1st order stream using the Harpex decoder [25] and the KU100 HRTF set. The compromise in spatial precision and choice of HRTF set is for illustration only. Examples should be listen to over a high-quality pair of headphones.

can now return to the topic of allure as a spatial manifestation. Using the English translation of allure as gait[5], we can apply a spatial version of embodied cognition (discussed in section 2). When viewing a moving object from afar, we may embody or feel a distinctive movement with which we are familiar, without actually ourselves moving, and long before we experience closer details to clarify the subject's actual identity. To spatially explore the sounding image in this way, the component sources discussed in the previous section and the nature of their behaviour as energy in space, need *somethings, but not all things*, in common. Possible approaches to create this variation include spectral decorrelation prior to encoding, variations in texture (grain), or moreover a thorough consideration of the spectral-temporal aspects. The resulting combinations as compound images are somewhat limitless, making this pre-encoding approach a winner in terms of compositional method.

Another important aspect of the component sources concerns how they together create *motion behaviour internal to the image*. One way to compose this kind of motion behaviour would be to add micro-movements, where the character and amplitude of the movement of each individual source is controlled by their spectral-temporal dynamics (top of figure 3). A partial correlation in spectral change will result in a partial correlation in motion. The complete image is then moved globally in the listener-space. In example 2a, an image consisting of four relatively similar sources approaches and then recedes. The image appears to increase in size as it approaches, yet the result is more akin to a simple spreading with minor changes in spectral information. In example 2b we hear the same image and global motion, with the difference that each source is wiggled a little in space (the global motion is 15 meters and the wiggle is between 0 - 0.5 meters). The speed and distance moved of the wiggle is controlled by the spectral centroid of each source. The difference between example 2a and 2b, is that in example 2b, the image appears subtly clearer in space. We can however also hear that when the image is at its closest, it begins to disassociate and we hear hints of individual source movement. This result is expected when compared to the image in example 1 consisting of 12 sources.

Simpler solutions can be equally successful. We can for example develop an approach that originates in source recording. Referring to section 3.1, as acoustic sources propagate sound in all directions, recording the image with multiple microphones captures an array of partly correlated sounds to feed into our ambisonics image. Acousmatic composers have for decades recorded sound objects using multiple microphones as a way to capture more than a stereo phantom image. I first transferred this approach to the ambisonics domain in the work Trade Winds (2004) [22]. A similar approach has been used in live electronics since the late 70's, serving as automatic natural spatialisation or timbral immersion [23]. For the purposes of imaging in ambisonics, not only will spectral variation be captured in a more effortless way than by modelling or synthesis, if we record with close microphone techniques in

[5] Translators of Schaeffer's work point out that the translation to 'gait' is not entirely accurate.

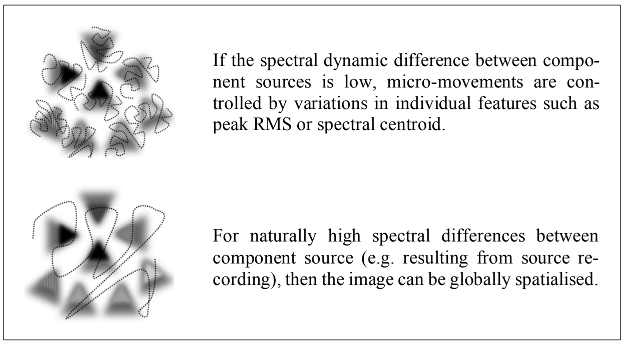

Figure 3. Motion behaviour inside the image.

the acousmatic tradition, then spectral and dynamic variations will be highly magnified, creating a sensation of movement inside the compound image. Although care must be taken to avoid a constant phase difference (which may result from microphones placed at difference distances and directions), naturally modulating phase variations are constructive in image movement. Microphone arrays can be placed either around or inside the sound object. In practical work, an A-format microphone can be useful, where the channels capture four directions (skipping the conversion to B-format). Next, instead of spatialisating each channel with a unique movement, we can track the running averages of the spectral-temporal group dynamics and use these values to influence movement of the complete image (bottom of figure 3), or simply control proximity. The source materials used in sound example 1 originate from this recording method. Sound example 3 plays four of these recorded channels spatialised across a 90-degree angle in front of the listener. No ambisonics spatial movement is applied.

These approaches connect sounds with musical structures, and in composition theory remind us of Emmerson's proposition concerning the relation of language to materials [24]: the materials offer an abstracted syntax (i.e. a method of organisation derived from the rules that construct the sounds) in all aspects including space. Abstracted syntax injects character to both the spatial image and its meaning within the composition. The discourse is aural (rather than mimetic or anecdotal). Abstracted syntax and aural discourse is where Emmerson positions Schaeffer's work.

5. CONCLUSIONS

Ambisonics as a spatialisation method for both contemporary electroacoustic music and commercial media is ubiquitous, and the intersection between composition and technology is an exciting area of investigation. Here, the acousmatic spatial-object can be reflected on with both technical and musical considerations.

In focusing on the musical, I have developed Schaeffer's idea of 'allure' into the spatial domain. Although Schaeffer formalised his ideas in the 1960's, which in some respects can be considered part of history rather than part of the present, we find interesting clues for our work. Sources that only remotely connect to real-world images need to find place and space in the experiential reality of tangible sound in composition, and in the concert or private space. By considering the perceptual and compositional implications of acousmatic images projected into HOA, we can then contextualise composition goals, technological opportunities, and how to create spatial images.

In focusing on the technical, we see that the increased use of ambisonics in VR and sound design has called for software solutions that automatically address some of the topics of this paper, and a few of the available options have been outlined. Such solutions are most useful when the eyes are leading how we hear. I hope to have illustrated that by taking compositional control over the available technology, rather than allowing the technology to dictate how we work, we are able to compose tangible 3D images for the ear alone. Abstract materials come alive, characterised by the projection of allure as a unique way of moving and behaving in space.

6. REFERENCES

[1] P. Schaeffer, Traité des objets musicaux (Treatise on Musical Objects, translation by Dack, J., and North, C)., Oakland, California: University of California Press, 1966 / 2017, pp. 443-446.

[2] Ibid., pp. 69-111.

[3] R.-I. Godøy, "Images of Sonic Objects," *Organised Sound,* vol. 15, no. 1, pp. 54-62, 2010.

[4] N. Barrett, "Creating tangible spatial-musical images from physical performance gestures.," in *Proceedings, NIME'15*, Louisiana, May 31-June 3, 2015.

[5] J. Daniel and S. Moreau, "Further Study of Sound Field Coding with Higher-Order Ambisonics," *116th Convention of the Audio Engineering Society, 8-11 May,* 2004.

[6] N. Barrett, "Spatio-musical composition strategies," *Organised Sound,* vol. 7, no. 3, pp. 313-323, 2002.

[7] M. Noisternig, F. Zotter and B. Katz, "Reconstructing sound source directivity in virtual acoustic environments," in *International Workshop on the Principles and Applications of Spatial Hearing*, World Scientific Publishing, 2011, pp. 357-372.

[8] J. Blauert, "Spatial Hearing," Cambridge, Massachusetts: MIT Press., 2001, pp. 39-50.

[9] P. Schaeffer, Traité des objets musicaux (Treatise on Musical Objects, translation by Dack, J., and North, C), Oakland, California: University of California Press, 1966 / 2017, p. 190.

[10] Ibid., p. 325.

[11] Ibid, pp. 464-467.

[12] M. Chion, Guide des objets sonores: Pierre Schaffer et la recherche musicale (Guide to Sound Objects, translation by Dack, J., and North, C.), http://www.ears.dmu.ac.uk, 1983 / 2009.

[13] H. Harrison, "Sound, space, sculpture: some thoughts on the 'what','how'and 'why'of sound diffusion," *Organised Sound,* vol. 3, no. 2, pp. 117-127, 1998.

[14] F. Zotter, M. Frank, M. Kronlachner and J.-W. Choi, "Efficient phantom source widening and diffuseness in ambisonics," *roc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 3-5 April,* 2014.

[15] T. Carpentier, "Ambisonic Spatial Blur," *Audio Engineering Society Convention 142,* 11 May 2017.

[16] E. Deleflie and G. Schiemer, "Spatial grains: Imbuing granular particles with spatial-domain information.," in *Proceedings of ACMC09, Improvise, The Australasian Computer Music Conference, 2-4 July.*, Queensland, 2009.

[17] T. Schmele and U. Sayın, "Controlling The Apparent Source Size In Ambisonics Using Decorrelation Filters," in *AES Conference on Spatial Reproduction*, Tokyo, July, 2018.

[18] D. Menzies, "W-Panning and O-Format, Tools for Object Spatialization," in *22nd International Conference of the Audio Engineering Society*, 2002.

[19] N. Barrett, "A Musical Journey towards Permanent High-Density Loudspeaker Arrays," *Computer Music Journal,* vol. 40, no. 4, p. 35–46, 2016.

[20] N. Barrett, "ICMC2019 Sound Examples in Binaural," 2018. [Online]. Available: http://users.notam02.no/~natashab/ICMC2019.zip.

[21] N. Barrett, Composer, *Dusk's Gait.* [Sound Recording]. www.natashabarrett.org/works/dusksgait.html. 2018.

[22] N. Barrett, Composer, *Trade Winds.* [Sound Recording]. Aurora . 2004 (2007).

[23] S. Emmerson, Living electronic music, Ashgate, 2007, pp. 128-130.

[24] S. Emmerson, "The Relation of Language to Materials," in *The Language of Electroacoustic Music*, S. Emmerson, Ed., London, Macmillan Press, 1986, pp. 18-40.

[25] S. Berge and N. Barrett, "A new method for B-format to binaural transcoding," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space.*, Tokyo, 2010.

[26] P. Schaeffer, In Search of a Concrete Music (translation by Dack, J., and North, C.), University of California Press, 1952 / 2012, p. 65.