SONIC ENVIRONMENTS

ACMC2016 | JULY 10-11 | BRISBANE

CONFERENCE PROCEEDINGS
sonicenvironmens.org
A CLASSIFICATION OF MULTI-POINT SPECTRAL SOUND SHAPES

Stuart James
Western Australian Academy of Performing Arts
Edith Cowan University

ABSTRACT

Previous research by the author has involved the investigation of sound shapes produced by the multi-point spatial diffusion of independent spectral bands. Fundamentally two implementations emerged through this research: one that primarily dealt with only the diffusion of spectra (i.e. spectral spatialisation) and another further extension of this approach that accounted for unique frequency-space distributions unfolding through time (i.e. timbre spatialisation implemented in the frequency domain). Through the process of exploring these possible sound shapes, a range of multi-point distributions emerged making it possible to form a categorical set of distinct multi-point distributions. The classifications were informed by the writings of Gary Kendall, Francis Rumsey, Robert Normandeau, Ewan Stefani, and Karen Lauke on spatiality, writings by Albert Bregman on auditory scene analysis (ASA), writings on directionality and immersion within the field of psychoacoustics, writings by Denis Smalley on spectromorphology, spatiomorphology, spatial texture, contiguous space, and non-contiguous space (i.e. zones), writings by Gary Kendall on spectral correlation and decorrelation, and writings by Trevor Wishart on spatial motion.

1. INTRODUCTION

The spatialisation of the spectral content of a sound enables the musical exploration of spatial percepts to a degree unobtainable with traditional point-source spatialisation techniques. While sound spatialisation has always been a fundamental part of the language of electroacoustic music, spectral spatialisation enables new timbral identities and morphologies to be explored (Kim-Boyle, 2008, p. 1). Whether imagined or experienced, both Ewan Stefani and Karen Lauke have described such spatial gestures:

Lauke and Stefani refer not only to timbre, but also to the notion of spatial de-composition and re-composition. Spectral analysis and resynthesis techniques may be seen as analogous to this process, traditionally allowing for the deconstruction and reconstruction of sound based on its frequency content (Lippe, 2003); a process that has been linked to the term spectromorphology (Smalley, 1994; 1997). This process of ‘decomposition’ is also one of the four necessary criteria for electronic music composition discussed by Karlheinz Stockhausen (1978). Just as sound can be de-constructed in various ways, its re-constitution can take not only the amplitude and phase of each frequency into consideration, but may also determine the spatial re-distribution of each frequency. Documentation of such a technique is not new. Robert Normandeau (2009) discusses a technique of decomposing a sound source into between 4 and 16 bands using bandpass filters, and spatialising these independently. David Topper et al (2002) also adopted a similar process when de-composing the sound of an input source into eight separate bands, and again spatialising these independently. Other research projects have used spectral analysis as a means of de-composing an audio signal into separate frequency bands, and then spatialising these independently (Barreiro, 2010; Keyes, 2004; Kim-Boyle, 2006, 2008; Lippe & Settel, 1993, 1999; Torchia & Lippe, 2003).

In parallel to these explorations in spectral and timbre diffusion, and in a closely related application, Gary Kendall (1994; 1995) documents a perceptual theory and application of signal decorrelation, a method that has been described for its ‘volumetric’ modelling of acoustic fields (Kaup et al., 1999). Time-varying spectral decorrelation techniques adopt time-varying all-pass filters that are used to filter the input signal. Kendall describes the effect of dynamic (or evolving) signal decorrelation as a:

...spatial effect akin to the sound of an environment with moving reflecting surfaces or moving sounds sources, such as the movement of leaves and branches in a forest ... and dynamic decorrelation imparts a quality of
liveliness to a [sound scene]. (Kendall, 1995, p. 77)

Due to the complex constructive and destructive interference of different frequencies across a sound scene, due to phase differences, the obtained output signals are perceptually equal but statistically orthogonal (Kendall, 1995). What is more, Kendall highlights the effect of decorrelated audio on the creation of diffuse sound scenes without the use of reverberation stating that the advantages of such a technique results in a reduction of image shift from different listening locations, the reduction in perceptivity of combing and colouration due to constructive and destructive interference, and reduction of the precedence effect\(^2\) (Haas, 1951; Lindemann, 1986; Kendall 1995). Spectral spatialisation, and timbre spatialisation in the frequency domain, can result in similar outcomes: both outcomes are immersive (James, 2015a), can contribute to a widening of the listener area or ‘sweet spot’ (Normandeau, 2009), are capable of creating diffuse sound scenes without reverberation, and can reduce the precedence effect.

The advantage of spectral spatialisation and timbre spatialisation over spectral decorrelation techniques is the ease from which the multi-point spatial diffusion can be controlled, as the spatial sound shapes can be highly directional, zone-like, or completely immersive, and may morph between these states with ease.\(^3\) Some of these spatial transformations, or spatiomorphologies (Smalley, 2007), may follow some of the scene-based transitions outlined by Trevor Wishart (1996) in his discussions on spatial motion. Some of these transitions discussed by Wishart include scene contraction and expansion, rotation, translation, swing, twist, spiral, and distortion, and these have each been applied to multi-point distributions of sound spectra by Stuart James (2015a). These transformations are linked closely with Smalley’s notions of spectromorphology and spatiomorphology (James, 2015b). However, in order to classify each sound shape, it is important to compare momentary instances of spectral distribution, and discuss these sound shapes within the theory and context of spatial music practice, those being notions of spatiality and the field of psychoacoustics.

2. TERMINOLOGY AND CONTEXT

Gary Kendall (2007) argues that the spatiality of electroacoustic music still lacks a definitive vocabulary. This is confirmed not only by psychoacousticians but also by conceptual artists:

Although there has been much value written about spatial attributes and the role of space, mainly by composers, the thinking is somewhat scattered, and as yet there is no substantial, unified text on the topic, nor any solid framework which might provide a reasonably secure basis for investigating space. (Smalley, 2007, p. 35)

Although interest in psychoacoustics emerged in the late 19th century, it has only been in recent years that perceptual research has been able to clarify a vocabulary for spatial attributes (Kendall, 2010; Yost, 2014). In 1990, Albert Bregman, a Canadian professor and researcher in experimental psychology, cognitive science and Gestalt psychology, designed and conceptually organised a field known as auditory scene analysis (ASA). Deriving attributes from the framework of ASA, Francis Rumsey (2002) uses the three ‘dimensional’ attributes: spatial width, distance and depth. Gary Kendall and Andrés Cabrera (2011) also discuss the attributes of direction and height. The spatial attributes that relate to our perception of immersiveness in reproduced sound—as outlined by Rumsey (2002) and Sazdov et al. (2007)—are envelopment, presence/spatial clarity and engulfment. Research by Robert Sazdov, Garth Paine and Kate Stevens (2007) has involved a perceptual investigation into envelopment, spatial clarity, and engulfment in reproduced multichannel audio. These attributes are listed in Table 1, where they are classified according to their dimensional or immersive quality, and they are further illustrated in Figure 1.

Table 1: Spatial Attributes

<table>
<thead>
<tr>
<th>Dimensional attributes</th>
<th>Immersive attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>direction</td>
<td>envelopment</td>
</tr>
<tr>
<td>distance</td>
<td>engulfment</td>
</tr>
<tr>
<td>width</td>
<td>presence/spatial clarity</td>
</tr>
<tr>
<td>depth</td>
<td></td>
</tr>
<tr>
<td>height</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. A visual illustration of auditory spatial attributes in relation to the listener. Note. Image source: Kendall, 2011.

---

1. Also referred to in literature within the field of psychoacoustics as soundfield, although a soundfield should specifically apply to Ambisonics.

2. Helmut Haas discovered that we can discern the sound source despite additional reflections at 10 dB louder than the original wave front, using the earliest arriving wave front. This principle is known as the Haas effect, the precedence effect or, in the psychoacoustic literature as the law of the first wavefront.

3. This is dependent on the control interface used, and the author has discussed this in-depth elsewhere. Refer to James, 2016A.
Interaural intensity differences (IID), interaural time differences (ITD) and spectral cues are responsible for determining the directionality of sound sources. Depth of field, and the perception of distance, is determined by an additional set of auditory cues. Spatial width is defined as the perceived spatial dimension or size of the sound source (Potard & Burnett, 2004). Scientific evaluations in the fields of acoustics and psychoacoustics often refer to apparent source width (ASW) (Keet, 1968; Morimoto & Maekawa, 1998; Sato & Ando, 1999). Listener envelopment (LEV) is an attribute that is used to define the sense of immersivity achieved in a reverberant environment, such as a concert hall, where sound seems to be arriving from all around the listener (Ando, 1998; Berg & Rumsey, 2001). Although LEV is largely treated as a 2D attribute, engulfment is an immersive spatial attribute associated with the sensation of being 'covered by sound', and is unique to 3D speaker configurations (Sazdov et al., 2007; Lynch & Sazdov, 2011). Spatial clarity has been also identified as an important spatial attribute (Nakayama et al., 1971), and is closely related to ‘directional selectivity’ as it is placed within the same subdivision and therefore can be interpreted as being linked to other established attributes such as sense of direction, localisation or source localisation (Gustavino & Katz, 2004).

In terms of the listening experience, timbre spatialisation challenges traditional notions of spatialisation, such as single point-source techniques, since the kinds of movements that result from such a process do not easily sit comfortably with existing written taxonomies of spatial motion such as those found in Wishart’s On Sonic Art (1996).

Although the processes by which this research explores spectral spatialisation and timbre spatialisation draw on panning algorithms that rely on the fundamental theories of sound localisation, these low-level processes, due to their polyphonic and multi-point nature in practice, give rise to a range of immersive spatial attributes including spatial width, spatial clarity and envelopment. The exploration of multi-point spatialisation of sound spectra allow for the exploration of all of these dimensional attributes and immersive attributes, that is the perceived spatial width, distance, direction, envelopment and spatial clarity of such distributions. Most importantly, in the process of spatialisation, such multi-point spatial distributions can shift between dimensional attributes and immersive attributes.

Albert Bregman’s ASA provides a framework that proposes several key inference processes that may inform the way in which we, as listeners, perceive the kinds of sound shapes that timbre spatialisation produces; that is, a number of independently spatialised frequency bands, or auditory streams. The theory of ASA suggests how a listener can perceive multiple streams of musical information through the processes of stream fusion and stream segregation. Any similarity in auditory cue, based on their vertical or horizontal organisation as they extend through time (i.e. pitch, timbre, location), may ultimately subserve streaming. In other words, streams will thus favor the grouping together of sounds that are perceptually similar, and segregate sounds which are perceptually dissimilar (Moore & Gockel, 2002). Auditory streaming theory describes audio objects as sequences displaying internal consistency or continuity, and ultimately serving the purpose of clustering related qualities (Kubovy & Valkenburg, 2001).

David Kim-Boyle (2008) stated that for most practical purposes when researching spectral spatialisation he found little need to explore more than 256 bands of frequency as the various individual trajectories cannot be discretely perceived but instead tend to coalesce into complex spatial gestalts. This issue of auditory fusion versus segregation largely has to do with the level of spatial correlation. Taking any broadband sound source and decomposing it into several, spatially distinct, point sources, if the correlation of movement of these distinct point sound sources is high, the human auditory system perceives them as a single auditory event, the location being perceived to be at the centre of gravity. The position of the centre of gravity depends on the positions and intensity gains of the point sources. However, if the point sound source movements are highly uncorrelated or weakly correlated, the human auditory system perceives the point sources as distinctly separated auditory streams, and results in the perception of a spatially wide sound source. However, if the point sources are densely distributed, it might not be possible to distinguish every single point source as a different stream because the human auditory system produces a

---

4 This has alternatively been referred to as interaural level difference (ILD) by some authors.
5 Introduced by the reflection of sound off the convolutions of the pinna, the shoulders and the upper torso (Kendall & Martens, 1984).
6 The perception of distance has been attributed to the loudness, the direct v. reflection ratio of a sound source, sound spectrum or frequency response due to the effects of air absorption, the initial time delay gap (ITDG) and movement (Howard & Angus, 2009). Further research in the multi-point distribution of spectra has accounted for distance cues by incorporating such auditory cues into the model. Refer to James (2016b).
7 Also referred to in psychoacoustic literature as spatial extent, source width or tonal volume. The spatial width of sound sources is a natural phenomenon; for example, a beach front, wind blowing in trees, a waterfall and so on.
8 It has been shown that the apparent source width (ASW) of a sound source depends on the value of the interaural cross correlation coefficient (IACC) (Morimoto, 2002), sound loudness (Boring, 1926), pitch, and signal duration (Perrott & Buell, 1982).
9 Also referred to in the literature as ‘clearness’. Letowski (1989) as part of his multilevel auditory assessment language classifies ‘clarity of sound texture’ as a subdivision of ‘distinctiveness’.
10 Catherine Gustavino and Brian Katz refer to spatial clarity in their colouration attribute. This makes use of the terms ‘muffled’ and ‘clear’ when rating spatial scenes, making it a timbral attribute (Sazdov et al., 2007).
final impression of a single, spatially large, sound source (Potart & Burnett, 2004). In the case of densely distributed streams, Topper et al. (2002) suggest that a listener experiences a ‘persistance of audition’ in that they are aware that auditory objects are moving, but are not completely aware of where and how. Bregman (1990) notes that conditions can be altered to make localisation easier or more difficult, so that “conflicting cues can vote on the grouping of acoustic components and that the assessed spatial location gets a vote with the other cues” (p. 305). This distinction is shown visually in Figure 2.

Figure 2. The variety of sound shapes created through timbre spatialisation has some relationship with decorrelation processes. The independent spatialisation of frequency bands generates multiple auditory streams. Beyond a threshold, the brain begins to fuse these separate streams into a single, yet spatially wide, sound scene. The numbers represent suggested loudspeaker distributions.

In addition to exploring the nature of spatial extent and immersion through the number of simultaneous auditory streams, a similar process of auditory assimilation was observed as a result of the global rate of change of the system. In this research the spatialisation is controlled by processes computed at audio rates, the potential rate of spectromorphology and spatiomorphology is considerably faster than if the system were using control rate signals, which are commonly used in spatial music practice. A pertinent example of audio-rate spatialisation is in Karlheinz Stockhausen’s Sirius (1975–77) where the sound moves so fast in various spatial trajectories that it seems to stand still, yet it has a motion that has been described as pulsation and vibration (Schmele & Gomez, 2012):

It is an entirely different kind of sound experience, because you are no longer aware of speakers, of sources of sound—the sound is everywhere. When you move your head even the slightest bit, it changes colour, because different distances occur between the two sound sources. (Schmele & Gomez, 2012, p. 26)

Spatial texture, as described by Smalley (1997, p. 124) is concerned primarily with how the spatial perspective is revealed over time. Such pulsating, vibrating or fluttering spatial textures are classified by the author as high-frequency spatial texture. Rates of change that occur at haptic rates tend to coalesce in a different way, as the rate of change allows the listener to discern the unfolding sequence of sound shapes in animation. These have been classified as low-frequency spatial texture. Results of both high frequency and low frequency spatial texture can be classed as being contiguous or non-contiguous depending on their spatial distribution, but are also ultimately concerned with the perceived fusion or segregation of multiple auditory streams.

The use of methods such as spectral spatialisation and timbre spatialisation allow for diverse range of spectromorphologies and spatiomorphologies to emerge. The following sections present the visualisation and categorisation of these sound shapes.

3. VISUALISATION

The visualisation of such processes requires a frequency–space representation, where colour conforms with frequency, and the spatial position is analogous to the auditory localization of each point. These are represented as colour plots over 2D horizontal speaker configurations. Figure 3a does indicate a speaker configuration, as numbered. The colour coding applied to these frequency–space plots follow the colour scale defined in Figure 3c.

Figure 3a. A frequency–space distribution over an eight-channel equidistant speaker configuration that only involves differences in azimuth. Lower frequencies (red) are displaced differently to the upper (blue) frequency bands in order to make the distribution of spectra as clear as possible. The numbers represent suggested loudspeaker distributions.

Figure 3b. A frequency–space distribution over any arbitrary speaker configuration involves azimuth and distance over a 2D speaker configuration. Whilst this is not so easily visualised on the page, this can also include elevation for sound shapes across 3D speaker configurations.

Figure 3c. A colour scale used to define the audible frequency range, and relate specific frequency ranges to spectral FFT bins.
4. SOUND SHAPES

Sound shapes and attributes that emerged through this research tended to fall into specific groups outlined by the framework established herein.

4.1. Single Point Source (Traditional Panning)

Single point-source spatialisation is possible with spectral spatialisation where all spatial coordinates mapped to all spectral bands are folded into a single point in space as shown in Figure 4a. This single point may be spatially transformed by translation across the Cartesian coordinate space (i.e. traditional point-source spatialisation), but other transformations like rotations of this point about its centre of gravity have no effect.

Whilst spectral spatialisation will result in all spectral bands folded into a single point in space, timbre spatialisation on the other hand may also fold this spectrum into a single frequency or band of frequencies. These frequencies would most conventionally be contiguous. Such sound shapes are considered to be contiguous and highly localised.

4.2. Spectral Zones and Multiple Point Sources

Multiple point-source spatialisation is possible using the technique of spectral spatialisation where spatial coordinates are segregated into a small finite number of spatial locations and mapped accordingly to all spectral bins. This process results in certain regions of the spectrum being located in specific zones of the sound scene as shown in Figure 4b. This would be considered to be a non-contiguous treatment if the separate zones are distinct enough to be perceived as separate auditory streams. These multiple groups of spectra may be spatialised individually to create effects like Normandeau has achieved by spatialising distinct bands of separated frequency across the sound scene. Spatial transformation may also be applied to all of these zones of spectra as a group, those transformations being contraction and expansion, rotation, translation, swinging, twisting, spiralling, and distortion across the Cartesian coordinate space.

In terms of timbre spatialisation, this process may also be applied to narrower bands of frequency such that the full spectrum of the audio input source is not reproduced. Such sound shapes are considered to be generally non-contiguous, highly localised, but can also incorporate attributes of spatial depth if different zones of sound are reproduced with varying perceived distances.

4.3. Spectral Diffusion (No Folds)

Spectral diffusion not only includes multiple groups of spectra placed in zones across the sound scene, but the exploration of spatial width encourages the possibility of spectra being ‘spread’ across space in a line (or potentially some other more abstract contour) as shown in Figure 4c. The simplest of this ‘spreading’ is a single linear distribution of frequency bands from low to high frequency across the sound scene. Spatial transformations may also be applied to this entire spectral distribution, those being contraction and expansion, rotation, translation, swinging, twisting, spiralling, and distortion across the Cartesian coordinate space.

Such sound shapes are considered to be contiguous, have a degree of spatial width (and are hence more enveloping), and depending on the distance of specific frequency bands, depth of field may be achieved where some frequencies are virtually placed closer to the listener than others.

4.4. Spectral Diffusion (One Fold)

Another possible sound shape is determined through the ‘folding’ of the frequency spectrum such that a number of frequencies are overlayed across space as shown in Figure 4d. This produces a further distinct shape that again may consist of a linear or non-linear distribution. Spatial transformation may also be applied to this entire spectral distribution, those being contraction and expansion, rotation, translation, swinging, twisting, spiralling, and distortion across the Cartesian coordinate space.

Such sound shapes are considered to be contiguous, have a degree of spatial width (and are hence more enveloping), and depending on the distance of specific frequency bands, depth of field may be achieved where some frequencies are virtually placed closer to the listener than others. The spectral folding increases the sense of envelopment, as perceptively each point in space is characterised by two distinct bands of frequency.

4.5. Spectral Diffusion (Multiple Folds)

When extending the ‘folding’ of the frequency spectrum such that a large number of frequencies are overlayed across a sound scene as shown in Figure 4e. Again such sound shapes are considered to be contiguous, have a degree of spatial width (and are hence more enveloping), and depending on the distance of specific frequency bands, depth of field may be achieved where some frequencies are virtually placed closer to the listener than others. The spectral folding here generates a further sense of envelopment.

Again spatial transformation may also be applied to all this spectral distribution, those being contraction and expansion, rotation, translation, swinging, twisting, spiralling, and distortion across the Cartesian coordinate space.

4.6. Low-Frequency Spatial Texture

Depending on the rate at which spectra are transformed, this will have a bearing on how spectromorphologies and spatiomorphologies are
perceived. Slower rates of change (i.e. haptic rate) result in slowly unfolding spectromorphologies across the sound scene as shown in Figure 4f. The continuous or discontinuous nature of these transformations will determine whether these spatial gestures are perceived as connected or disconnected. In these cases the spectromorphologies produced can either present smooth or jagged contours, characterised by the author as low-frequency spatial and spectral texture. These smooth or jagged contours may be timbrally and spatially distinct enough to be perceived as being fused or segregated auditory streams.

4.7. High-Frequency Spatial Texture

Depending on the rate at which spectra are transformed, this will have a bearing on how spectromorphologies and spatiomorphologies are perceived. Faster rates of change (i.e. audio rate changes of spatial attributes) result in pulsating and fluttering sound scenes that present a high degree of immersion as shown in Figure 4g. Due to the vast number of frequency bands, and the speed at which the location of these bands are determined, such sound shapes are perceived as being fused spatial images, but often feature some spatial movement reminiscent of those discussed in literature pertaining to spectral decorrelation. Noisy distributions (the use of white noise to determine spatial distribution) has some similarities to spectral decorrelation in that the results sound perceptively equal but are statistically orthogonal.

4.8. Other

Other attributes that appeared to be relevant included spatial width, spatial height or elevated spatial width or height ASW\(^1\), spatial depth, immersion, spectral quality, and spectral width. The spatial width of the sound shape can be confined to a smaller region of a sound scene, and may be modified by a scaling transformation applied to all spatial coordinates as shown in Figure 4h. For systems where elevated speakers are used, the same principle of spatial width can be applied to the elevated dimension to explore the attribute of elevated ASW. The spatial depth of the sound scene is determined by the incorporation of multiple distance cues as shown in Figure 3b. Immersion is most prevalent in sound shapes that exhibit the layering and spreading of spectra across space, and may also be enhanced through high-frequency spatial texture, increases of apparent source width (ASW), and increasing the folding (particularly using non-linear and random distributions) of spectral bands across the sound scene. Spectral quality is ultimately determined by the instantaneous state of how spectra are distributed spatially across the sound scene. Spectral width is determined by the range of frequency used. The way in which these evolve over time may be influenced by spatial transformations and by the exploration of low- or high-frequency spatial texture.

---

\(^1\) It is worth noting here that Dr Hyunkook Lee at the University of Huddersfield is currently researching the perceptual rendering of vertical image width for 3D multichannel audio as part of an EPSRC-funded research project.
5. CONCLUSION

Previous research by the author involved inquiry into the sound shapes produced by the multi-point spatial diffusion of independent spectral bands. Through the process of exploring these possible sound shapes, a range of multi-point distributions emerged making it possible to form a categorical set of distinct multi-point distributions related specifically to the spatial attributions outlined by Kendall, Rumsey, Szadov et al., Smalley, and Wishart. Delineation has been made based not only on the contiguous versus non-contiguous nature and the correlated versus de-correlated nature of these sound shapes, but mention has been made of their perceptive attributions with respect to audio scene analysis and Bregman’s theory of auditory streaming; that is, whether the sound shapes are perceived as comprising of separate streams through a process of stream segregation (i.e. non-contiguous sound shapes comprised of spatial zones), or whether these fuse together to form a single spatial image (i.e. contiguous sound shapes). Relevant spatial attributes are also tabulated as means of expressing the spatial domain from which these shapes can be transformed by spectromorphology and spatiomorphology. Whilst this article presents a pragmatic categorical list, future research would benefit from blindfold listener evaluations of such distributions in order to further clarify the delineation of such sound shapes with respect to cognitive awareness and auditory perception, and further refine the language used to define such distributions.

6. REFERENCES


Lippe, C. & Settel, Z. 1999. Low Dimensional Audio Rate Control of FFT-Based Processing. Institute of Electrical and Electronics Engineers (IEEE) ASSP Workshop, Mohonk, pp. 95-98.


James, S. 2015a. Spectromorphology and Spatiomorphology: Wave Terrain Synthesis as a Framework for Controlling Timbre Spatialisation in the Frequency-Domain (Ph.D Exegesis, Edith Cowan University)


Rumsey, F. 2002. Spatial Quality Evaluation for reproduced Sound: Terminology, Meaning, and a Scene-


