2016

Multi-point nonlinear spatial distribution of effects across the soundfield

Stuart James
Edith Cowan University, s.james@ecu.edu.au

James, S. (2016). Multi-Point Nonlinear Spatial Distributions of Effects across the Soundfield. In *Proceedings of the International Computer Music Conference* (Vol. 130, p. 129). Available here.

Multi-Point Nonlinear Spatial Distribution of Effects across the Soundfield

Stuart James
Edith Cowan University
s.james@ecu.edu.au

ABSTRACT

This paper outlines a method of applying non-linear processing and effects to multi-point spatial distributions of sound spectra. The technique is based on previous research by the author on non-linear spatial distributions of spectra, that is, timbre spatialisation in the frequency domain. One of the primary applications here is the further elaboration of timbre spatialisation in the frequency domain to account for distance cues incorporating loudness attenuation, reverb, and filtration. Further to this, the same approach may also give rise to more non-linear distributions of processing and effects across multi-point spatial distributions such as audio distortions and harmonic exciters, delays, and other such parallel processes used within a spatial context.

1. INTRODUCTION

Controlling large multi-parameter systems has always been bound by evaluating on one side the extremities of performer specificity versus generality on the other. Is it possible to intentionally control thousands of parameters simultaneously in performance, particularly when each parameter may require an assortment of attributes such as source localization, source distance, source width, loudness, and frequency? Certainly traditional approaches to live performance using a standard mixing console present difficulties when diffusing multiple sound sources across a multi-loudspeaker system. As Jonty Harrison (2005) has stated on this issue:

> If you've got an eight-channel source, and every channel of the eight has a fader, how do you do crossfades? You haven't got enough hands! (Mooney (Ed.), 2005, Appendix 2) [1]

The author proposed a solution that involved mapping audio signals to some audio-rate multi-channel panning routines developed by the author.[1] The use of audio signals for control allowed for both synchrony and adequate timing resolution, without necessarily compromising data precision. Three audio signals were used to determine the spatial localization cues azimuth, distance, and elevation/zenith. These often comprised of a vector of Cartesian (x, y, z) coordinates. In order to control the state of

---

[1] The author implemented audio-rate models of both Ambisonic Equivalent Panning (AEP) and Distance-based Amplitude Panning (DBAP)

independent spectra, these audio signals are de-interleaved. For example, to control 1024 spectral bands independently, 1024 parameter values are de-interleaved every 1024 audio samples [2].

The author also extended this to include a table lookup stage that would be used to determine how frequencies are distributed across space. In this way, a graphics file or video could be used to control this distribution in real-time. This novel process was described by the author as using *Wave Terrain Synthesis* as a framework for controlling another process, in this case timbre spatialisation in the frequency domain [3, 4].
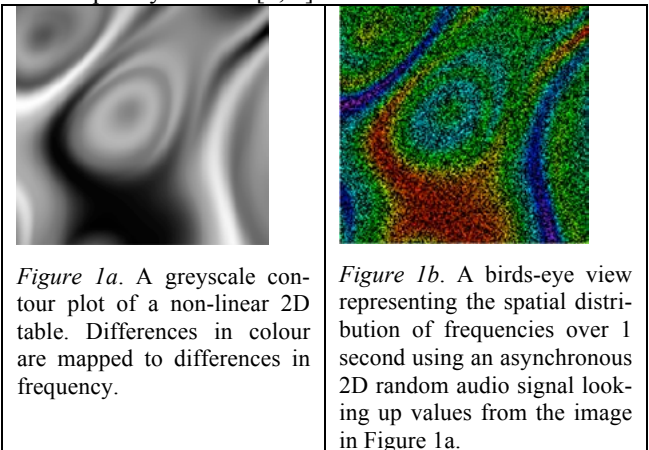

Figure 1a. A greyscale contour plot of a non-linear 2D table. Differences in colour are mapped to differences in frequency.

Figure 1b. A birds-eye view representing the spatial distribution of frequencies over 1 second using an asynchronous 2D random audio signal looking up values from the image in Figure 1a.

Schumacher and Bresson (2010) use the term 'spatial sound synthesis' to denote any sound synthesis process that is extended to the spatial domain [5]. Whilst timbre spatialisation [4, 10] falls into this category, other techniques include spatial swarm granulation [6], sinusoidal modulation synthesis [7], spectral spatialisation [8, 9], and spatio-operational spectral (SOS) synthesis [11].

2. TIMBRE SPATIALISATION IN THE FREQUENCY DOMAIN

The use of *Wave Terrain Synthesis* for controlling such a system relies on both the state of a stationary or evolving audio-rate *trajectory*, and the stationary or evolving state of a haptic-rate *terrain*. In this section some of these combinations of terrain and trajectory types are discussed in practice before extending the process to explore the impression of distance cues and other increasingly more non-linear approaches to spatial effects. Generally the results fall into the immersive category, but results can also be quite localised too.

For a single stationary *trajectory* over a colored terrain surface (a density plot using the color spectrum to describe the contour) only a single band of frequency is produced in the relative position of the virtual stationary point as shown in Figure 2a. Figure 2b shows the spectral processing functions (SPFs) that are produced for the four loudspeakers. These are color coded to illustrate the spectral distribution for each speaker. Since the point is closest to speaker 4 in Figure 2a, most of the energy accumulates in one speaker as shown in Figure 2b. In this case the amplitude ratio of this frequency is over 30 times, correlating with an increase in level of approximately +30 dB.
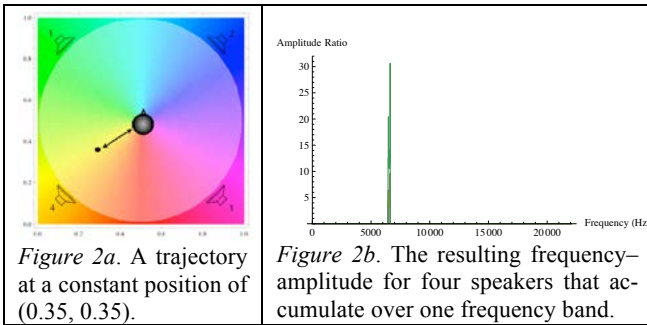

Figure 2a. A trajectory at a constant position of (0.35, 0.35).

Figure 2b. The resulting frequency–amplitude for four speakers that accumulate over one frequency band.

For a circular *trajectory* across the listener field, synchronized to the frequency of the FFT, and such that the radius is equidistant about the virtual central (ideal) listening position, generates an even spread of frequencies around the listener as shown in Figure 3b. We notice here that there are four bands of frequency separated by the speakers with which they coincide. The panning algorithm ultimately determines the relative amplitude weighting of components across the speaker array. After the smoothing process (spectral centroid smoothing and linear-phase filtration) the frequency bands shift in level to a generalised weighting of four or an increase of +12 dB. Since this difference is substantial, the smoothing algorithms adopt an auto-normalise option that recalibrates automatically for large level differences introduced by the spatialisation process. This is calculated based on the relative loudness of the input source to be spatialised, and the resulting output level of the multi-channel audio.
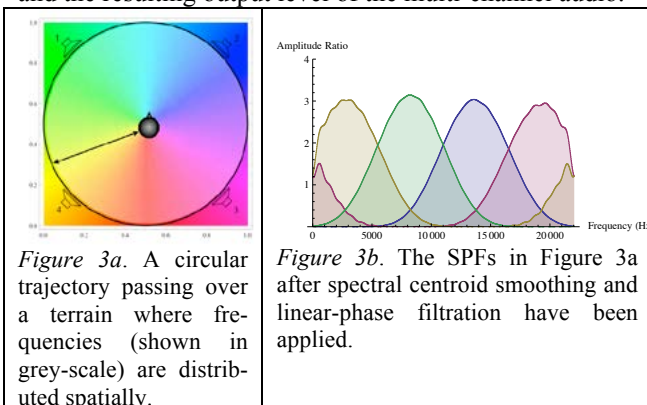

Figure 3a. A circular trajectory passing over a terrain where frequencies (shown in grey-scale) are distributed spatially.

Figure 3b. The SPFs in Figure 3a after spectral centroid smoothing and linear-phase filtration have been applied.

A very unique outcome arises when the *terrain* and the *trajectory* curve are symmetrical about the vertical or horizontal axes, resulting in the same SPF being produced in multiple speakers. Any asymmetry in either the terrain or trajectory will result in different SPF functions for all speakers. Figure 4 shows a scenario where the

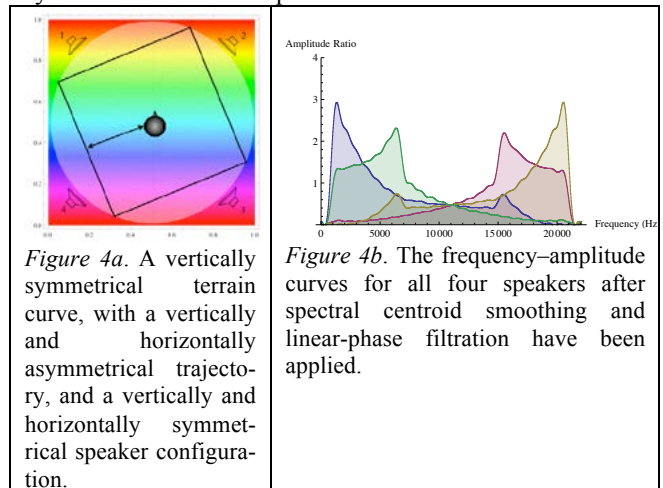SPFs for all speakers are different, yet still exhibit some symmetrical relationships.

Figure 4a. A vertically symmetrical terrain curve, with a vertically and horizontally asymmetrical trajectory, and a vertically and horizontally symmetrical speaker configuration.

Figure 4b. The frequency–amplitude curves for all four speakers after spectral centroid smoothing and linear-phase filtration have been applied.

This scenario does not apply to terrain surfaces and/or trajectories that are not symmetrical over the horizontal or vertical axes. Sound shapes generated by non-symmetrical relationships result in all speakers having vastly different timbres as shown in Figure 5.
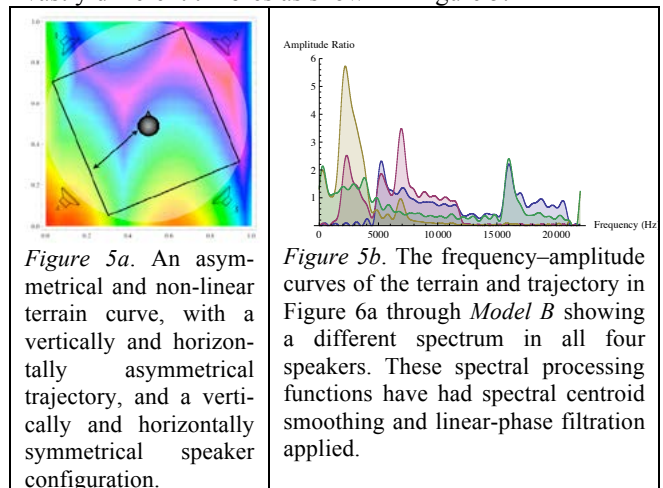

Figure 5a. An asymmetrical and non-linear terrain curve, with a vertically and horizontally asymmetrical trajectory, and a vertically and horizontally symmetrical speaker configuration.

Figure 5b. The frequency–amplitude curves of the terrain and trajectory in Figure 6a through *Model B* showing a different spectrum in all four speakers. These spectral processing functions have had spectral centroid smoothing and linear-phase filtration applied.

Noisier signals increase the potential for describing a sound shape in more detail due to their more effective space-filling properties. Figure 6 shows a high-frequency asymmetrical trajectory used over a non-linear and asymmetrical terrain curve, resulting in a much more detailed series of SPFs generated.
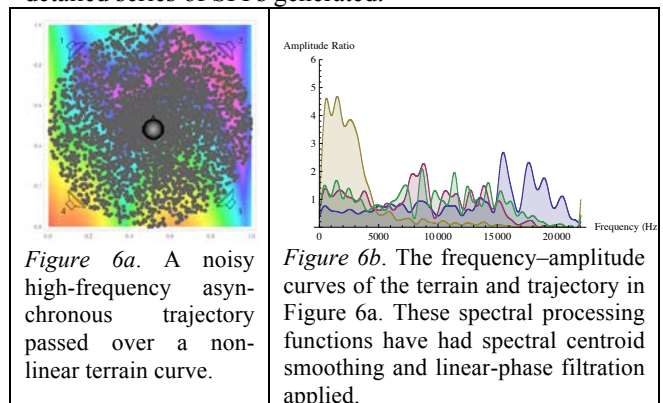

Figure 6a. A noisy high-frequency asynchronous trajectory passed over a non-linear terrain curve.

Figure 6b. The frequency–amplitude curves of the terrain and trajectory in Figure 6a. These spectral processing functions have had spectral centroid smoothing and linear-phase filtration applied.

The spatial resolution of these sound shapes can increase drastically with larger numbers of loudspeakers. In Figure 7, we see the same contour distributed between 1, 2, 8 and 32 speakers. The higher the number of loud-

speakers, the more spatial resolution, hence the spectral bands become increasingly separated. This enables the frequency response curves to represent the states 'in between'. As the number of speakers increases we observe increasing detail in each subsequent area of the spatial field determined by their respective set of SPF functions.
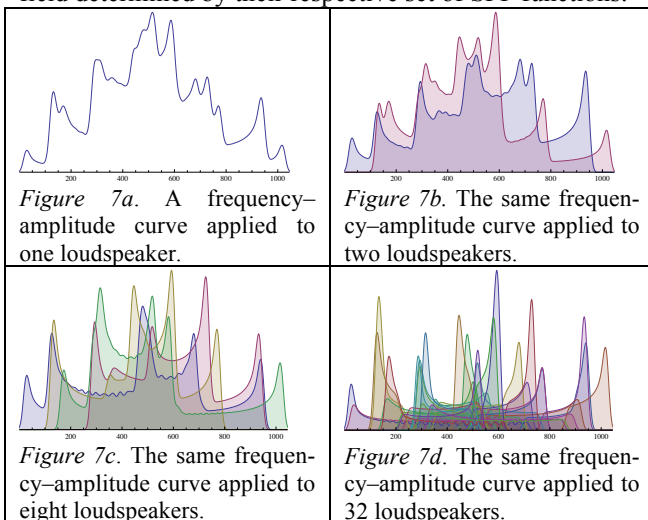

Figure 7a. A frequency–amplitude curve applied to one loudspeaker.

Figure 7b. The same frequency–amplitude curve applied to two loudspeakers.

Figure 7c. The same frequency–amplitude curve applied to eight loudspeakers.

Figure 7d. The same frequency–amplitude curve applied to 32 loudspeakers.

3. DISTANCE CUES

One of the further lines of inquiry that emerged from this research involved integrating distance cues into such a model. What is commonly referred to as 'localisation' research is often only concerned with the direction of a source, whereas the perceived location of a sound source in a natural environment has two relatively independent dimensions—both direction and distance [12]. *Interaural intensity differences* (IIDs), *interaural time difference* (ITDs), and *spectral cues*, are significant in establishing a source sound's direction, but they do not take into consideration the perception of distance.[2] 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 [13].

Most software implementations that simulate direction and distance cues do not take into consideration the wide number of indicators for perceiving distance, as the algorithms responsible for panning sources (generally) only take into consideration differences in loudness; that is, they are often simply matrix mixers that control the various weights, or relative loudness, assigned to different speakers. However there is a small number of software implementations designed to additionally incorporate some of these other indicators for distance perception. These include implementations like *ViMiC* [14], *Spatialisateur* [15], and *OMPrisma* [16]. For example, *OMPrisma*, by Marlon Schumacher and Jean Bresson [17], includes pre-processing modules to increase the impression of distance and motion of a sound source. The effect of air absorption is accounted for using a second-order Butterworth low-pass filter, Doppler effects are simulated using a moving write-head delay line, and the decrease in amplitude (as a function of distance) is accomplished with a simple gain-stage unit.

3.1 Spatial Width

In addition to the spatial localization cues azimuth, distance, and elevation/zenith, the panning algorithms used in this research also included a further parameter determining the spatial width of each spectral bin. Spatial width[3] is considered to be another significant perceptible spatial attribute, and is defined as the perceived spatial dimension or size of the sound source [19]. The spatial width of sound sources is a natural phenomenon; for example, a beach front, wind blowing in trees, a waterfall and so on. Spatial width was incorporated in the model after observing the same approach used in implementations of Ambisonic Equivalent Panning, such as ICST[4] ambisonic panners for MaxMSP [18]. It should be made clear that Ambisonics algorithms do not render distance cues, however documentation by Neukom and Schacher [20] and its implementation in the ICST Ambisonics library demonstrate how the algorithm has been extended to account for distance. One of these relationships is the binding of spatial width with the distance of a sound source. The ICST implementation binds the order of directivity to the distance of each point, so as sources move further away from the centre they become narrower, and when they move closer they are rendered with greater spatial width, and if they are panned centre they are omnipresent. This is all dependent on the order of directivity of the AEP algorithm, as shown in Figure 8. Applying this at audio-rates with a polyphonic parameter system, like spectral processing, creates a complex spatial soundfield where different spectral bands have different orders of directivity.

Similarly other panning techniques such as Distance-based Amplitude Panning (DBAP) have provision for the amount of spatial blurring, which inadvertently increases the immersive effect, effectively spreading localized point-source movements to zones or regions of a multi-speaker array. Again, each spectral band can be rendered with a different spatial blur, resulting in a complex multi-parameter organization.
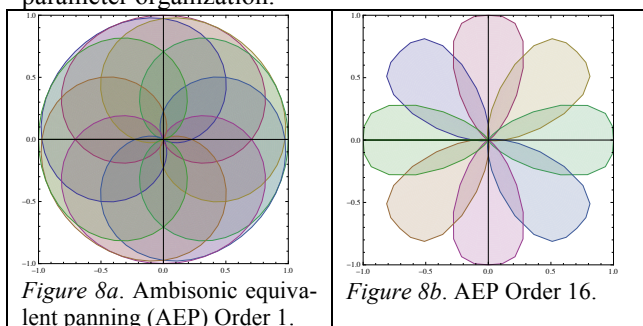

Figure 8a. Ambisonic equivalent panning (AEP) Order 1.

Figure 8b. AEP Order 16.

Whilst this could be determined solely by the radial distances of the intended diffusion, a further lookup stage could be used to determine spatial width across a 2D plane, either by a conventional circular distribution as

---

[2] The attributes that assist in the perception of distance are sometimes referred to as *distance quality*.

[3] Also referred to in psychoacoustic literature as *spatial extent*, *source width* or *tonal volume*.

[4] The Institute for Computer Music and Sound Technology in Zürich, Switzerland.

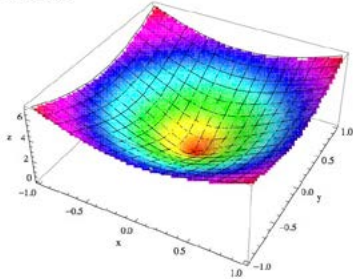shown in Figure 9, or one that is significantly more non-linear.

Figure 9. A circular distribution determining the order of directivity for different spatial coordinates (x, y).

3.2 Loudness

The role of loudness with respect to the perception of distance is inextricably linked with a sound sources relative drop in energy over distances, measured in decibels per metre (dB/m). The inverse distance law states that sound pressure (amplitude) falls inversely proportional to the distance from the sound source [24]. Distant sound sources have a lower loudness than close ones. This aspect can be evaluated especially easily for sound sources with which the listener is already familiar. It has also been found that closely moving sound sources create a different *interaural intensity difference* (ILD) in the ears than more distance sources [13].

However before considering the relative amplitudes generated across the multichannel system, we have to consider the amplitudes generated for each loudspeaker, keeping in mind the non-linearities of the panning algorithms used. For example, a complicating factor for the AEP model is that when incorporating more loudspeakers, and also modulation of the order of directivity, the resulting amplitude ranges change drastically too. Therefore implementations such as ICST account for both centre attenuation (dB) and distance attenuation (dB) (as well as the centre size). Centre attenuation is required to counteract the order of directivity when it is 0.[5] The distance attenuation serves to ensure that for larger virtual distances, the appropriate roll-off is applied. Some distance attenuation curves, with their associated parameter settings, are shown in Figure 10.
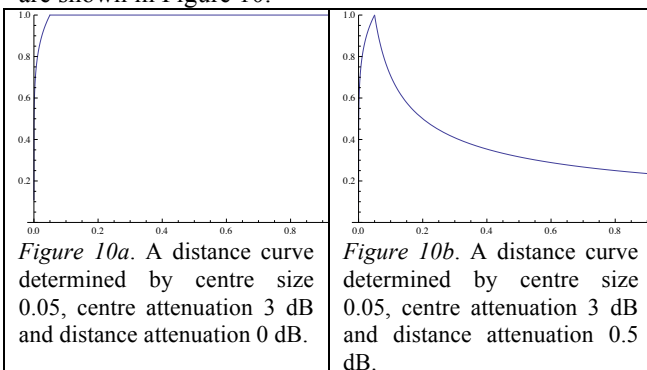

Figure 10a. A distance curve determined by centre size 0.05, centre attenuation 3 dB and distance attenuation 0 dB.

Figure 10b. A distance curve determined by centre size 0.05, centre attenuation 3 dB and distance attenuation 0.5 dB.

The frequency–amplitude curves generated in some cases can feature strong energy on certain bands of fre-

---

[5] When the order of directivity is 0, the amplitude is 1 in all loudspeakers. Therefore for larger loudspeaker systems this accumulates based on number of speakers used.

quency, and this ultimately depends on the rate of change of the trajectory curve. In other words, stationary points in the *terrain* or *trajectory* are the reason for this accumulation of energy in certain regions of the frequency spectrum (see Figure 11a). Calibrating appropriate loudness attenuation curves across this 2D (or 3D system in the case of elevated cues) depend on relatively linear distributions of frequency across space. In order to achieve this, tests involved the use of a flat linear terrain surfaces, and a 2D random audio-rate trajectory with effective space-filling properties. Calibration of the distance as applied to *timbre spatialisation* can be achieved using the combination of a white noise *trajectory* over a simple linear *terrain* function. Figure 11b shows the standard frequency–space visualisation used in the authors research, and the ideal position of a listener (centre), where the distance of low frequencies highlighted (above) and low frequencies (below) are more distant than the mid-range frequencies (in the middle) that should sound perceptively louder.
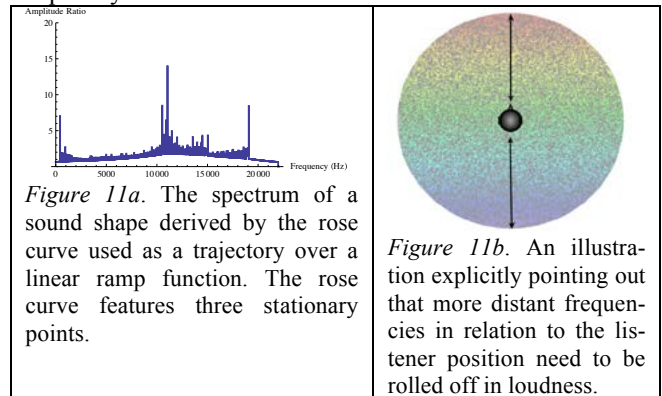

Figure 11a. The spectrum of a sound shape derived by the rose curve used as a trajectory over a linear ramp function. The rose curve features three stationary points.

Figure 11b. An illustration explicitly pointing out that more distant frequencies in relation to the listener position need to be rolled off in loudness.

By reading the resulting frequency–amplitude curves from this process, it is possible to determine to what extent frequencies that are further away from the centre position are attenuated as a result of their relative distance from the listener, as shown in Figure 12a. These frequency–amplitude curves can be used to calibrate the distance roll-off curve and centre size of AEP. The combined use of the centroid smoothing and a linear-phase low-pass filter can also help to smooth out the peaks in the SPF in order to better gauge the roll-off in each instance. These smoothed frequency–amplitude plots are shown in Figures 12b. With a centre size of one and a roll-off of 3 dB, the impression of distance is subtle but evident. The use of the low-pass filter can also remove the comb filtering effects of the SPFs that result from computing the histogram.
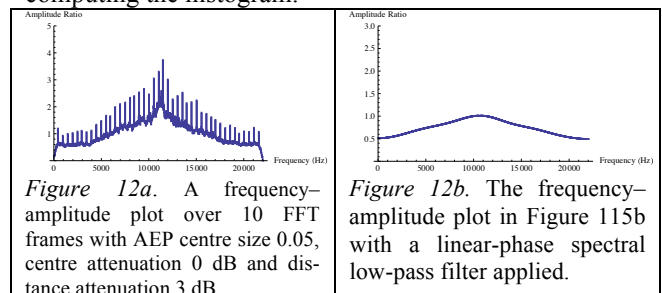

Figure 12a. A frequency–amplitude plot over 10 FFT frames with AEP centre size 0.05, centre attenuation 0 dB and distance attenuation 3 dB

Figure 12b. The frequency–amplitude plot in Figure 115b with a linear-phase spectral low-pass filter applied.

As is the case with encoding spatial width, a 2D or 3D table can be used to lookup the relative loudness (or amplitude scaling) over a nominal distance.

3.3 Air Absorption

The sound spectrum can also be an indicator of distance since high frequencies are more quickly damped by air than low frequencies. Consequently, a distant sound source sounds more muffled than a close one, due to the attenuation of high frequencies. For sound with a known and limited spectrum—for example, human speech—the distance can be estimated roughly with the listener's prior knowledge of the perceived sound [25]. The implementation here effectively involves a parallel process that would essentially split the spectral bands based on a distance ratio. This involves an amplitude scaling function that is applied as the SPF functions are generated for each respective loudspeaker. By separating the spectra into two groups, one can be left a group of spectra that are unaffected (dry), whilst the other group is processed in some way (wet). In the case of air absorption, this would involve convolution filtering of the parallel group in order to attenuate high frequencies. As a result of this, perceptively the processing would appear to be applied increasingly more for distant spectra.

Figure 13a. Amplitude scalar for unprocessed spectra

Figure 13b. Amplitude scalar for processed spectra
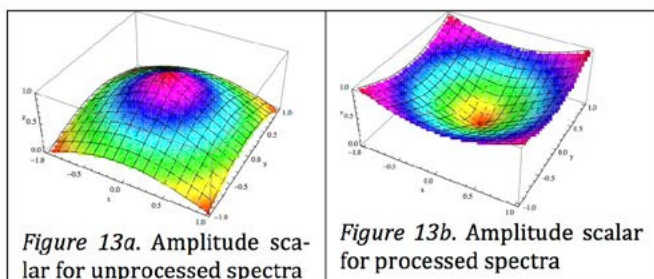
3.4 Direct versus Reflection Ratio

The direct *v.* reflection ratio is a phenomenon that applies mostly to enclosed rooms and spaces. Typically two types of sound arrive at a listener: the direct sound source and the reflected sound. Reflected sound is sound that has been reflected at least once at a wall before arriving at the listener. In this way the ratio between direct sound and reflected sound can be an indicator of the distance of the sound source [13].

A way to integrate reverberation in such a multi-point model could be achieved in a similar way to the application of convolution filtration for simulating the effects of air absorption over larger distances. By separating the spectra into two groups, a dry and wet multi-point set, it is possibly to apply reverberation proportionally to the distant of each point of sound spectra from the central listening position. The amount of reverberation applied is therefore dependent on the distance quality of each frequency band.

The reverberation used may also allow for some adjustments in terms of the ratio of early reflection versus reverb tale, as well as the amount of pre-delay applied to the early reflections. If the pre-delay is short it may be indicative of a more distant sound source, versus a longer pre-delay indicating a first reflection that is heard off a nearby wall. This is often referred to as the Initial Time Delay Gap (ITDG). The ITDG describes the time diffe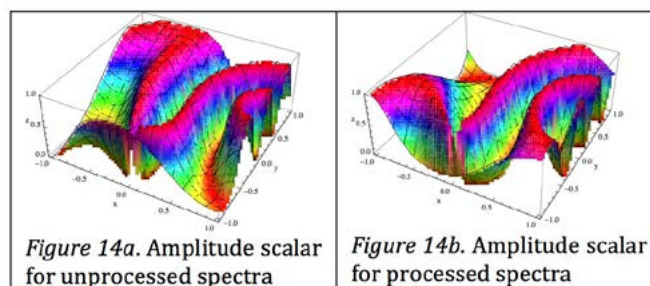rence between the arrival of the direct sound and first strong reflection at the listener. Nearby sound sources create a relatively large ITDG, with the first reflections having a longer path to the listener. When the source is far away, the direct and reflected sound waves have more similar path lengths.

The ITDG can be compensated for with the use of spectral delays, such that more distant frequency bands will be subjected to a different ITDG than a frequency band that is, in a virtual sense, closer to the listener. This aspect adds considerably more awareness of depth in the resulting spatialisation.

4. NONLINEAR SPATIAL DISTRIBUTION OF AUDIO EFFECTS

Another outcome of this same parallel process is firstly they could be used to apply other kinds of effects to a multi-point spatial distribution, and secondly they don't have to follow a distribution that is dependent on a central listener position, but rather aimed at exploring immersive and evolving transitions of effects such as delays, distortions, harmonic exciters, over a soundfield.

The fundamental process is the same here, where a spectral distribution is separated into an unprocessed group and a processed group. Figure 14 shows some nonlinear ways in which such a parallel process could manifest over a complex spatial sound shape.

Figure 14a. Amplitude scalar for unprocessed spectra

Figure 14b. Amplitude scalar for processed spectra

5. CONCLUSIONS

Exploration of techniques that evoke a stronger sensation of distance in multi-point spatialisation, such as timbre spatialisation in the frequency domain, have resulted in more engaging spatial sound shapes with a stronger sense of depth over the soundfield. By applying some of these processes in parallel, it was also found that the same approach could be used to control other signal processes that are not specifically distance-dependent, but follow some other more novel and non-linear distribution across the soundfield. Further research could be focused on the movement of sound sources, particularly the effect known as 'Doppler shift'. The source *radial velocity*—the speed of a sound source moving through space—will affect the pitch of the sound due to the compression or expansion of the sound's wavelength as it travels through the air towards the listener [22]. Such effects may be possible through frequency modulating specific partials through the use of specific all-pass filters [23]. Furthermore, blindfold listener evaluation of such effects are essential in both evaluating the effectiveness, and optimizing the perceived effect of such processes.

6. REFERENCES

[1] Mooney, J. (Ed.) (2005). An Interview with Professor Jonty Harrison, In J. Mooney *Sound Diffusion Systems for the Live Performance of Electroacoustic Music* (Appendix 2) (Unpublished doctoral thesis), University of Sheffield. Retrieved from http://www.james-mooney.co.uk/publications (accessed May 15 2011).

[2] James, S. (2016). A Multi-Point 2D Interface: Audio-rate Signals for Controlling Complex Multi-Parametric Sound Synthesis. *Submitted to New Interfaces for Music Expression.*

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

[4] James, S. (2015). Spectromorphology and Spatiomorphology of Sound Shapes: audio-rate AEP and DBAP panning of spectra. *Proceedings of the 2015 International Computer Music Conference*, Denton, Texas.

[5] Schumacher, M. & Bresson, J. (2010). Compositional Control of Periphonic Sound Spatialization. *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*.

[6] Wilson, S. (2008). Spatial Swarm Granulation. *Proceedings of the 2008 International Computer Music Conference*. Belfast.

[7] Cabrera, A. & Kendall, G. (2013). Multichannel Control of Spatial Extent Through Sinusoidal Partial Modulation (SPM). *Proceedings of the Sound and Music Computing Conference 2013*, Stockholm, 532-537. Retrieved from http://smcnetwork.org/system/files/MULTICHANNEL%20CONTROL%20OF%20SPATIAL%20EXTENTTHROUGH%20SINUSOIDAL%20PARTIAL%20MODULATION(SPM).pdf (accessed January 10 2015).

[8] Kim-Boyle, D. (2006). Spectral and Granular Spatialization with Boids. *Proceedings of the 2006 International Computer Music Conference*, New Orleans, 139-142.

[9] Kim-Boyle, D. (2008). Spectral Spatialization: An Overview. *Proceedings of the 2008 International Computer Music Conference*, Belfast, 1-7.

[10] Normandeau, R. (2009). Timbre Spatialisation: The Medium is the Space. *Organised Sound,* 14(3).

[11] Topper, D., Burtner, M. & Serafin, S. (2002). Spatio-Operational Spectral (S.O.S.) Synthesis. *Proceedings of the 5th International Conference on Digital Audio Effects*, Hamburg, Germany.

[12] Kendall, G. & Martens, W. L. (1984). Simulating the Cues of Spatial Hearing in Natural Environments. *Proceedings of the 1984 International Computer Music Conference*, Paris, 111-126.

[13] Howard, D. & Angus, J. (2009). *Acoustics and Psychoacoustics: Fourth Edition*. Burlington, MA: Focal Press.

[14] Peters, N., Matthews, T., Braasch, J., & McAdams, S. (2008). Spatial sound rendering in Max/MSP with ViMiC. *Proceedings of the 2008 International Computer Music Conference*, Belfast.

[15] IRCAM (Institut De Reserche Et Coordination Acoustique), Retrieved 10th Jan 2015 from http://www.ircam.fr/1043.html?&L=1

[16] Bresson, J. (n.d.). *bresson:projects:spatialisation*. Retrieved 10th Jan 2015 from http://repmus.ircam.fr/bresson/projects/spatialisation

[17] Schumacher, M. & Bresson, J. (2010). Compositional Control of Periphonic Sound Spatialization. *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*.

[18] The Institute for Computer Music and Sound Technology. (n.d.). *ZHdK: Ambisonic Externals for MaxMSP*. Retrieved 10th Jan 2015 from https://www.zhdk.ch/index.php?id=icst_ambisonicsexternals

[19] Potard, G. & Burnett, I. (2004). Decorrelation Techniques for the Rendering of apparent Sound Source width in 3D audio displays. *The 7th International Conference on Digital Audio Effects*.

[20] Neukom, M. & Schacher, J. (2008). Ambisonics Equivalent Panning. *Proceedings of the 2008 International Computer Music Conference*, Belfast.

[21] Lossius, T., Baltazar, P. & de la Hogue, T. (2009). DBAP - Distance-Based Amplitude Panning. *Proceedings of the International Computer Music Conference*, Montreal, 17-21.

[22] Chowning, J. (1971). The Simulation of Moving Sound Sources. *Journal of the Audio Engineering Society*, 19, 2–6.

[23] Surges, G. & Smyth, T. Spectral Distortion Using Second-Order Allpass Filters. *Proceedings, 10th Sound and Music Computing Conference*, 2013. Stockholm, Sweden: SMC.

[24] Everest, F. A, & Pohlmann, K. (2014). Master Handbook of Acoustics, Sixth Edition. McGraw-Hill Education, TAB.

[25] Harris, C. (1966). The Absorption of Sound in Air versus Humidity and Temperature. *The Journal of the Acoustical Society of America*, 40.