Three Approaches to the Dynamic Multi-channel Spatialization of Stereo Signals

Dr. Christopher J. Keyes
Department of Music and Fine Arts, Hong Kong Baptist University
ckeyes@hkbu.edu.hk

Abstract
The dynamic multi-channel spatialization of stereo signals can bring a new dimension to the sound diffusion of acoustical instruments in concert performances as well as providing depth and user interactivity to the routine listening of sound recordings at home. This paper examines three approaches to this particular aspect of sound spatialization, using: 1) a time-domain granular synthesis model, 2) a frequency-domain spectral balancing algorithm, and 3) an amplitude panning array. Emphasis will lie not in recreating previous perceptual experiences but on the creative potential for new experiences, even with older music.

1 Introduction
Thanks largely to the movie industry, multi-channel sound is currently enjoying something of a renaissance, an especially appropriate term in that listeners are now more frequently enjoying the kind of antiphonal listening experience Giovanni Gabrielli’s audience did in the late 16th century. One major difference, however, lies in our current ability to simulate an instrument’s trajectory through a three dimensional space and, with real-time FFT-based analysis/resynthesis methods, even project different regions of an instrument’s harmonic spectra to different virtual locations.

For various reasons, however, most approaches to multi-channel audio have focused either on the localization of monophonic sound sources in 2 and 3 dimensional spaces, or on environmental simulation. My own interest in creating rich spatializations of acoustical instruments in the concert hall have lead to approaches that employ stereo images with, I believe, excellent results. Software written that employ these approaches have also transformed my own enjoyment of recorded music from static stereo experiences to greatly enhanced, dynamic, and interactive multi-channel experiences. This paper will explore three such approaches, their effects, and their possible applications to alternate listening situations.

2 Advantages of using stereo sources, multiple microphones, and signal decorrelation
When diffused to multiple channels, sound spatialization and real-time digital signal processing of acoustic instruments using multiple microphones has numerous advantages over monophonic miking and diffusion, many of which originate from the decorrelation of the audio signals (here used to denote differentiated signals that sound identical). Gary Kendall discusses these at length in “The Decorrelation of Audio Signals and Its Impact on Spatial Imagery” (1995). In that article he highlights the effect of decorrelated audio on the creation of diffuse sound fields without reverberation, the reduction in perceptivity of combing and coloration (due to constructive and destructive interference), reduction of image shift from different listening locations, and reduction of the precedence effect (Haas 1951) or the law of the first wave front. (Lindemann 1986) The later two effects are also noted in “Volumetric Modeling of Acoustic Fields…” (Kaup et al., 1999) as researchers at CNMAT grappled with optimizing the greater listening area of their Sound Spatialization Theatre.

Other reasons to explore multiple miking techniques in the concert hall lie squarely on the production of a more interesting and dynamic sound image, which our ears seem to enjoy. In an instrument such as the piano, for example, microphones can be placed over the upper and lower strings such that if the pianist were to play a scale, even with fixed left/right panning, the audience would experience the sound apparently moving from one speaker to the other. This rather simple micro-spatialization, when subjected to additional processing can lead to a very rich final spatialization for which monophonic signals would fall short. Often two microphones on an instrument such as the flute can better capture the sound radiation patterns inherent to the instrument and provide a fresh listening experience, as well as the decorrelated advantages mentioned above.

Yet another reason to investigate the multi-channel spatialization of stereo sound sources lies precisely in their ubiquity. Record shops and home listening libraries abound with recorded stereo images, most of which are the result of
careful consideration and balancing. With the spatialization techniques that follow, this careful work can be preserved while still providing an enhanced experience of it to the listener, and with listener interactivity.

3.1 Granularization Approach

In previous work with real-time granular time expansion/compression I found the effect of multi-channel spatialization of each of the grains became one of the most rewarding aspects of the work (Keyes 2003). To be useful as a spatialization tool and divorced from the (then sought after) effects of time stretching, however, required a rethinking of the grain engines and particularly the kind of buffers used.

Although Curtis Roads gives a far more comprehensive explanation of granular synthesis and granulation in his book “Microsound” (2001), the basic approach can be seen in figure 1 below.

![Figure 1. Each grain window is sent to a different channel.](image)

The program breaks each channel of an incoming time-domain signal into windowed grains and allocates each grain to a particular output channel. Each grain is controlled by a separate engine (see figure 2) which also allows for an independent time delay, helpful for time alignment of the speaker system as well as further decorrelation, and transposition of each grain up to +/-25 cents. With the addition of just a few cents difference between grains the apparent spatialization becomes far more exaggerated. Note that Max/MSP’s ‘tapin~’ and ‘tapout~’ objects create and read from a circular buffer/delay line.

![Figure 2. Grain engine for granularized spatialization realized in the Max/MSP environment.](image)

As can be seen from the user interface (figure 3) the program also allows control of ‘Grain Density’, how far apart each window is from the previous window, and ‘Random Grain Position’, helping to smooth modulation effects with smaller grain sizes and greater densities.

To realize this with stereo sound sources while maintaining (to a surprisingly large degree) the original stereo image, grain engines are paired to the ‘left’ and ‘right’ input channels (for the purposes of this paper I will use the terms ‘left’ and ‘right’ to denote the different stereo input channels), and then sent to stereo pairs in the performance space. For this approach a horse-shoe or cubic array of loudspeakers is often optimal.

With grain sizes between 20 and 100 milliseconds the granularized output produces the typical choppy, blurry sound associated with granularization. Lower grain densities and higher random grain positions achieve a kind of reverberant sound also noticed by Curtis Roads (2001). The higher the random grain position the more the output resembles reverberation in an asymmetrical space.

However larger grain sizes, modest densities, and lower random positions leave the sound relatively unaltered. Although there seems to be no real change in the sound, the amplitude envelop of each of the grains results in each channel having its own unique articulation of the sound. When spatialized to two or more paired channels, the effect becomes one of much greater depth with a pronounced enveloping quality and a multi-directional dynamic.

![Figure 3. User interface for the 4 channel version of the Granular Spatializer.](image)

3.2 Frequency-domain Approach

This approach uses a FFT-based analysis/resynthesis method which, as the processing power of personal computers rise and prices drop, are becoming ever more practical. Ryan H. Torchia and Cort Lippe discuss the spatial redistribution of spectral energy in their “Techniques for Multi-Channel Real-Time Spatial Distribution Using Frequency-Domain Processing” presented at the ICMC
2003 in Singapore (Torchia 2003). They use the Max/MSP “Multislider” to pan specific frequency bins of a monophonic input to either stereo or quad locations. They also note inherent problems in multi-dimensional controllers, interfaces, and representation using this method for more than two panning locations.

The Spectral Spatializer takes slightly different approach using frequency-domain filtering to spatialize (one could also use the term ‘animate’) stereo sources over an even number of playback channels. This can be accomplished by applying a dynamic equalization curve to one channel while applying the inverse curve to the opposite channel. Example 4 shows a Max/MSP engine for accomplishing the filtering, similar to Z. Settel and C. Lippe’s Forbidden Planet example in Max/MSP (Zicarelli 1997).

![Figure 4. Engine for frequency-domain filtering by convolution of the input signals with signals based on equalization curves.](image)

The left signal is convolved with a frequency-domain signal based on the equalization curve indicated in dark blue on the graphic user interface (shown in example 5) while the right channel is convolved with the inverse of that curve, automatically drawn in light blue on the opposite graph. These allow real-time control of up to 256 frequency bins and can be drawn either with a mouse (see example 5 ‘Ch 1-2’), with any 2-dimensional MIDI controller, or via automation, parsing through the table bin by bin. Automation controls include a ‘random walk’ based on the ‘drunk’ object (see example 5 ‘Ch 3-4’) as well as sine and saw-tooth wave forms.

![Figure 5. GUI of Spectral Spatializer with equalization curves on odd numbered output channels and their inverse on even numbered output channels.](image)

The result is a dynamic spatialization that, baring extreme equalization curves, maintains the stereo image and total spectral energies of the input source while creating a subtle but captivating aural animation of those sources. The output can also be described as having a distinct sense of motion but without image shift.

### 3.3 Amplitude Panning Array

At a concert of electro-acoustic music held in Manchester at the 1998 International Society of Contemporary Music I gained a renewed respect for the careful use of amplitude panning in the multi-channel spatialization of stereo signals. At this particular concert all of the pieces were realized for stereo playback, and all of them were diffused over a 10 channel sound system. Those responsible for the diffusion read from scores they had made to depict various aspects of change in the music and were dutifully rehearsed. The results were spectacular, perfectly matched for the sound system and acoustics of that particular hall, and I constantly found myself in disbelief that the works were actually emanating from a stereo DAT machine.

One all too common difficulty for the diffusion, however, was the lack of an interface or controller to deal with this effectively. As is often the case, those responsible for the diffusion were sitting behind a mixer with all 10 fingers posed anxiously over faders, straining to maintain even panning curves as they worked in real time.

To capture this mode of spatialization while providing a more useful user interface I used two, ‘double-quad’ amplitude panning elements, one of which is seen in figure 6. These multiply each input by a square root curve to achieve smooth signal attenuation across 4 channels of spatialization.
To preserve the stereo image the user interface takes the panning coordinates of the left signal and inverts them for the right signal (see figure 7), either along the x-axis or along both the x and y axes. If inverted along both axes the complete stereo image is maintained so long as the panning is kept along the outsides of the box. When inverted along the x-axis only the complete stereo image is maintained whenever the source is panned along the left or right edges. It should be noted that although the interface is square, when the channels are panned along the edges of the boxes the panning curve used keeps the intensity of the signal constant, thus giving the illusion of the sound traveling in an arc around listener. For a deeper and more dynamic sound field, a delay (and thus decorrelation) can be added to one of the inputs channels, and this value can vary over time.

As with the Spectral Spatializer, this implementation also includes automation controls which may direct the panning of each input along two sets of 4 channels. These automation controls include a pull-down menu for which corners the input channels will move to, and a fade time in which the move is to be accomplished. This guarantees a smooth panning time and also frees up the controllers hands to work with the other set. There are also controls for geometric patterns, including circles, ellipses, sine waves, and others. To avoid the problem of these patterns becoming too predictable a low frequency oscillator signal can be added to the signal driving the function, thus providing deviations from the prescribed curves. Experiments in making these patterns frequency and/or amplitude dependent have also proved effective, though often manual control provides the best balance in following metrics, mood, and structure of the works being spatialized.

### 3.4 The Approaches and “Performance”

This last observation brings us to a central point: that manual control (as well as selected automation) can become a part of the ephemeral realization of the work and thus a lend a different perspective on each hearing of the work. In a concert situation one can imagine a “diffusion artist” making adjustments for the particular performance space, instruments, (available play back equipment, hardware and software) and even the perceived mood of the audience, thus avoiding fixed-position spatializations that may not be optimal for a given situation. In a home listening environment one can imagine a remote control (perhaps similar to those now used for equalization) allowing the listener much the same opportunity and thus including a role for interactivity in listening to music taking it beyond a strictly passive experience. These could be especially useful for multi-channel devices such as DVD SACD players.

### 4 Conclusions

Speaking strictly of decorrelation, Kendal notes (1995): “Dynamic variation produces 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 field…” And what quality could be more inviting than liveliness to impart on a sound image? Certainly “depth” and “spaciousness” are also terms revered throughout the audio world. The approaches stated above can go a long way to achieving these effects (and without, heaven forbid, a 192 kHz sampling rate).

Any one of the three approaches to the multi-channel spatialization of stereo signals mentioned above, and of course their use in combination, can go beyond the recreation of perceptual experiences and can create new experiences that are both dynamic and variable for each hearing of a work. They also rely on equipment and resources commonly available at this point in time.
References