

# 5.1 & 2.0 Interspersion: A Reality

Robert Reams, Neural Audio

## Introduction

Until 5.1 music production technologies and techniques mature, discrete 5.1 music content will be relatively scarce. This means that much of the music being broadcast on the modern “surround station” will be “stereo” (much like earlier mono music being broadcast on stereo stations).

The addition of 5.1 content to the broadcast plant creates additional challenges as 5.1 relies on a transport format that is either non-editable or non-interoperable within the stereo infrastructure, causing a financial and technical burden that seemingly outweighs the benefit.

Broadcasting individual 5.1 and stereo mixes in the broadcast stream is doable, but creates logistic and budget challenges for both the broadcast facility and the content provider, not to mention a really small bit budget for the content. Everybody, that is, the content provider, the broadcaster, and the consumer, loses.

The solution to this “chicken and egg” problem is a system that allows the seamless interoperability of stereo and surround content.

While matrix methodologies have satisfied television and movie stereo/surround interoperability challenges for the past two decades, they do not satisfy the music distributor’s and consumer’s expectation of what is now called “discrete” 5.1.

The solution for music interoperability in the broadcast environment demands higher performance. Two-dimensional spatial compression allows the distributor or broadcaster the ability to capture original N-source content (in this case, 5.1) and “downmix” it to a 2.0 channel format.

N-channel rendering methodologies allows the consumer to render the content, regardless of the original format, in any spatial format they choose; mono, stereo, 3 channel stereo, quad, 5.1, 10.2, up to 256 channels.

In the absence of the spatial rendering feature in legacy receivers (which will be the case during transition toward 2.0/5.1 interoperability) the content is received and perceived as “stereo” on existing receivers.

## Capture and Encoding (Satisfying the content provider’s expectations)

Audio engineers have noticed for years that certain types of music content “doesn’t code well”.

This is because music, whether it is 5.1 or stereo, is not generally mixed with lossy compression in mind. This phenomenon, having little impact on linear formats, is showing itself as broadcast data rates and download time speed become considerations in the “value” of the content. Does this mean that “artistic freedom” will become corrupted by the limitations of distribution? Not necessarily. The job of the mix engineer has *always* been to capture the artist’s perspective and “translate” it to the medium of distribution. The “medium of distribution” has been in flux since the dawn of recording and reproduction and will be in flux until people are born without ears. There are simply new advantages, trade-offs and tools to understand.

### **Spatial Descriptors**

The process of downmixing original source 5.1 to 2.0 stereo is not trivial. The process is known to be mysterious and unpredictable. There doesn’t seem to be any “rules” or “procedures” that achieve good results regardless of the content. There is a general understanding of spatial descriptors but no apparent way of preserving them in the mixdown process.

To get a grasp on how to achieve an artifact free downmix, we must consider three generally accepted spatial descriptors that are used to define image in spatial audio coding. These are:

Interaural Time Difference (ITD’s). This parameter defines the difference of time between a pair of audio sources that affect the lateral displacement of an image or the mean time difference between the two ears, taken from the group delay of the impulse responses.

Interaural Level Difference (ILD’s). This parameter defines the difference of level between a pair of audio sources that also affect the lateral displacement of an image. The ILD is derived by subtracting left and right energy spectra, as computed as energy summations.

Interaural Cross Correlation (ICC). This parameter defines the measure of linear association between the left and right channels. Correlation is most easily determined by computing the centroid of coherence as a function of spectra.

### **Transforming ITD**

In order to achieve a downmix free of frequency domain artifacts, time based spatial attributes (ITD’s) must be removed and substituted with the psychoacoustic equivalent of intensity (ILD’s) for the lateral descriptor while the resultant depth component resulting from the ITD is translated into pure coherence (ICC). The advantages of this methodology are enormous.

- 1) The number of spatial descriptors has been reduced from three to two
- 2) The remaining two are statistically unrelated (orthogonal) and thus unconditionally stable
- 3) Downmixing without ITD’s eliminates frequency domain artifacts

The content is then downmixed into two fine structures. The 5.1 image envelope is transformed to a *pure two dimensional* version of the original 5.1 image envelope. The azimuth of the original image envelope is represented by inter-channel intensity differences. The front-to-back, or “depth”, of the original image envelope is represented by average inter-channel coherence (normalized cross-correlation).

### **Downmixing and Lossy Compression**

Downmixing followed by lossy compression is even more difficult. The reason is that modern

lossy compression is “spatial” in nature, that is, modern codecs such as AAC and its variants, depend on certain spatial “redundancies” to improve coding efficiency that translates into fewer artifacts.

The reduction of the original image envelope to pure 2-D descriptors allows the downmix to be performed in a way that is both stereo and mono compatible. Descriptor purity is a pre-cursor to accurate translation of the content image when rendered back to multichannel.

In addition, 2-D downmixing is an excellent choice because of its similarity to the image construct of naturally occurring 2-D stereo and compatibility with already prevalent Lo/Ro and Lt/Rt content.

Legacy stereo is easily transformed into 2-D stereo by reducing naturally occurring spatial descriptors to intensity and coherence via “image packing”. Image packing augments spatial redundancy, further improving the efficiency of the core codec.

### **Production Simplicity**

There are and always will be many “configurations” of content capture and mixing. Some methods are better than others as quality differences will always be budget and time driven.

For production “in the field”, as with live broadcasts, many engineers are mixing with portable consoles. 2-D stereo works well with quality analog or digital consoles. Routing is accomplished by buss assignment and “odd/even” panning (Buss 1-center, Buss 3/4-left/right, Buss 5/6-left surround/right surround, Buss 7-LFE, etc.) with the submaster outputs inputted to a 2-D downmix appliance. The 2-D output of the appliance is routed to the console “2-mix” either directly or through a pair of channel inputs.

If the console features “in-place” soloing (as many do), one may even audition the spatial placement of individual channels. The existing console stereo monitoring output is simply routed into a 5.1 rendering appliance and monitored in “5.1”. This is an excellent environment for content capture, dubbing, mixing/editing and stereo/mono compliance checks.

### **DAW’s**

The same scheme may be used in the quiet and safety of a control room. DAW’s and editors with adequate routing and I/O fare equally well. ‘Sacred cow’ stereo processing still functions. Nothing of value is obsoleted while new artistic vistas are enabled, all, with appropriate “reality check” facilities (stereo and mono compatibility) in place.

### **Distribution realities**

2-D stereo is naturally well suited for existing supra structures using distribution technologies typically employed within cable, internet, terrestrial and satellite broadcast. As the spatial information is contained in the *waveform* of the audio, 2-D stereo performs well under analog/digital conversion and codec concatenation. 2-D stereo, being appropriately “packed”, is coded with higher efficiency than regular stereo as the modern codec’s redundancy functions are more effectively utilized.

## **Real World**

Substantial testing has been focused on the realities of the distribution path. The following slide illustrates a real world test “fixture” for benchmarking the performance of a spatial audio coder. Note that the spatial information contained in the audio must traverse a “linear PCM-MPEG1 L2 (ISDN)-analog (BTS Router)-MPEG1 L2 (Sat)-PCM (Editor)-Linear PCM” path with the spatial information intact. While seeming severe, this is typical of the journey that spatial content will take. Slight discreteness impairments are a small price to pay for unconditional compatibility and stability of spatial content.

## **Decoding and Rendering (Satisfying the consumer’s expectations)**

Upon decoding, the image envelope of the original 5.1 content is re-synthesized based on the intensity/coherence information contained in the watermark of the two fine structures. Using this methodology, an impression of the original source 5.1 content is rendered from the two downmixed audio channels with a high degree of merit.

The 5.1 rendering is accomplished by a (patent pending) programmable, transform based spatial rendering system. SEE (Spatial Environment Engine) can render any two dimensional audio source (both 5.1 and stereo are 2-D) into as few as 2 to as many as 256 outputs with a high degree of perceived separation.

The spatial elements of 2.0 stereo (or Lt/Rt) are segregated based on the 2-D image envelope naturally residing in the content, nothing is either created or destroyed. “Re-downmixing” of the 5.1 rendering of stereo back to 2.0 stereo (which happens quite often in the HDTV industry) results in appropriate reconstruction of the original stereo content with the stereo image intact.

## **Production/Editing**

All downmixing and rendering duties of the broadcaster are handled by the Harris/Neural 5225 surround production appliance.

Existing stereo editing systems are 100% compatible with 5225 downmixed 5.1

Neural spatial compression survives editing and even conversion to analog. After 5.1 content elements are downmixed to 2.0 they may be stored in the stereo archive (server) where they may be imported into the stereo editor and treated like stereo elements. Cutting, pasting and mixing of the downmixed 5.1 elements with stereo 2.0 elements is allowed.

Downmixed 5.1 elements may be loaded, stored or recorded to any cart, sampler or rec/rep. AES or analog I/O is allowed.

The 5225 works well with 5.1 editors. The editor 5.1 outputs may be routed directly into the 5225's downmix inputs. The 2.0 downmix output is routed to the 5225 spatial rendering input as

well as the 2.0 track on the editor. The session may be monitored in real time through the 5225. The session 2.0 file may be dumped onto the server afterwards.

## **Monitoring**

Encoded content may be broadcast, stored, loaded, sampled, recorded or distributed through existing 2.0 infrastructures. It may be rendered to 5.1 at any point in the broadcast chain for production QC or broadcast “confidence monitoring”.

5.1/2.0 mixes may be tested for 5.1 compatibility at any time using the 5225's 2.0 to 5.1 rendering facilities.

Content may be tested anywhere from post capture to post receiver using the 5225 rendering facilities.

Session work may be monitored in realtime through the 5225 while tracking.

## **Storage, Archiving and the Server**

After capture, downmixed 5.1 content may be stored in the existing server (compressed or not) and treated as any other stereo content, thus avoiding the horrors of re-tooling the server.

At this time, three automation system claims the ability to store a discreet 5.1 channel .wav format. No known automation system can currently store, edit or reproduce bitstream encoded surround. If a radio station decides they want to start archiving their library to a linear 5.1 channel .wav format, it must purchase enough storage to support the 31 Mbit/ minute rate that linear 5.1 storage requires. This would typically result in about 1 terabyte of storage needed per/ radio station. This gets *very* expensive with RAID redundancy and regular tape backups. Although some corners could be cut for production rooms, linear 5.1 storage is just one place within the station infrastructure where cost and complexity could be impacted.

Assuming the audio server has enough horsepower to stream multiple, multi-channel audio files simultaneously to enable cross-fades, new sound cards would be needed if the audio console only accepted the standard multichannel AES/EBU format. Assuming the studio is entirely IP based, the station may be able to skip this step and stream the audio over TCP/IP directly to the console. Unfortunately, very few broadcasters either have consoles that can accept a multi-channel AES/EBU or TCP/IP streams. (Thanks to Dave Casey for info)

## **Broadcast Processing**

Mono, stereo or 5.1 content of unknown pedigree can cause trouble with broadcast codecs if not properly treated. Image packing and noise reduction are tantamount to maximizing the performance of today's perceptual coders.

Level and spectral (loudness and signature sound) processing may be performed commonly on both 2.0 and downmixed 5.1 content as both are represented in 2-D with two channels.

The stereo and interspersed 5.1 content are 100% compatible with the Harris/Neural Neustar and UltraLink codec conditioners. Placing the Neustar between the content “play-out” and the IBOC exciter results in HDC compatible content that is controlled, consistent and “renderable” to a 2.0 or 5.1 spatial environment.

## **Benefits**

The methodology demonstrates a high degree of immunity to lossy compression, D/A and A/D conversion. It may be stored or broadcast in either digital or analog.

To broadcasters, this level of backward compatibility is highly attractive as most existing infrastructures, including production and storage, are 2.0. This greatly simplifies the inevitable integration of 2.0 and 5.1 content on both the broadcaster and consumer sides.

The use of intensity/coherence watermarking allows the spatial codec to dedicate significantly higher bits/sec to the fine structures. At ultra low data rates (48kB/s) this would increase the number of bits available for the fine structure by 33% as there is no side data to transport. There is no argument that the AAC codec and its derivatives “sound better” at 64kB/s vs. the 48kB/s or 48kB/s vs. the 32kB/s that would be allowed if the (16kB/s of) side information approach is used. In the realm of lossy compression for broadcast 16kB/s is a lot of bits to use for data other than the fine structure of the audio.

## **Automotive adoption**

Automotive engineers are adopting a consumer version of the Neural’s 2-D downmix/SEE rendering technology as the automotive audio infrastructure is also 2.0. SEE drastically reduces the cost of implementing 5.1, thus driving (pun intended) the ubiquity of the automotive 5.1 system. Any 5225/Neustar treated content that is introduced to the SEE based automotive audio infrastructure will be rendered in “5.1”. Original sourced 5.1 will be rendered as the original source 5.1 and stereo will be rendered in a 5.1 spatial environment.

## **Home Theater adoption**

Receiver manufacturers are adopting SEE rendering technology as a stable and “correct” method of rendering (upmixing). The concept of replacing the “phantom” sources of legacy stereo with “hard” sources (loudspeakers) is a sensible way of improving image stability (discreteness) without adding “special effects”. This allows the consumer to leverage their existing 5.1 loudspeaker array to improve the reproduction of their stereo library without taking artistic license. Apparently this application has significant legs.

## **Conclusion**

Wide spread acceptance of 5.1 radio broadcasting isn't as far away as once thought. 2.0 and 5.1 content interoperability doesn't have to be scary, expensive or dangerous. Modern digital signal processing can modernize the "work horse" content distribution infrastructure every bit as effectively as replace it.

The key is planned, staged transitioning from where the broadcaster is now to where the broadcaster wants to be at a rate of adoption that makes tactical and fiscal sense. The rate of transition should parallel the efforts of the content providers (record labels and artists) as well as manufacturers of home and automotive audio systems.