

Alternatives to 5.1 for Multichannel Music

*Originally commissioned for John Sunier's Audiophile Audition and published in October 2003, this short article is Richard Elen's answer to the resolution:
"Resolved: That the speaker-based 5.1-Channel ITU Standard is fine for movie surround sound but inappropriate for SSfM [Surround Sound for Music].
A better alternative would be _____"*

It's not hard to see why a great many people are of the opinion that while 5.1 may be fine for movie theaters, it is not the best solution for musicsurround. It should not be forgotten that 5.1 was developed to deal with certain problems that beset analog movie theaters - and these problems simply are not present in a modern digital distribution chain such as is available with technologies like DVD-Audio. The Center Front was required in movie theaters because front speakers too far apart led to a hole in the middle and loss of movie dialog. The LFE was required to carry low frequency EFFECTS - such as T.Rex footfalls and asteroids crashing into the Earth - and these do not occur in music: not even in "heavy rock" (sorry, I couldn't resist). The effects channel was necessary to avoid intermodulation effects that would have occurred if sub-bass had been added to other channels. It is arguably beneficial to use the LFE in a digital distribution environment - for sub-bass movie effects only - to avoid problems with headroom, but as every channel in a DVD-Audio disc, for example, is capable of delivering the entire audible range, there should be no need for it in normal musical applications.

So we can say that the LFE is not required for music and could instead be used for something else more interesting - and some record companies such as Telarc, Chesky, MDG and DivoX, use it for height information. We have seen that the CF is, also, not particularly necessary. Indeed many music engineers, brought up on a virtual front stage delivered by two speakers at sixty degrees, prefer the "virtual center" that this configuration provides over stuffing a signal up the center front so that it sticks out like a sorethumb - because the trouble with simply panning sounds to a CF channel is that they are no longer integrated with the front stage. A third SPEAKER - not a third CHANNEL - at center front can, however, be used to decode a stereo front stage more accurately, and in a completely integrated fashion, using a technology such as Trifield - either in the studio or in the playback system.

But all this is tap-dancing around the fundamental problem. Certainly, it must be remembered that whatever better ideas we may think we have, we are stuck with 5.1 and its descendants for the foreseeable future, and we need to learn to live with it and remain compatible with it until it goes away, which may be some time. The real problem is that we have become used to something much more insidious than mere 5.1: we have come to believe that 'one-to-one mapping' is all there is. 5.1 is an example of this, but it is only a special case.

Since the days of quad, we have been led to believe that the ultimate in surround sound involved capturing sound with n microphones, transferring their signals via n channels, and replaying them with n loudspeakers placed in something like the directions the mics were pointed in. What the world was waiting for, according to this idea, was the availability of distribution media with n high-resolution channels to do the job properly. This, I am afraid, is complete rubbish.

What we really ought to be doing is to capture and/or mix the sound in the most artistically and technologically satisfactory way possible. This might involve one mic or many, depending on the project and the intent of the production team. The resulting signals should be transmitted in the most effective and efficient way possible: this does not mean one channel per microphone, it means representing a multi-dimensional soundfield in the most efficient way. And finally at the other end, these efficient distribution channels need to be decoded to drive an appropriate number of speakers, the feed for each speaker being derived as a function of its location in the listening environment.

The most obvious and best-known method of doing this is Ambisonics, though it is no doubt not the only method of achieving this goal. However, using simple first-order Ambisonics as an example, we can see how this concept works in practise. To begin with, envisage the three-dimensional soundfield being captured by a suitable microphone array or multitrack recording with multi-channel panpots or a combination of the two.

The resulting signals are now encoded into a series of sum and difference signals, essentially similar to a three-dimensional development of the Blumlein X-Y technique: one channel carries the sum of all the dimensional signals, ie Left+Right+Front+Back+Up+Down, while other channels carry the differences: L-R, F-B, U-D. This signal set is called B-Format, and is an extremely efficient way of representing the surround-sound we hear - note that it only uses four channels to carry everything needed to recreate a full three-dimensional (with height) soundfield.

At the receiving end, the B-Format signals are decoded to suit a multi-speaker array that is practical for the listening environment. More speakers may be better; height might be nice; but essentially given some basic ground-rules, with a reasonable number of speakers in reasonable places, you can get excellent results - in an ordinary living room, or a home theater, or even in a movie theater or auditorium. The incoming B-Format signals are decoded for the speaker array, and the speakers can be more or less where you like, within reason.

This is all very well, but what about the 5.1 compatibility I mentioned earlier? This is taken care of, thanks to the work behind a paper presented by R&D staff from Meridian Audio Ltd at the recent AES Banff conference. Built into MLP, the lossless packing system at the heart of DVD-Audio, is the ability to code hierarchical surround information (such as Ambisonic information derived from B-Format) and flag it in the metadata, such that the result can be played direct into a 5.1 loudspeaker system with no special equipment whatsoever and be completely compatible. However, with a suitable decoder, switched in automatically by a flag in the datastream if desired, the same information can be decoded for the listener's specific speaker array, using Ambisonic technology. The transmitted signal delivers a hierarchy of information, that can be decoded according to the equipment available - or simply used as standard 5.1 speaker feeds. The best of both worlds.

A simple version of this concept exists in the Trifield technology I mentioned earlier. Trifield, based on Ambisonic research by Dr Geoff Barton and Michael Gerzon, essentially decodes a 2-channel stereo mix for three loudspeakers. As in the more sophisticated system described above, the original multichannel/multimic or whatever source, mixed to stereo, is carried via two channels, but decoded to three loudspeakers that work together to generate a fully integrated stereo sound stage that is superior in imaging and image stability than either two-channel, two-speaker stereo or three-channel "panpotted mono". But Trifield is simply the beginning. Because with hierarchical coding, you can have your 5.1 cake and eat it too - right now.

