2024-01-31 multi-channel audio part 2

Last time, we left off at the fact that modern films are distributed with their audio in multiple formats. Most of the time, there is a stereo version of the audio, and a multi-channel version of the audio that is perhaps 5.1 or 7.1 and compressed using one of several codecs that were designed within the film industry for this purpose.

But that was all about film, in physical form. In the modern world, films go out to theaters in the form of Digital Cinema Packages, a somewhat elaborate format that basically comes down to an encrypted motion JPEG 2000 stream with PCM audio. There are a lot of details there that I don't know very well and I don't want to get hung up on anyway, because I want to talk about the consumer experience.

As a consumer, there are a lot of ways you get movies. If you are a weirdo, you might buy a Blu-Ray disc. Optical discs are a nice case, because they tend to conform to a specification that allows relatively few options (so that players are reasonable to implement). Blu-Ray are allowed to encode their audio as linear PCM [1], Dolby Digital, Dolby TrueHD, DTS, DTS-HD, or DRA.

DRA is a common standard in the Chinese market but not in the US (that's where I live), so I'll ignore it. That still leaves three basic families of codecs, each of which have some variations. One of the interesting things about the Blu-Ray specification is that PCM audio can incorporate up to eight channels. The Blu-Ray spec allows up to 27,648 Kbps of audio, so it's actually quite feasible to do uncompressed, 24-bit, 96 kHz, 7.1 audio on a Blu-Ray disc. This is an unusual capability in a consumer standard and makes the terribly named Blu-Ray High Fidelity Pure Audio standard for Blu-Ray audio discs make more sense. Stick a pin in that, though, because you're going to have a tough time actually playing uncompressed 7.1.

On the other hand, you might use a streaming service. There's about a million of those and half of them have inane names ending in Plus, so I'm going to simplify by pretending that we're back in 2012 and Netflix is all that really matters. We can infer from Netflix help articles that Netflix delivers audio as AAC or Dolby Digital.

Or, consider the case of video files that you obtained by legal means. I looked at a few of the movies on my NAS to take a rough sampling. Most older films, and some newer ones, have stereo AAC audio. Some have what VLC describes as A52 aka AC3. A/52 is an ATSC standard that is equivalent to AC3, and AC-3 (hyphen inconsistent) is sort of the older name of Dolby Digital or the name of the underlying transport stream format, depending on how you squint at it. Less common, in my hodgepodge sample, is DTS, but I can find a few.

VLC typically describes the DTS and Dolby Digital as 3F2M/LFE, which is a somewhat eccentric (and I think specific to VLC) notation for 5.1 surround. An interesting detail is that VLC differentiates 3F2M/LFE and 3F2R/LFE, both 5.1, but with the two "surround" channels assigned to either side or rear positions. While 5.1 configurations with the surround channels to the side seem to be more standard, you could potentially put the two surround channels to the rear. Some formats have channel mapping metadata that can differentiate the two.

Because there is no rest for the weary, there is some inconsistency between "5.1 side" and "5.1 rear" in different standards and formats. At the end of the day, most applications don't really differentiate. I tend to consider surround channels on the side to be "correct," in that movie theaters are configured that way and thus it's ostensibly the design target for films. One of few true specifications I could find for general use, rather than design standards specific to theaters like THX, is ITU-R BS 775. It states that the surround channels of a 5.1 configuration should be mostly to the side, but slightly behind the listener.

That digression aside, it's unsurprising that a video file could contain a multi-channel stream. Most video containers today can support basically arbitrary numbers of streams, and you could put uncompressed multichannel audio into such a container if you wanted. And yet, multi-channel audio in films almost always comes in the form of a Dolby Digital or DTS stream. Why is that? Well, in part, because of tradition: they used to be the formats used by theaters, although digital cinema has somewhat changed that situation and the consumer versions have usually been a little different in the details. But the point stands, films are usually mastered in Dolby or DTS, so the "home video" release goes out with Dolby or DTS.

Another reason, though, is the problem of interconnections.

Let's talk a bit about interconnections. In a previous era of consumer audio, the age of "hi-fi," component systems dominated residential living rooms. In a component system, you had various audio sources that connected to a device that came to be known as a "receiver" since it typically had an FM/AM radio receiver integrated. It is perhaps more accurate to refer to it as an amplifier since that's the main role it serves in most modern systems, but there's also an increasing tendency to think of their input selection and DSP features as part of a preamp. The device itself is sometimes referred to as a preamp, in audiophile circles, when component amplifiers are used to drive the actual speakers. You can see that in these conventional component systems you need to move audio signals between devices. This kind of set up, though, is not common in households with fewer than four bathrooms and one swimming pool.

Most consumers today seem to have a television and, hopefully, some sort of audio device like a soundbar. Sometimes there are no audio interconnections at all! Often the only audio interconnection is from the TV to the soundbar via HDMI. Sometimes it's wireless! So audio interconnects as a topic can feel a touch antiquated today, but these interconnects still matter a lot in practice. First, they are often either the same as something used in industry or similar to something used in industry. Second, despite the increasing prevalence of 5.1 and 7.1 soundbar systems with wireless satellites, the kind of people with a large Blu-Ray collection are still likely to have a component home theater system. Third, legacy audio interconnects don't die that quickly, because a lot of people have an older video game console or something that they want to work with their new TV and soundbar, so manufacturers tend to throw in one or two audio interconnects even if they don't expect most consumers to use them.

So let's think about how to transport multi-channel audio. An ancient tradition in consumer audio says that stereo audio will be sent between components on two sets of two-conductor cables terminated by RCA connectors. The RCA connector dates back to to the Radio Corporation of America and, apparently, at least 1937. It remains in widespread service today. There are a surprising number of variations in this interconnect, in practice.

For one, the audio cables may be coaxial or just zipped up in a common jacket. Coaxial audio cables are a bit more expensive and a lot less flexible but admit less noise. There is a lot of confusion in this area because a particular digital transport we'll talk about later specified coaxial cables terminated in RCA connectors, but then is frequently used with non-coaxial cables terminated in RCA connectors, and for reasonable lengths usually still works fine. This has lead to a lot of consumer confusion and people thinking that any cable with RCA connectors is coaxial, when in fact, most of them are not. Virtually all of them are not. Unless you

specifically paid more money to get a coaxial one, it's not, and even then sometimes it's not, because Amazon is a hotbed of scams.

Second, though these connections are routinely described as "line level" as if that means something, there is remarkably little standardization of the actual signaling. There are various conventions like 1.7v peak-to-peak and 2v peak-to-peak and about 1v peak-to-peak, and few consumer manufacturers bother to tell you which convention they have followed. There are also a surprising number of ways of expressing signaling levels, involving different measurement bases (peak vs RMS) and units (dBv vs dBu), making it a little difficult to interpret specifications when they are provided. This whole mess is just one of the reasons you find yourself having to make volume adjustments for different sources, or having to tune input levels on receivers with that option [2].

But that's all sort of a tangent, the point here is multi-channel audio. You could, conceptually, move 5.1 over six RCA cables, or 7.1 over eight RCA cables. Home theater receivers used to give you this option, but much like analog HDTV connections, it has largely disappeared.

There is one other analog option: remember Pro Logic, from the film soundtracks? that matrixed five channels into the analog stereo? Some analog formats like VHS and LaserDisc often had a Pro Logic soundtrack that could be "decoded" (really dematrixed) by a receiver with that capability, which used to be common. In this case you can transport multi-channel audio over your normal two RCA cables. The matrixing technique was always sort of cheating, though, and produces inferior results to actual multichannel interconnects. It's no longer common either.

Much like video, audio interconnects today have gone digital. Consumer digital audio really took flight with the elegantly named Sony/Philips Digital Interface, or S/PDIF. S/PDIF specifies a digital format that is extremely similar to, but not quite the same as, a professional digital interconnect called AES3. AES3 is typically carried on a three-conductor (balanced) cable with XLR connectors, though, which are too big an expensive for consumer equipment. In one of the weirder decisions in the history of consumer electronics, one that I can only imagine came out of an intractable political fight, S/PDIF specified two completely different physical transports: one electrical, and one optical.

The electrical format should be transmitted over a coaxial cable with RCA connectors. In practice it is often used over non-coaxial cables with RCA connectors, which will usually work fine if the length is short and nothing nearby is too electrically noisy. S/PDIF over non-coaxial cables is "fine" in the same way that HDMI cables longer than you are tall are "fine." If it doesn't work reliably, try a more expensive cable and you'll probably be good.

The optical format is used with cheap plastic optical cables terminated in a square connector called Toslink, originally for Toshiba Link, after the manufacturer that gave us the optical variant. Toslink is one of those great disappointments in consumer products. Despite the theoretical advantages of an optical interconnect, the extremely cheap cables used with Toslink mean it's mostly just worse than the electrical transport, especially when it comes to range [3].

But the oddity of S/PDIF's sibling formats isn't the interesting thing here. Let's talk about the actual S/PDIF bitstream, the very-AES3-like format the audio actually needs to get through.

S/PDIF was basically designed for CDs, and so it comfortably carries CD audio: two channels of 16 bit samples at 44.1kHz. In fact, it can comfortably go further, carrying 20 (or with the right equipment even 24) bit samples at the 48 kHz sampling rate more common of digital audio other than CDs. That's for two channels, though. Make the leap to six channels for 5.1 and you are well beyond the capabilities of an S/PDIF transceiver.

You see where this is going? compression.

See, the problems that Dolby Digital and DTS solved, of fitting multichannel audio onto the limited space of a 35mm film print, also very much exist in the world of S/PDIF. CDs brought us uncompressed digital audio remarkably early on, but also set sort of a constraint on the bitrate of digital audio streams that ensured the opposite in the world of multi-channel theatrical sound. It sort of makes sense, anyway. DTS soundtracks came on CDs!

Of course even S/PDIF is looking rather long in the tooth today. I don't think I use it at all any more, which is not something I expected to be saying this soon. Today, though, all of my audio sources and sinks are either analog or have HDMI. HDMI is the de facto norm for consumer digital audio today.

HDMI is a complex thing when it comes to audio or, really, just about anything. Details like eARC and the specific HDMI version have all kinds of impacts on what kind of audio can be carried, and the same is true for video as well. I am going to spare a lengthy diversion into the many variants of HDMI, which seem almost as numerous as those of USB, and talk about HDMI 2.1.

Unsurprisingly, considering the numerous extra conductors and newer line coding, HDMI offers a lot more bandwidth for audio than S/PDIF. In fact, you can transport 8 channels of uncompressed 24-bit PCM at 192kHz. That's about 37 Mbps, which is not that fast for a data transport but sure is pretty fast for an audio cable. Considering the bandwidth requirements for 4K video at 120Hz, though, it's only a minor ask. With HDMI, compression of audio is no longer necessary.

But we still usually do it.

Why? Well, basically everything can handle Dolby Digital or DTS, and so films are mostly mastered to Dolby Digital or DTS, and so we mostly use Dolby Digital or DTS. That's just the way of things.

One of the interesting implications of this whole thing is that audio stacks have to deal with multiple formats and figure out which format is in use. That's not really new, with Dolby Pro Logic you either had to turn it on/off with a switch or the receiver had to try to infer whether or not Pro Logic had been used to matrix a multichannel soundtrack to stereo. For S/PDIF, IEC 61937 standardizes a format that can be used to encapsulate a compressed audio stream with sufficient metadata to determine the type of compression. HDMI adopts the same standard to identify compressed audio streams (and, in general, HDMI audio is pretty much in the same bitstream format as good old S/PDIF, but you can have a lot more of it).

In practice, there are a lot of headaches around this format switching. For one, home theater receivers have to switch between decoding modes. They mostly do this transparently and without any fuss, but I've owned a couple that had occasional issues with losing track of which format was in use, leading to dropouts. Maybe related to signal dropouts but my current receiver has the same problem with internal sources, so it seems more like a software bug of some sort.

It's a lot more complicated when you get out of dedicated home theater devices, though. Consider the audio stack of a general-purpose operating system. First, PCs rarely have S/PDIF outputs, so we are virtually always talking about HDMI. For a surprisingly long time, common video cards had no support for audio over HDMI. This is fortunately a problem of the past, but unfortunately ubiquitous audio over HDMI means that your graphics drivers are now involved in the transport of audio, and graphics drivers are notoriously bad at reliably producing video, much less dealing with audio as a side business. I shudder to think of the hours of my life I have lost dealing with defects of AMD's DTS support.

Things are weird on the host software side, though. The operating system does not normally

handle sound in formats even resembling Dolby Digital or DTS. So, when you play a video file with audio encoded in one of those formats, a "passthrough" feature is typically used to deliver the compressed stream directly to the audio (often actually video) device, without normal operating system intervention. We are reaching the point where this mostly just works but you will still notice some symptoms of the underlying complexity.

On Linux, it's possible to get this working, but in part because of licensing issues I don't think any distros will do it right out of the box. My knowledge may be out of date as I haven't tried for some time, but I am still seeing Kodi forum threads about bash scripts to bypass PulseAudio, so things seem mostly unchanged.

There are other frustrations, as well. For one, the whole architecture of multichannel audio interconnection is based around sinks detecting the mode used by the source. That means that your home theater receiver should figure out what your video player is doing, but your video player has no idea what your home theater receiver is doing. This manifests in maddening ways. Consider, for example, the number of blog posts I ran across (while searching for something else!) about how to make Netflix less quiet by disabling surround sound.

If Netflix has 5.1 audio they deliver it; they don't know what your speaker setup is. But what if you don't have 5.1 speakers? In principal you could downmix the 5.1 back to stereo, and a lot of home theater receivers have DSP modes that do this (and in general downmix 5.1 or 7.1 to whatever speaker channels are active, good for people with less common setups like my own 3.1). But you'd have to turn that on, which means having a receiver or soundbar or whatever that is capable, understanding the issue, and knowing how to enable that mode. That is way more than your average Netflix watcher wants to think about any of this. In practice, setting the Netflix player to only ever provide stereo audio is an easier fix.

The use of compressed multichannel formats that are decoded in the receiver rather than the computer playing back introduces other problems as well, like source equalization. If you have a computer connected to a home theater receiver (which is a ridiculous thing to do and yet here I am), you have two completely parallel audio stacks: "normal" audio that passes through the OS sound server and goes to the receiver as PCM, and "surround sound" that bypasses the OS sound server and goes to the receiver as Dolby Digital or DTS. It is very easy to have differences in levels, adjustments, latency, etc. between these two paths. The level problem here is just one of the several factors in the perennial "Plex is too quiet" forum threads [4].

Finally, let's talk about what may be, to some readers, the elephant in the room. I keep talking about Dolby Digital and DTS, but both are 5.1 formats, and 5.1 is going out of fashion in the movie world. Sure, there's Dolby Digital Plus which is 7.1, but it's so similar to the non-plus variant that there isn't much use in addressing them separately. Insert the "Plus" after Dolby Digital in the proceeding paragraphs if it makes you feel better.

But there are two significantly different formats appearing on more and more film releases, especially in the relatively space-unconstrained Blu-Ray versions: lossless surround sound and object-based surround sound.

First, lossless is basically what it sounds like. Dolby TrueHD and DTS-HD are both formats that present 7.1 surround with only lossless compression, at the cost of a higher bitrate than older media and interconnects support. HDMI can easily handle these, and if you have a fairly new setup of a Blu-Ray player and recent home theater receiver connected by HDMI you should be able to enjoy a lossless digital soundtrack on films that were released with one. That's sort of the end of that topic, it's nothing that revolutionary.

But what about object-based surround sound? I'm using that somewhat lengthy term to try to avoid singling out one commercial product, but, well, there's basically one commercial

product: Dolby Atmos. Atmos is heralded as a revolution in surround sound in a way that makes it sort of hard to know what it actually is. Here's the basic idea: instead of mastering a soundtrack by mixing audio sources into channels, you master a soundtrack by specifying the physical location (in cartesian coordinates) of each sound source.

When the audio is played back, an Atmos decoder then mixes the audio into channels on the fly, using whatever channels are available. Atmos allows the same soundtrack to be used by theaters with a variety of different speaker configurations, and as a result, makes it practical for theaters to expand into much higher channel counts.

Theaters aren't nearly as important a part of the film industry as they used to be, though, and unsurprisingly Atmos is heavily advertised for consumer equipment as well. How exactly does that work?

Atmos is conveyed on consumer equipment as 7.1 Dolby Digital Plus or Dolby TrueHD with extra metadata.

If you know anything about HDR video, also known as SDR video with extra metadata, you will find this unsurprising. But some might be confused. The thing is, the vast majority of consumers don't have Atmos equipment, and with lossless compression soundtracks are starting to get very large so including two complete copies isn't very appealing. The consumer encoding of Atmos was selected to have direct backward compatibility to 7.1 systems, allowing normal playback on pre-Atmos equipment.

For Atmos-capable equipment, an extra PCM-like subchannel (at a reduced bitrate compared to the audio channels) is used to describe the 3D position of specific sound sources. Consumer Atmos decoders cannot support as many objects as the theatrical version, so part of the process of mastering an Atmos film for home release is clustering nearby objects into groups that are then treated as a single object by the consumer Atmos decoder. One way to think about this is that Atmos is downmixed to 7.1, and in the process a metadata stream is created that can be used to upmix back to Atmos mostly correctly. If it sounds kind of like matrix encoding it kind of is, in effect, which is perhaps part of why Dolby's marketing materials are so insistent that it is not matrix encoding. To be fair it is a completely different implementation, but has a similar effect of reducing the channel separation compared to the original source.

Also I don't think Atmos has really taken off in home setups? I might just be out of date here, I think half the soundbars on the market today claim Atmos support and amazing feats with their five channels two of which are pointed up. I'm just pretty skeptical of the whole "we have made fewer, smaller speakers behave as if they were more, bigger speakers" school of audio products. Sorry Dr. Bose, there's just no replacement for displacement.

[1] The term Linear PCM or LPCM is used to clarify that no companding has been performed. This is useful because PCM originated for the telephone network, which uses companding as standard. LPCM clarifies that neither -law companding nor A-law companding has been performed. I will mostly just use PCM because I'm talking about movies and stuff, where companding digital audio is rare.

[2] There is also the matter of magnetic sources like turntables and microphones that produce much lower output levels than a typical "line level." Ideally you need a preamplifier with adjustable gain for these, although in the case of turntables there are generally accepted gain levels for the two common types of cartridges. A lot of preamplifiers either let you choose from those two or give you no control at all. Traditionally a receiver would have a built-in preamplifier to bring up the level of the signal on the turntable inputs, but a lot of newer receivers have left this out to save money, which leads to hipsters with vinyl collections having to really crank the volume.

[3] I don't feel like I should have to say this, but in the world of audio, I probably do: if it works, it doesn't matter! The problem with optical is that it develops reliability problems over shorter lengths than the electrical format. If you aren't getting missing samples (dropouts) in the audio, though, it's working fine and changing around cables isn't going to get you anything. In practice the length limitations on optical don't tend to matter very much anyway, since the average distance between two pieces of a component home theater system is, what, ten inches?

[4] Among the myriad other factors here is the more difficult problem that movies mix most of the dialog into the center channel while most viewers don't have a center channel. That means you need to remix the center channel into left and right to recover dialog. So-called professionals mastering Blu-Ray releases don't always get this right, and you're in even more trouble if you're having to do it yourself.