Please email any questions or comments to tom.hays@technicolor.com

Enjoy!

Hi, and welcome. This session is Dynamic Range and Mix Levels: Where we're at in 2011.

For starters, where we're at has evolved from where we were a few years ago. There are more and more games these days that utilize dynamic space to make rich, engrossing soundscapes, that use quiet moments to suck you in, and loud moments to shock you.

Of course, that's not right for every game – there are many types of game out there, and not all of them should have that sort of up/down, push/pull. But even with them, we're seeing fewer games that slam the DACs and generally ignore output gain.

So in today's brief session, we're going to look at some background, talk about a few of the issues that remain when it comes to dynamics, compare notes with other media, and look at some ways that might work if we want to move towards some standard practices.

The earliest video games often had essentially one bit of dynamic range: on or off, 1 or zero.

Through the '80s and early '90s, as games moved more into home consoles and PCs, their sound subsystems became more sophisticated. Sample playback engines had 4-bit resolution at first, for 16 glorious steps of dynamic range, until around 1992 with the advent of the Sound Blaster's 8 bits of awesome.

Fight the Noise!

Any of you who have worked with 8-bit sound know that the only way to get decent fidelity is to get rid of all the dynamics, because quiet material is drowned in quantization noise.

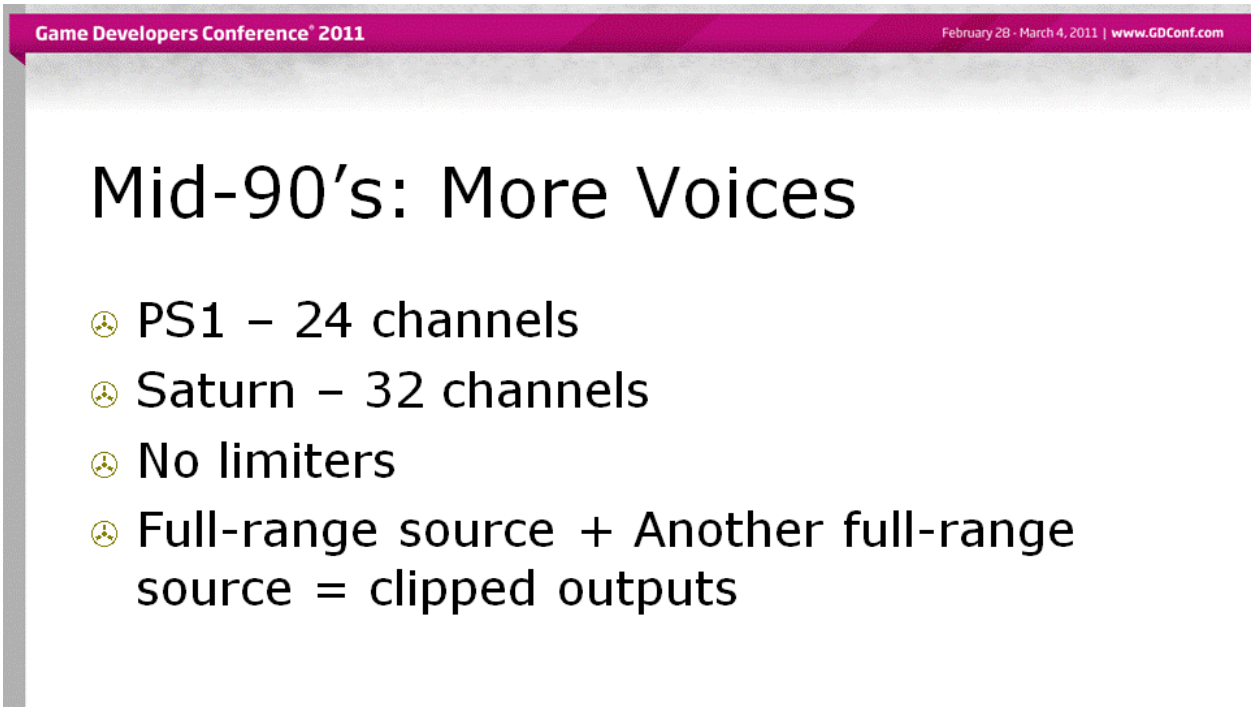
That particular issue went well beyond games back in the 80s and 90s. This was before cheap PCs with multitrack recording freeware, so many of us were recording music on fairly gamey analogue gear, such as multitrack cassette decks. When DAT decks came along, the early models that were commonly used sounded pretty grainy in their quieter ranges, so we learned to hit them hard.

To make a long story short, by the time we started to adjust to 16-bit mixers in game consoles and sound cards, there was a big group of us who had all learned that the only way to get decent sound was to stay in the top few dB of available dynamic range.

This was compounded by what was happening in music production, and goes on to this day:

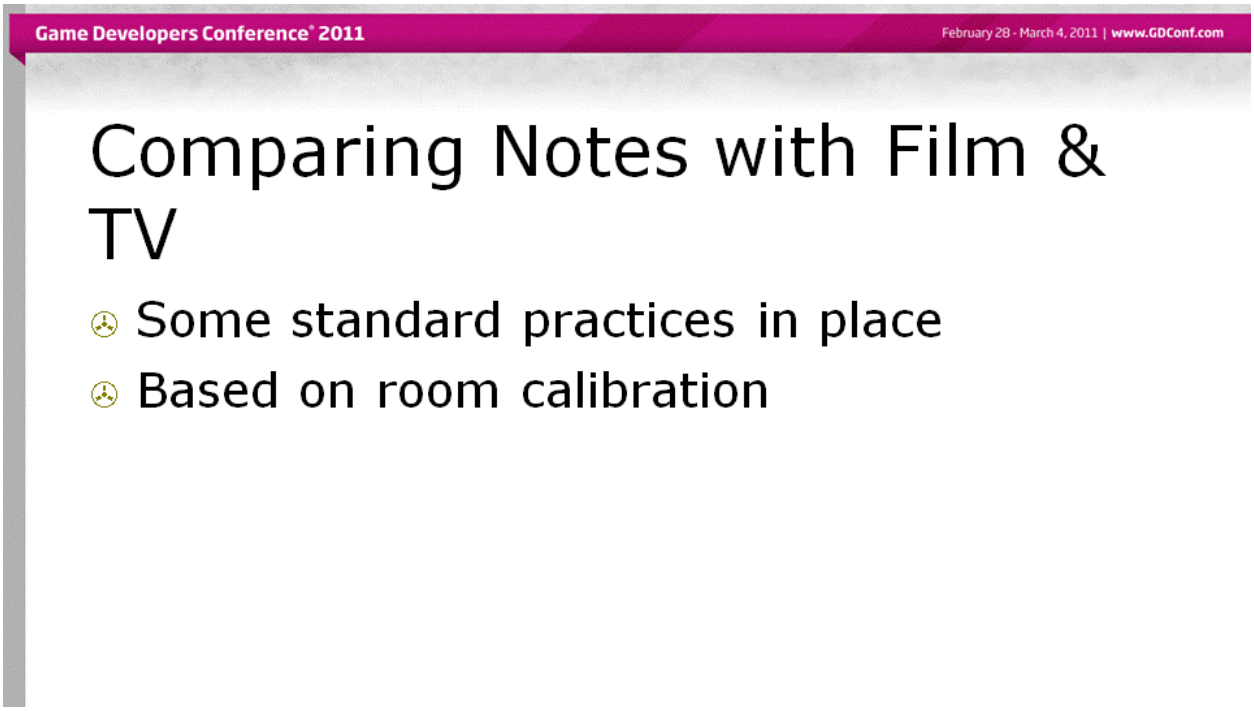
Volume Wars

Volume wars. We saw some of this back then in games as well: to the extent that we could control dynamics in a game, there was always at least some pressure to keep the outputs up for what were seen as competitive reasons.



In the mid-90's, games started to get to where they were able to load and play a much larger number of sounds at once. The Sega Saturn, for example, had 32 voices and the original PlayStation had 24. These had no bus or output limiters. So, if you played multiple full-scale sounds, they'd sometimes slam together and distort the outputs.

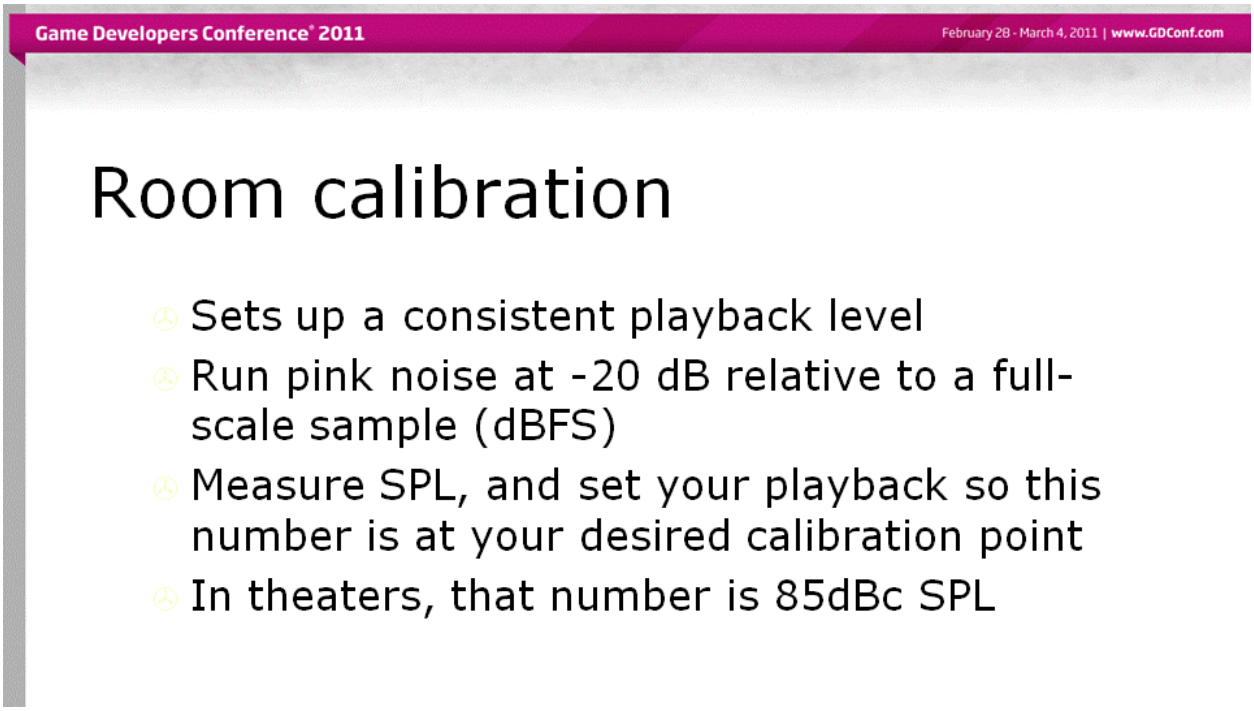
On the Playstation and PS 2, each ADPCM channel had a maximum output of -6dB, something we discovered to our chagrin when using them for premixed cutscene soundtracks. There were similar limits built into the Xbox. In a way, it was like the same kind of training-wheels safety margin that's built into some prosumer mixing gear. This wasn't the intent, necessarily, but in a time when many games were trying to play more than one full-scale sound at the same time with little or no attenuation, this architecture served as a brute-force way to reduce output distortion.



As memory and channel count increased, another thing that happened is that we started to see some similarities between some games and film, not just in cinematics but in gameplay. This led some game sound people to look more closely at practices common to film sound.

Also, as videogames evolved, and it took a little less math to get sounds into them, some people from film and TV started to dabble in games.

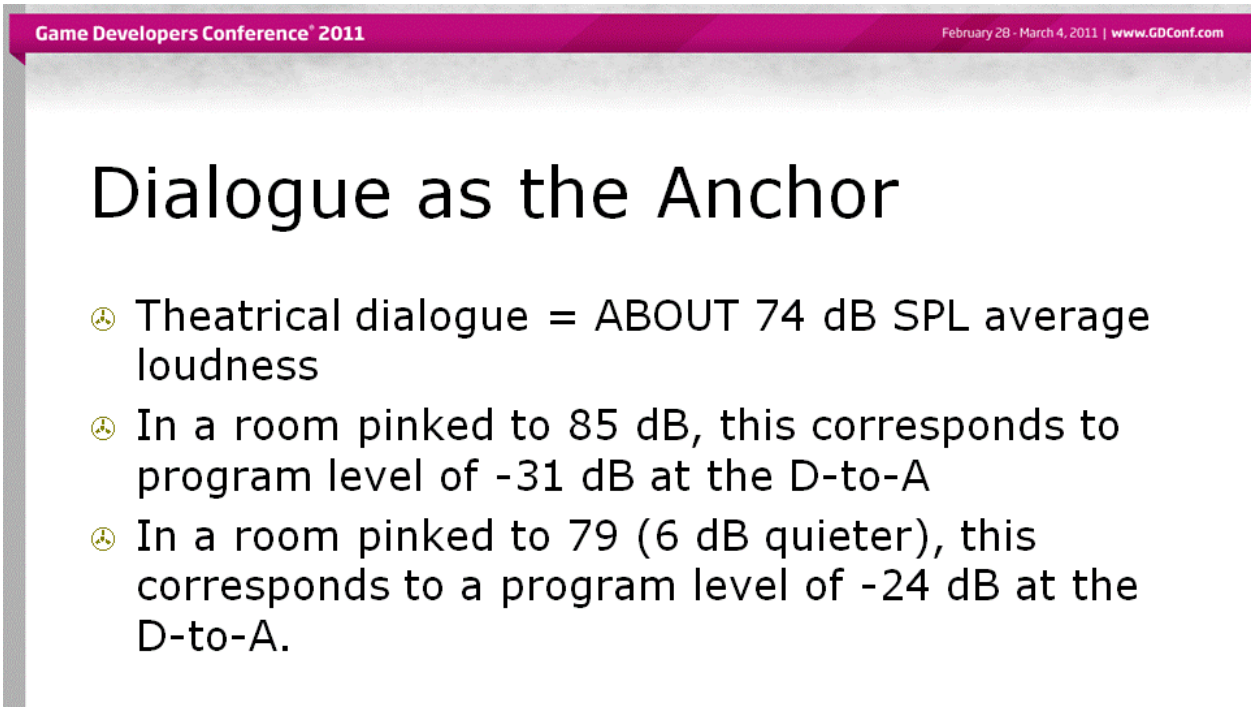
Film and TV are by no means immune from the loudness wars, but there are some standards in place that keep things from getting too far off the reservation. At the core of this is the practice of mixing in rooms with calibrated playback level. Calibrated level in a room's playback system means that pink noise played at a certain level relative to a full-scale sample in the program material will be heard at a set volume. A given amplitude in the waveform at the D-to-As translates to a set SPL in the room. Touching the master playback volume in a film dubbing stage is a big no-no.



A room's calibration level gives you some idea of what headroom you should expect in a mix done there. Once you get used to how different things sound in your room, you are likely to make mixes that coexist with what you worked on in the same room the week before.

This works pretty well in film, and is widespread practice in TV as well. THX's game guys proposed that we adapt film's room calibration standard, back in 2005. The biggest objection many had was that we tend to mix games, as well as TV, in rooms where the calibration level is lower, to create mixes with less dynamics. This is desirable for most home environments, and really most playback settings outside of a movie theater.

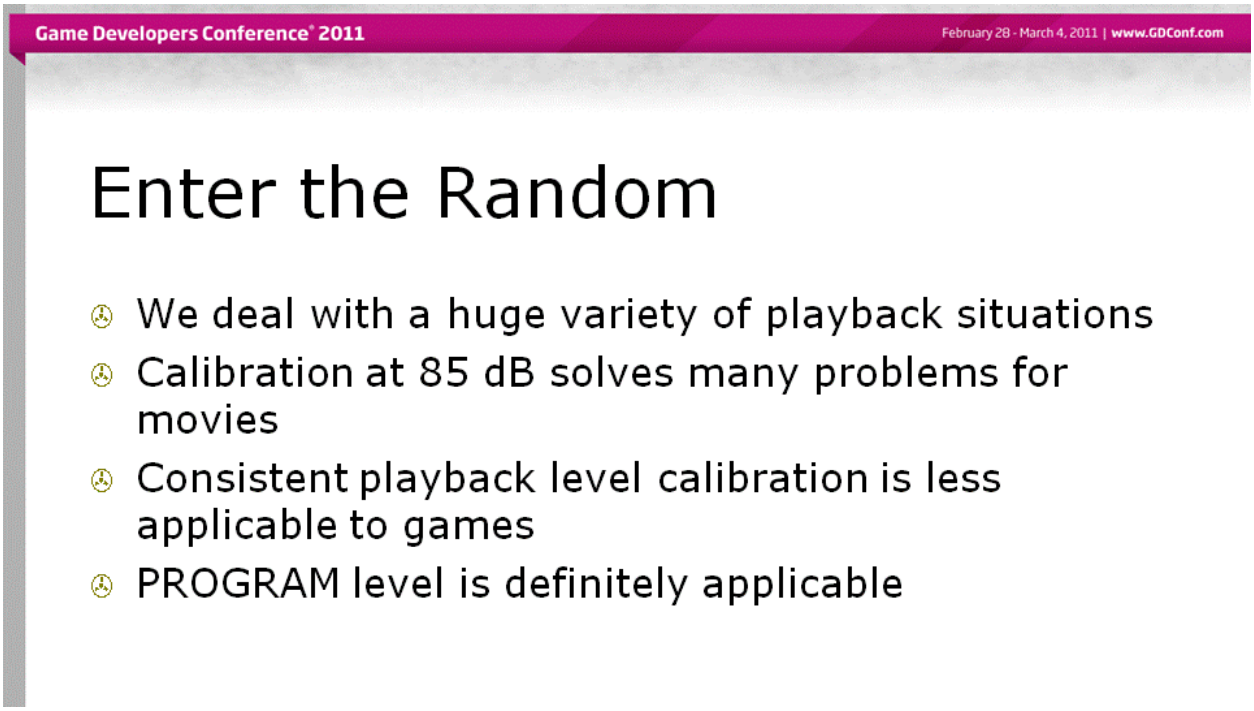
The theatrical standard is made for huge rooms. Theaters are calibrated by running pink noise at a program level of -20 db RMS through each individual main speaker, and adjusting the volume in the room so that a meter reads the sound pressure at 85 dB, c-weighted. That means the loudest peak can theoretically hit 105 dB SPL, which is hella loud no matter where you are.



Much of this is based upon where people generally agree conversational dialogue ought to sit, level-wise. In a big room like a theater, this seems to be somewhere in the ballpark of 74 dB SPL average loudness. If you look at what's recorded in a film mix, the textbook area to find dialogue's average loudness is at -27 to -31, relative to a full-scale sample.

So, in a room where, roughly speaking, -20 in the program material plays at 85 in the air, program material that's at -31 will play in the room at 74. So – again, roughly speaking - that's the relationship between calibration level and dialogue loudness in a theater.

It stands to reason, then, that if we turn the system down, and we want to still hear dialogue at the same level, we need to turn that dialogue up in the program material. This is common practice in rooms set up to mix for smaller environments than the theater. Most TV mix stages are calibrated using pink noise that's 3 to 10 dB quieter than the calibration noise in a theater, with the most common values being 79 and 78. The assumption is that the explosions will be less likely to drown out the voice if the headroom between the dialogue and the loudest possible sound is reduced. You'll typically find that when you measure the program material for high-end TV shows, dialogue measures about -24 dB average loudness. That's 5 or 6 dB hotter than is typical of theatrical program material. This directly reflects the difference in room calibration.

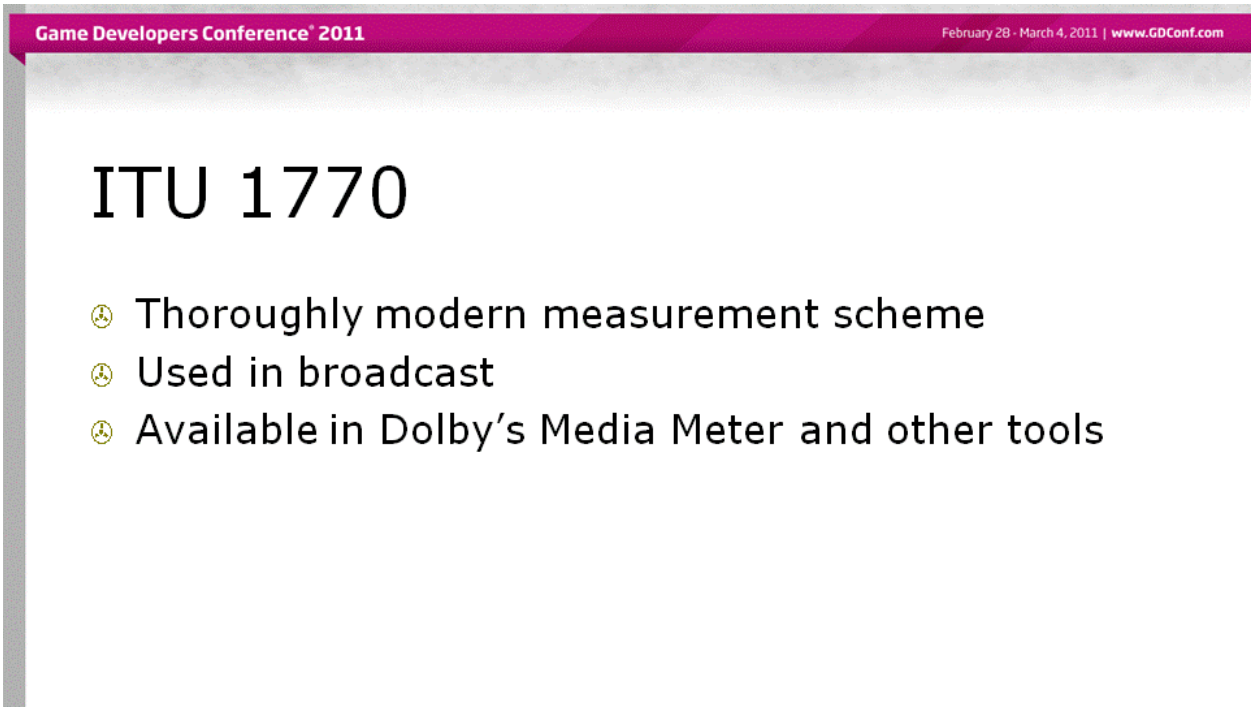


But, what happens when we get out of even these smaller mixing stages, into the offices and other small, non-ideal environments where much of our work winds up getting done? In some cases, that "correct" level for dialogue, and therefore the rest of the mix, sounds oppressively loud, if the room's too small or the speakers are too close.

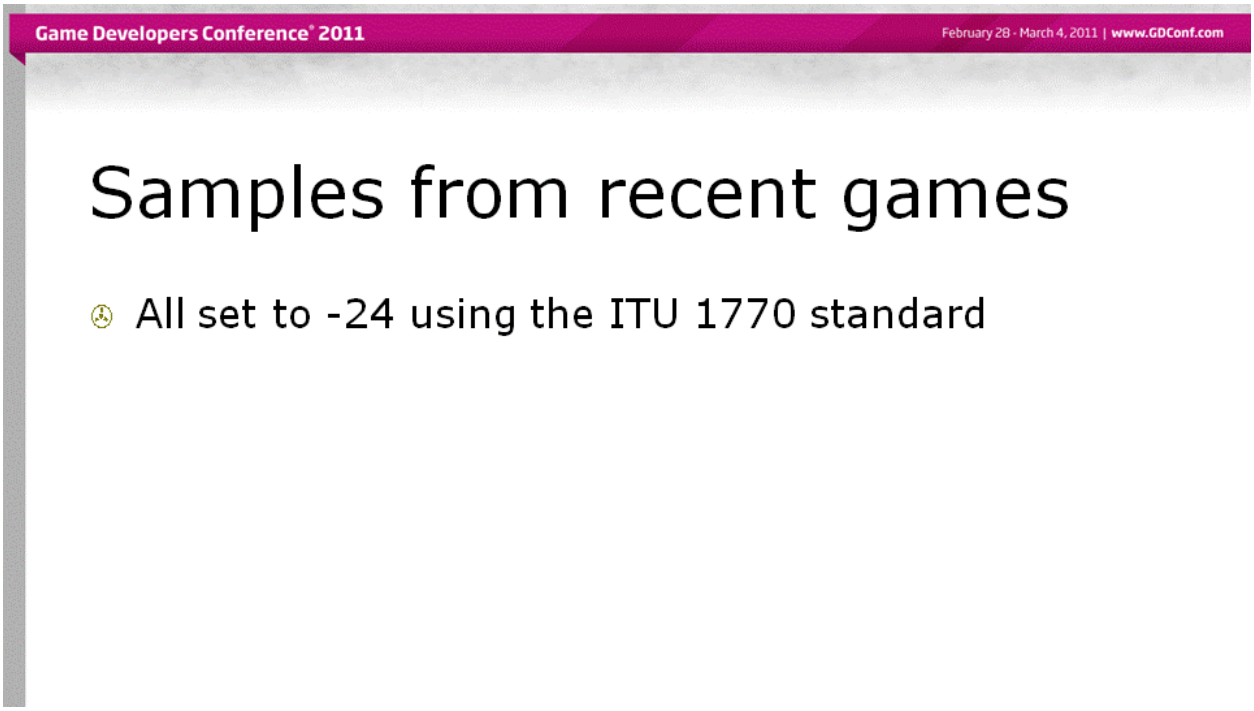
This leaves us with a problem when it comes to setting a consistent metric for dynamics, given the huge variety of settings and playback systems we face.

Room calibration has worked well for movies. It's not perfect - projectionists will sometimes sneak down the playback on excessively-loud movies, especially in smaller theaters, or where the playback systems aren't so great. Part of this is directors who go for overly-macho SPL in their final mixes, but part of this goes back to our problem of variable room size.

My thought after looking at all of this is that any kind of predictable measurement that's repeatable and relevant between a number of different playback and listening situations should be handled on the program side, not the playback side.



This isn't an original thought, so I don't want to take credit. Broadcast standards are on the program side, not the playback side. Charles Deenen from EA proposed going this route on our panel at the last AES convention. As he pointed out, there is a standard that works quite well across a big variety of program material: ITU 1770. This is a very advanced, modern measurement system that goes beyond what Leq and other previous systems did, and incorporates all sorts of things, including what the head does when sounds come from different angles and much more. It's pretty robust.



The next part of this presentation cannot be shared at this time, due to rights issues. Sorry!

Email any questions or comments to tom.hays@technicolor.com