Beyond Coding, 3D and beyond.

Let's take audio out of the 20th century.

James D. Johnston

Chief Scientist

DTS, Inc.

What is the state of the world?

- We have 2.0 and 5.1 delivery
  - Not all channels are always independently coded
    - This means that envelopment and dispersion are lost
    - This will create problems for 3D delivery
  - We have a reasonable, but imperfect, level measurement system
    - But we don't adapt it to the listening situation very well
    - In fact, we tend to assume "worst possible case"
      - This reduces quality elsewhere
  - We don't include material-specific acoustic metadata.
  - Delivery isn't scalable in any easy fashion

New Issues

- We would like one stream to accommodate 2.0, 5.1, 7.1 audio, with independent coding where the bit rate is available
- We want to be able to reduce the channel count easily
- We want to be able to scale the bit rate easily
- We want to be able to virtualize the multichannel for headphones

More new issues

- In the home, we will have smarter room characterization systems
  - We will need to adapt to the actual rendering site equipment
  - We will be able to know when:
    - There is no center speaker
    - The 5.1 is in a line above the TV set
    - The sub is turned off
    - etc

And even more

- We have the ability to create 5.1 and 7.1 mixes, now, with amplitude/delay panning, and with proper dispersive content for enveloping sounds
  - But then we have to code them right
    - Using codecs that would strip or take this information back out, or simplify it, is a non-starter
    - “Coupling channels” and 5 2 2 5 systems can not, by their very nature, get this right
- We would like to know more about the playback environment
- We would like to know (at the playback end) more about the intended acoustic environment

Now what?

- It’s time to remove the extreme loudness processing from the head end, and put it into the playback unit.
  - This means that in a car, you can “squish it flat”
  - In a quite home, you can have a proper dynamic range
  - You can do a dramatic “scene” and not lose the intended sensation
- It’s important for that processing to know about the playback environment as well as the content creator’s intent, in real time.

We want scalable coding

- Meaning, in particular, we want coding that can handle independent channels at higher rates, but drop down conveniently to 2-channels at lower rates
- We want decoders able to get all 5 channels in a mobile device, including time and dispersive cues, and render them into a headphone virtualizer.
- This is a difficult problem. Channel scaling is as yet only partially solved, and not solved in the sense of independent coding of channels.

We Need to Adapt to the Endpoint

- We need to be able to have enough information in the bitstream to do this in a smart way.
  - We need to know direct vs. dispersive content in each channel
  - We need to know the intended rendering position (usually standard, of course)
  - We need to know some data about the intended acoustic environment so that we can properly render that in the home environment.

We need to adapt to the network

- When bandwidth is available, we must use it.

- When bandwidth is limited, we must cope with the issue, but
  - We must NOT compromise the basic data
  - We must NOT force high-rate material to the lowest common denominator.
- This means metadata beyond channel count

The point?

- We have at least reasonably conquered the video problem, now we need to deal with the attendant audio issues.
- We need a good, real “3d” audio for the ‘3d’ video (which is really “2d plus distance”)
- We can not afford to take a step back in verisimilitude
  - If we do, the internet will bury us.
  - So, we need to pay attention to the future, and not the past, in audio processing and transmission.