



**dts**<sup>®</sup>

Digital Entertainment

# Next-Generation 3-D Audio

## Creation, Transmission and Reproduction

***Jean-Marc Jot***  
***Sr. Director, Audio Processing R&D***  
***DTS, Inc.***

*February 15, 2011*

## Objectives

- Analyze the technological opportunities that present themselves to the broadcast industry for elevating the consumer's audio experience beyond the current state of the art
- Derive some key new feature requirements of a next-generation audio coding format.

## Overview

- Current state of the art
- Opportunities for improvement in the audio experience
- New content creation paradigms: from “surround” to “3-D” soundtrack production
- Reproduction: conveying the intention of the content creator to the consumer
- Storage and delivery formats and requirements
- Recap / recommendations

## Attributes of high-fidelity audio reproduction

... conveying the creator's intent to the listener

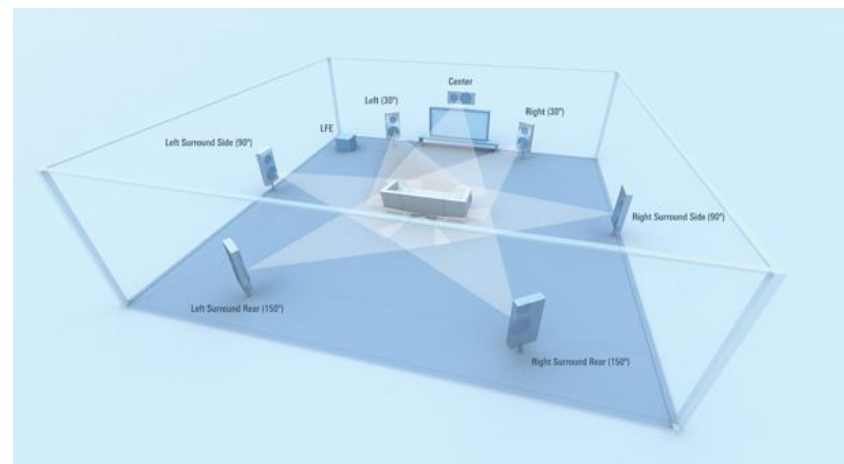
- Loudness and dynamics
- Voice/dialogue intelligibility
- Bandwidth and spectrum
- Spatialization

... suspension of disbelief

- Eliminate perceptible technology artifacts

## Current state of the art: spatial reproduction in A/V broadcast

- Frontal image presentation
- Frontal or horizontal surround audio presentation
  - the “dimensionality” of the audio experience is constrained by the limitations of the standard loudspeaker layouts
  - the assumed layouts determine the recording formats
  - content production tools are constrained by the recording formats



### Opportunities for improvement...

- The number of playback channels does not determine spatial fidelity
- Format-agnostic spatial audio.

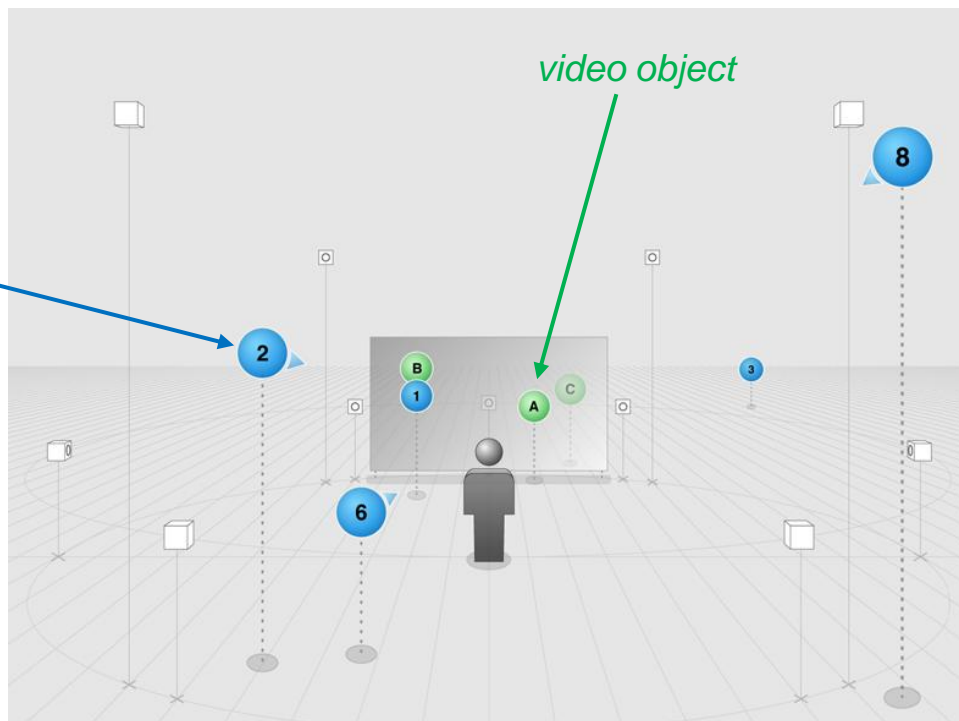
## Opportunity for improvement: beyond surround sound

... reproducing the perceptual attributes that characterize natural, 3-D audio

- elevation of sound events (audio objects)
- proximity and depth of sound events
- envelopment in diffuse sounds.

*audio object*

*video object*



## Current state of the art: fidelity of reproduction

- In the theater:
  - consistent, controlled audio and image presentation
- In consumer environments:
  - inconsistent, uncontrolled presentation of the coded/transmitted audio signals
    - Image dimensions, audio/video spatial congruence
    - Transducer capabilities and configurations
    - Environment acoustics and noise characteristics

### Unfortunate consequence...

- Ensure consistency by compromising fidelity at the source
  - Content production constrained for acceptability in worst-case playback conditions

## Opportunity for improvement: beyond coding

... controlled fidelity of audio scene reproduction

- loudness/dynamics/spectral fidelity
- dialogue intelligibility
- spatial fidelity
  - spatial coincidence between image and frontal sound stage, when possible

## Content creation: from “surround” to “3-D” soundtrack production

... two possible approaches:

- High-resolution multichannel recording formats
  - elevated loudspeaker-feed channels (e.g. NHK 22.2)
  - high-order Ambisonics
- Object-based soundtrack formats
  - examples:
    - interactive/game audio API standards (OpenAL, OpenSL ES)
    - synchronized multimedia standards (MPEG-4 Audio BIFS, 3D MIDI)
  - principles:
    - “unmixed” audio soundtrack → “rendering” at playback time
    - scene description: “audio objects” + rendering metadata/parameters
    - agnostic to spatial audio playback configuration
    - comparable to mixing automation (“session” format)

## Reproduction: conveying the intention of the creator to the consumer

- Fixed post-processing
  - playback system calibration/correction
    - at the factory (mobile/TV/Automotive) or in the home theater (A/V receivers)
  - surround-sound virtualization for headphones or frontal loudspeakers
    - does not take full advantage of 3-D audio capabilities of the playback system
- Content-aware post-processing
  - “blind” post-processing
    - processing “cues” driven by running audio signal analysis during playback
  - “guided” post-processing
    - using “cues” provided as metadata / ancillary data
      - offline signal analysis
      - producer-generated information

## Examples of “blind” post-processing

- Loudness & Dynamics control in the end-user environment
  - Volume leveling (DTS Symmetry)
  - Environment noise compensation (DTS Clear Audio)
  - Loudness maximization (DTS Boost)
  - Spectral loudness correction (DTS Enhance)
- Dialogue enhancement
  - applications: late-night listening, hearing-impaired listeners
  - limitations: assumes that dialogue is center-panned and isolated

### Advantages of post-processing

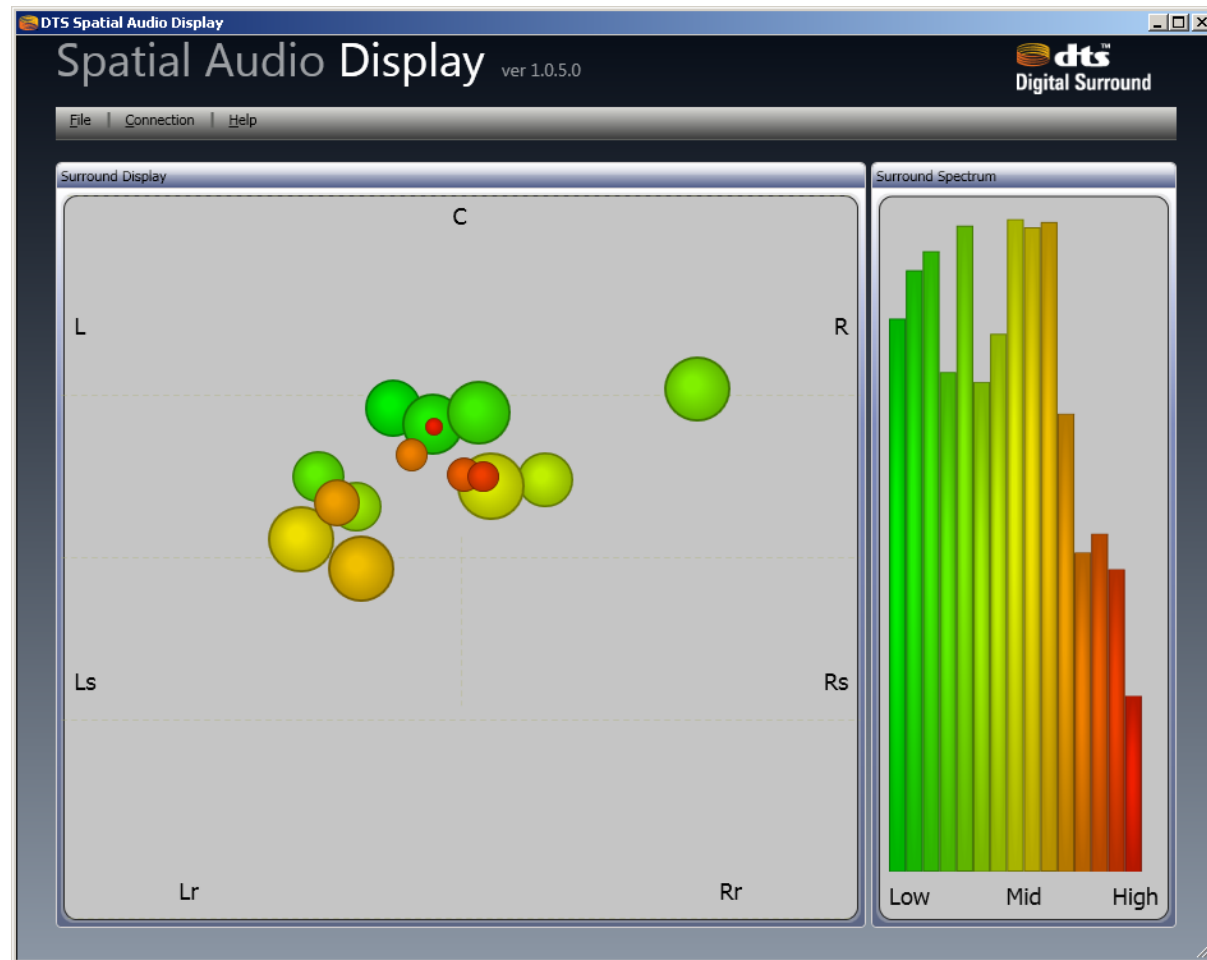
- Minimally obtrusive processing to fit playback conditions
- Uncompromised content dynamics (no need for DRC at the source)
- Uncompromised audio fidelity in favorable playback conditions

## Examples of “blind” post-processing

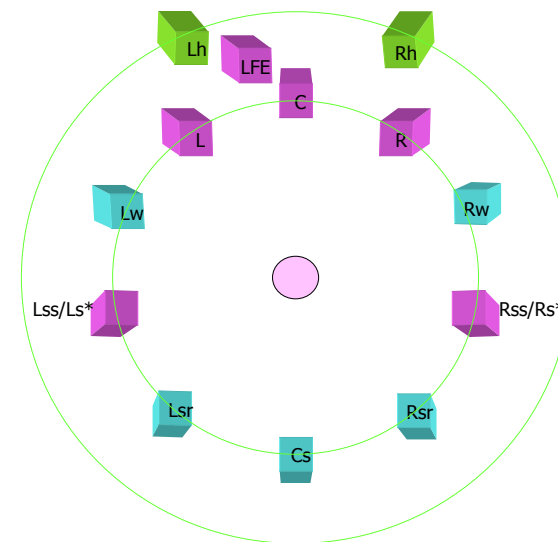
- Spatial audio format conversion
  - Virtualization
  - Stereo-to-surround up-mix
    - time-domain algorithms
    - frequency-domain algorithms (DTS Neural Surround, DTS Neo:X)
      - improved source separation/”discreteness”
  - Surround-to-3D up-mix
    - Loudspeaker “filling”
    - Generation of discrete height cues
      - format-specific height encoding/decoding
        - » not compatible with surround playback configurations
      - generic height cue extraction (DTS Neo:X)
  - Future perspective: flexible multi-channel format conversion

## Surround-to-3D up-mix

- Interior panning can be interpreted as elevation



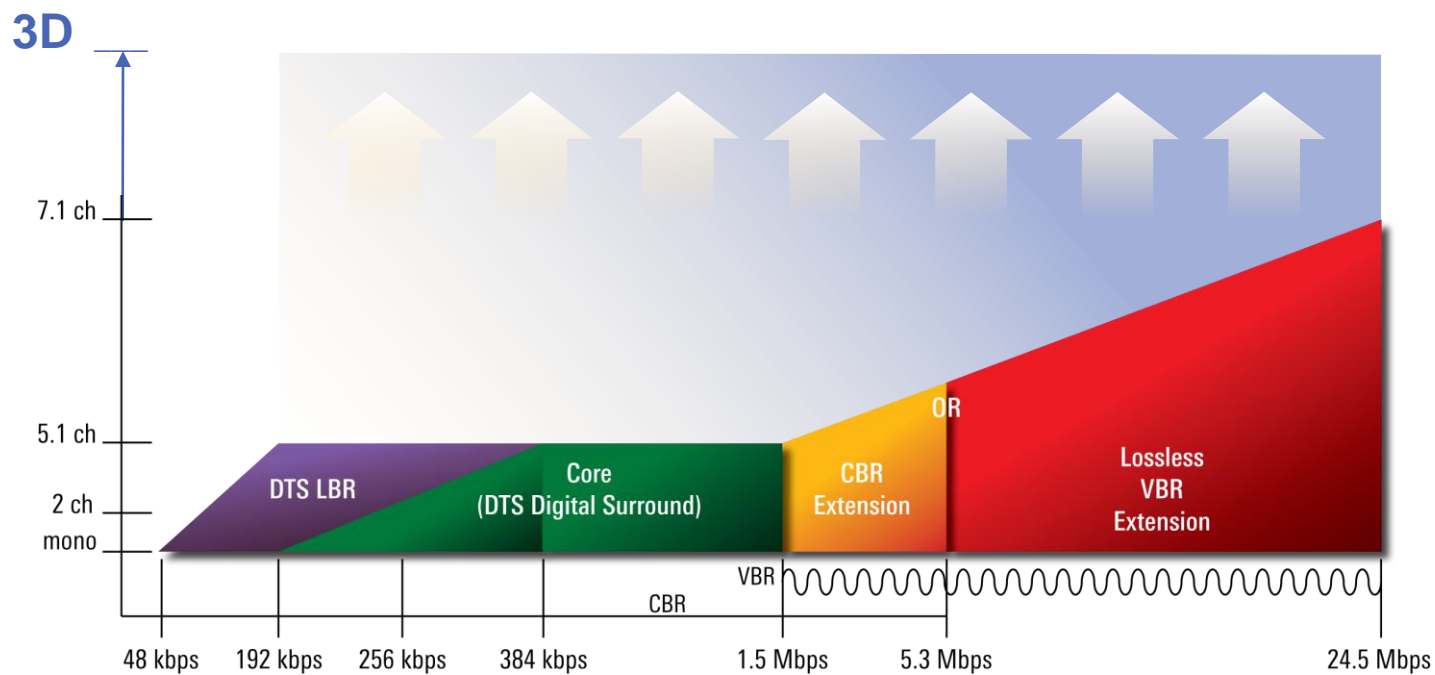
## Surround-to-3D up-mix



- Flexible multi-channel up-mixing solution (from 2.0 input to 11.1 output.)
- Surround-to-3D\* enhancement solution compatible with BD media. (\* - hemispherical 3D)

## Surround-to-3D up-mix

- Seamless soundtrack “upgrade” from Surround to hemispherical 3-D audio by post-processing adapted to the playback configuration.



## Examples of “guided” post-processing

- Audio metadata/cues
  - dialog normalization
  - reference/target SPL
  - ...
- Audio objects
  - applications
    - dialogue enhancement
    - optimal spatial audio rendering for the consumer’s playback configuration
  - approaches
    - frequency-domain spatial audio coding (MPEG SAOC)
      - limited separation
    - discrete audio coding
      - full separation, but potentially higher data rate

## Storage and delivery formats and requirements

- Backward compatibility
  - consumer decoders and playback configurations
  - production and broadcast infrastructures
- Universality, forward compatibility
  - create/store once, play everywhere (including future consumer systems/devices)
- Scalability
  - bit rate
  - end-user device capabilities

## Recap / recommendations (1)

- Bit-stream requirements
  - backward-compatible core + new extensions
    - include full soundtrack in legacy formats: 2.0, 5.1, 7.1
      - example: DTS-HD
        - » multiple layers of down-mix stored in one bit stream
  - include audio objects
    - dialogue objects
      - dialogue enhancement for hearing impaired, night listening
      - multilingual support
        - » requires full object separation
    - soundtrack elements selected for optimal spatial presentation
    - future-proof content production format
  - include metadata
    - target SPL
    - guided post-processing

## Recap / recommendations (2)

- Tying audio coding and post-processing
  - predictable post-processing
    - one content master for all playback configurations / user environments
    - enables/encourages uncompromised fidelity at content production
    - requires ability to monitor with post-processing at content production
  - applications
    - spatial audio format conversion/compatibility
    - loudness processing, environment noise compensation
    - dialogue enhancement

Thank You

The background features a series of white, overlapping, wavy lines that create a sense of motion and depth. These lines are set against a solid black background. At the bottom of the image, there is a horizontal grey bar containing the text "SCIENCE • TECHNOLOGY • ART" in a clean, white, sans-serif font.

SCIENCE • TECHNOLOGY • ART