Advanced Audio BIFS: Environmental Spatialization of Audio in MPEG-4 version 2

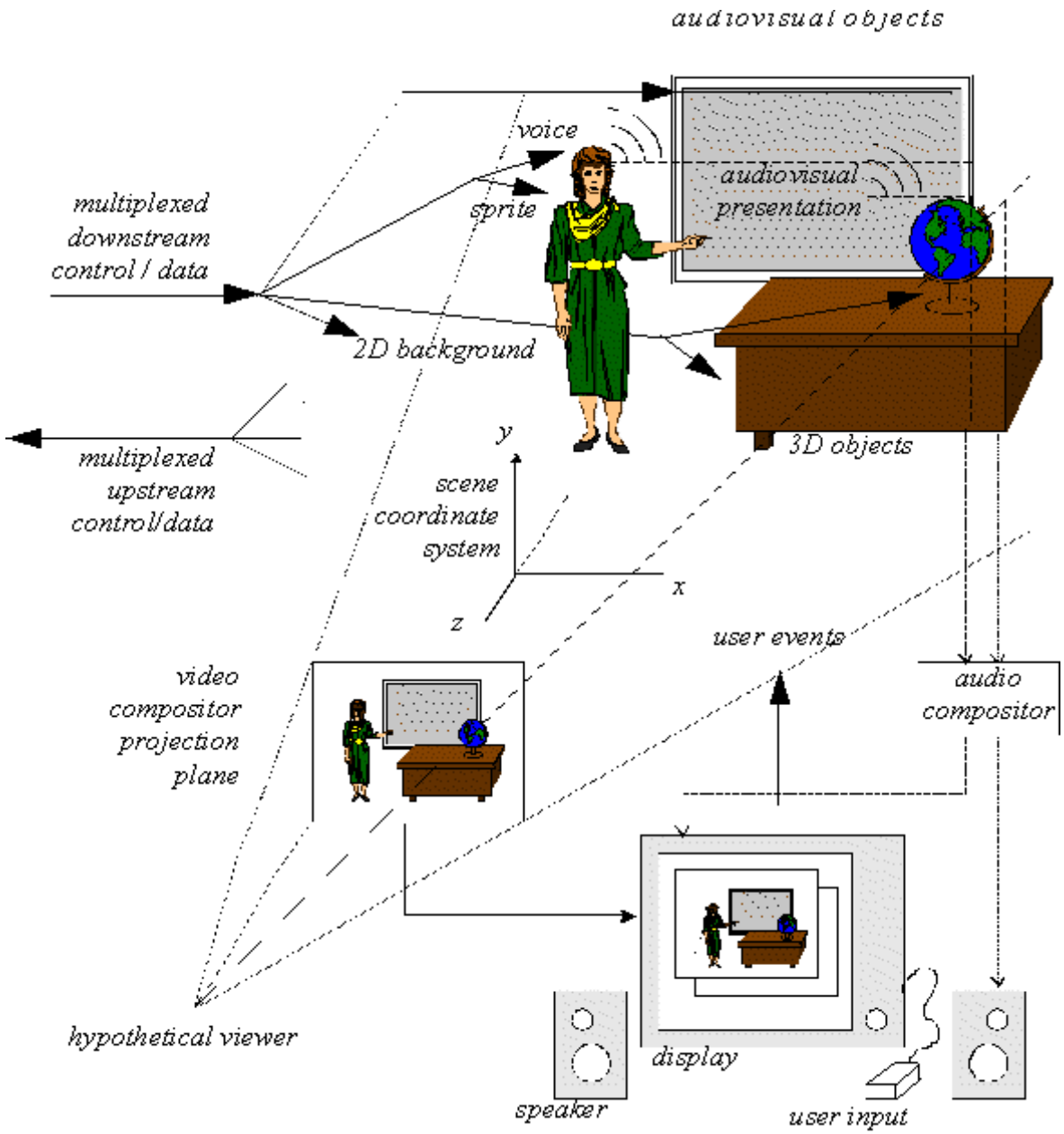
Jean-Marc Jot

Creative Advanced Technology Center

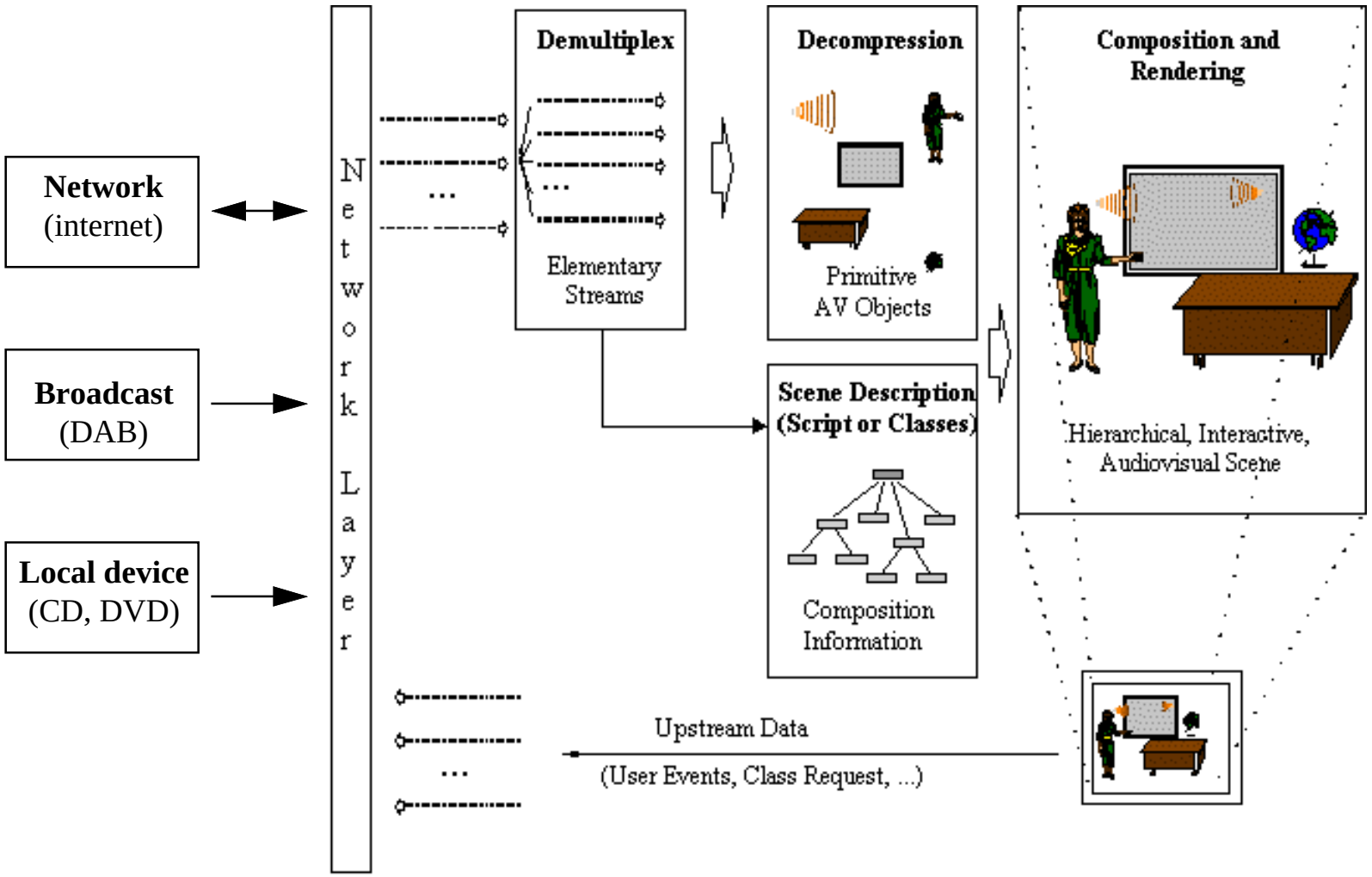
jmj@atc.creative.com

Basic Ideas

- Virtual worlds, Audio/visual coherence.
- Scenes, viewpoints, rooms, sound sources
- Interaction of user/viewer with virtual world
- Real-time mixing/compositing of audio streams
- Authoring $\rightarrow$ storing/transmitting $\rightarrow$ rendering
- Audio Content description is independent from playback format (nb of channels, layout)


audiovisual objects
voice
sprite
audiovisual presentation
multiplexed downstream control / data
2D background
3D objects
multiplexed upstream control/data
y
scene coordinate system
x
z
user events
video compositor projection plane
audio compositor
hypothetical viewer
display
speaker
user input

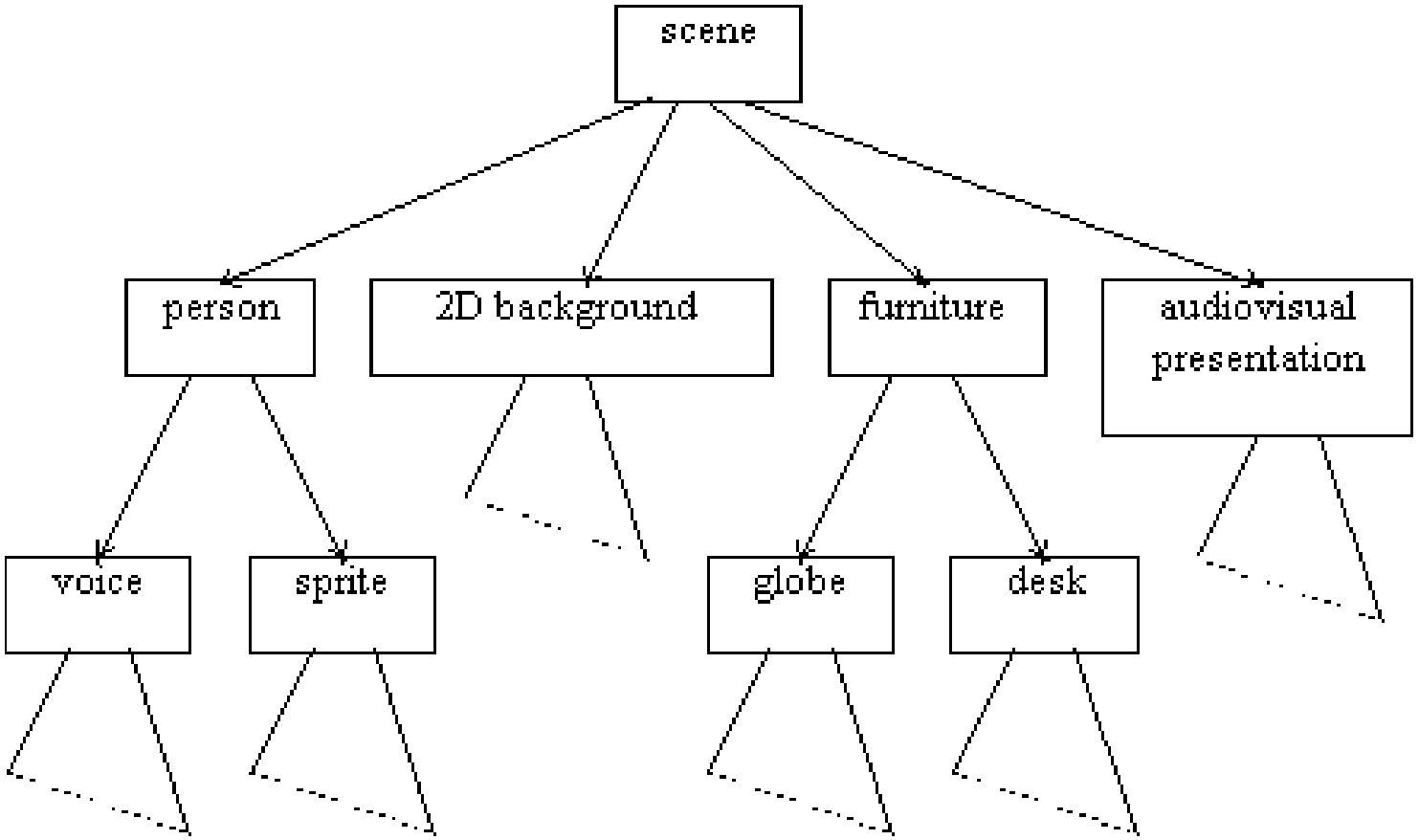
MPEG-4 Terminal



MPEG-4 BIFS
BInary Format for Scene description

- **Superset of VRML 2.0**
  - scene graph → nodes → fields
- **Functions**
  - compose MPEG-4 media objects into scenes
  - describe user interaction with objects
  - animate MPEG-4 objects
- **Binary format**
  - compression of scene data for transmission/storage
  - fast parsing for decoding BIFS stream

Scene Graph


scene
person
2D background
furniture
audiovisual presentation
voice
sprite
globe
desk

MPEG-4 AudioBIFS

- **Compositing** (post-processing + mixing)

  of elementary audio streams produced by audio decoders

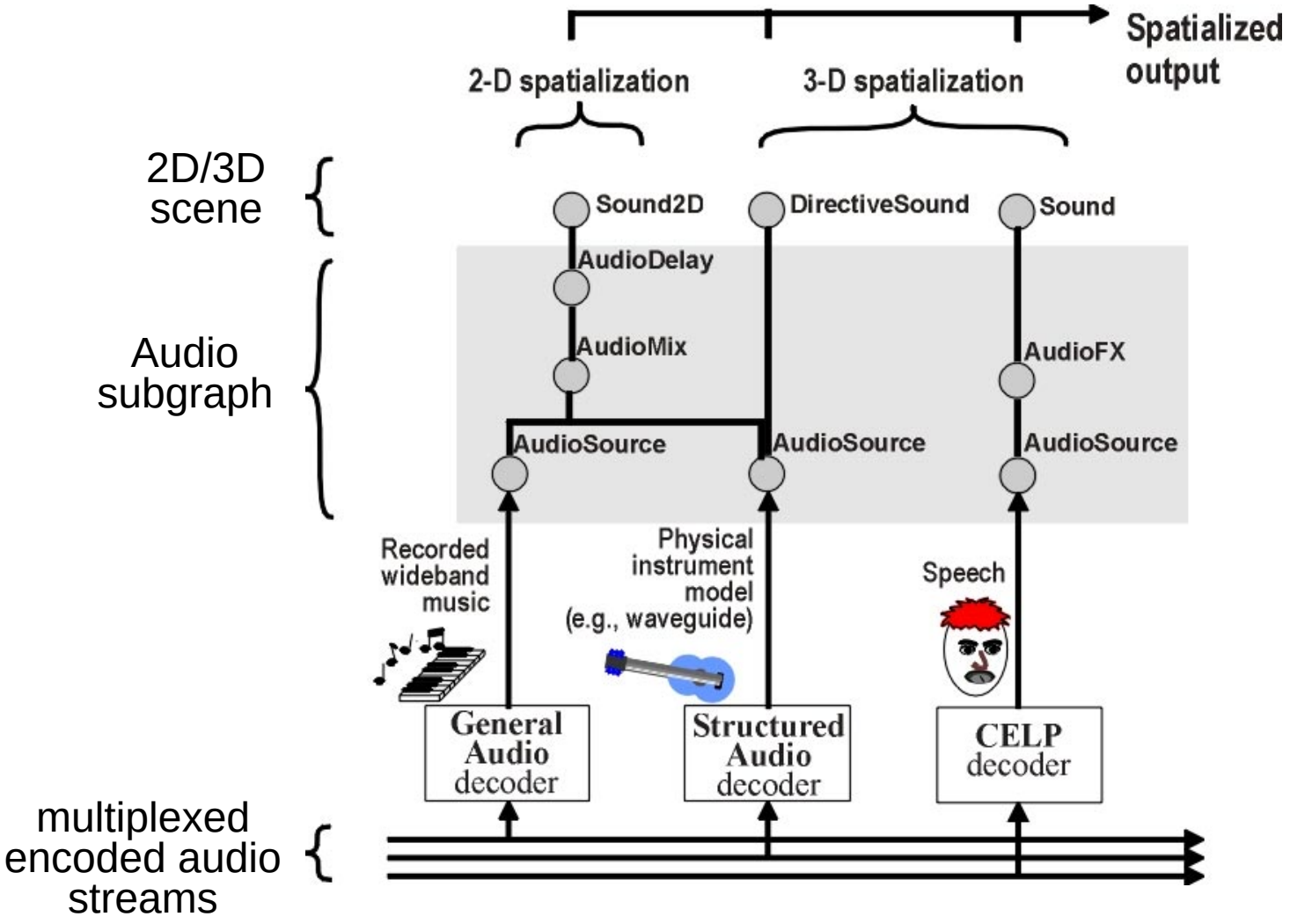
- **Interactivity:** scene author can allow end user to
  - change viewing/listening point (navigation in the scene)
  - drag objects in the scene to a different position
  - trigger a cascade of events by clicking on a specific object

- Two parts in AudioBIFS scene graph:
  - **2D/3D scene** (audio + visual)

    positions sounds in 3D space, independent of playback format
  - **Audio subgraph:** DSP signal flow (patching paradygm)

    mix several streams, apply effects described in SAOL language

AudioBIFS Scene Graph



(Riitta Vaananen)

Audio Mixing/Compositing in MPEG-4

- **MPEG-4 version 1**
  - (1) Audio subgraph — *Music/soundtrack post-production*, *Foley effects, sound design*
  - (2) 3D scene graph — *Immersive virtual worlds*, *Interactive 3D games*
- **MPEG-4 version 2: Advanced AudioBIFS**
  - (2a) Physical approach ← *VR, predictive acoustics*
  - (2b) Perceptual approach ← *Immersive computer music*
    - Frequency dependent sound-source directivity model
    - Environmental auditory cues: reverberation, reflections, occlusion by obstacles (transmission or diffraction)
    - Two complementary approaches: physical, perceptual

Physical ESA nodes

- **DirectiveSound**
  - **source** (allows connection to an audio stream)
  - **location**, **direction**, **intensity**, **speedOfSound** (for Doppler effect)
  - **angles**, **directivity** filters (filter = {coefs} or {frequencies/gains})
  - **useAirabs**, **spatialize**, **roomEffect** flags
- **AcousticScene**
  - late reverberation **reverbTime**, **reverbLevel, reverbDelay**
  - **center**, **size** of bounding box (which must contain **DirectiveSound** nodes and **ListeningPoint** or **ViewPoint**)
- **AcousticMaterial**
  - attached to **IndexedFaceSet** of polygons and **AcousticScene**
  - acoustic *reflectivity* and *transmissivity* filters
  - visual properties (*color*, *transparency*...)

Physical Approach

Demonstration

DIVA Project
(Helsinki Univ. of Technology)

Perceptual ESA nodes

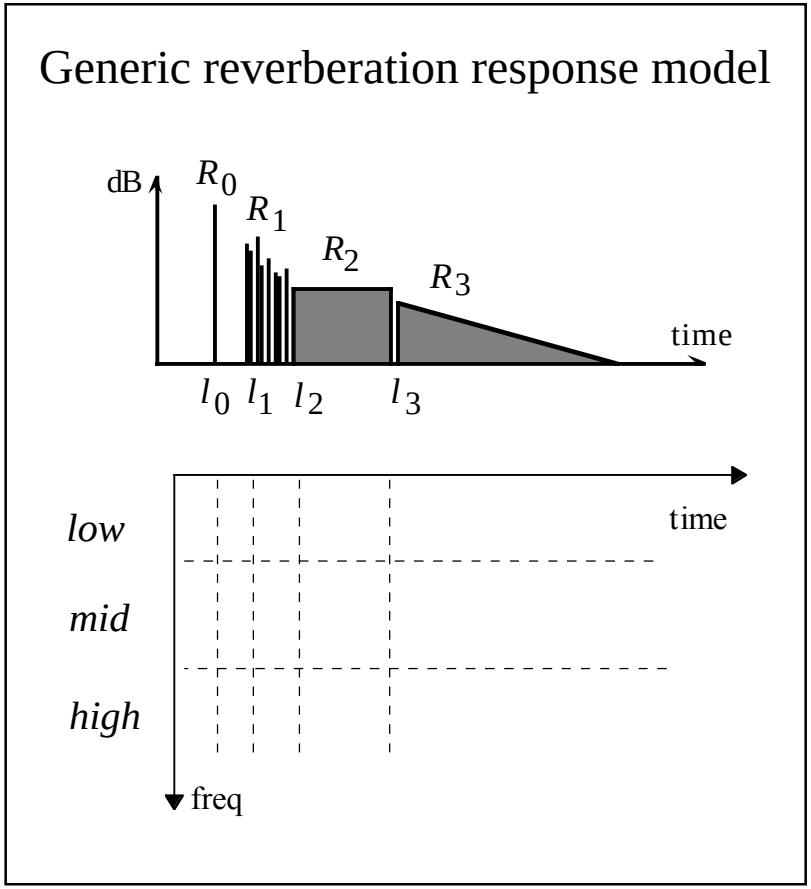
- **DirectiveSound**

  (same as in physical approach)

- **PerceptualParameters**

  (child of **DirectiveSound**)
  - timeLimits, modalDensity, freqLow, freqHigh
  - direct and input filters, diffuse-field filter
  - perceptual parameters

    source presence, warmth, brilliance, room presence, envelopment
    running reverberance, late reverberance, heavyness, liveness
  - refDistance



Perceptual Approach

Demonstration

Spatialisateur
(IRCAM - France Telecom)

What is in the MPEG-4 standard?

Information contained in the standard

- BIFS node definitions and protocol (normative)

  including formulas: perceptual params → energy distribution

- Implementation examples (not normative)

  source code (both approaches)

  appendix (perceptual approach)

Not in the standard (except as non-normative illustration)

- 3D positional audio rendering method
- Artificial reverberation algorithm
- Physical room modeling method

Conclusion/Perspectives

- **Video conferencing, group communications, Multiplayer online games**
- **Audio/Video Broadcast**
- **Music/soundtracks**

  Content representation independent from playback format

  – need for DSP ressources in the terminal (e.g. PC soundcard)

  – need to transmit/record individual source channels can be satisfied thanks to compression
    – natural audio → 1:10
    – synthetic audio → down to 1:100

Contributions to ESA in MPEG-4

Physical approach:

- Riitta Vaananen (HUT) — Co-editor of MPEG-4 v2 (Systems part)
  Implementation of physical approach
- Jyri Huopaniemi (Nokia)

Perceptual approach:

- Jean-Bernard Rault (FT) — Implementation of perceptual approach
  Marc Emerit (FT)
- Jean-Marc Jot (Creative)
  Olivier Warusfel (IRCAM)

References

- MPEG-4 v2 Systems FDAM (w3054)
- MPEG-4 Overview (w2995)

  http://drogo.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm
- Savioja & al.: DIVA project (HUT)

  JAES, vol. 47, no. 9, pp. 675-705, Sept. 1999.
  ICAD 97. http://www.icad.org/websiteV2.0/Conferences/ICAD97/
- Jot & al.: Spatialisateur (IRCAM/FT)

  Multimedia Systems, Vol. 7, no. 1, 1999.
  ICMC 97. http://mediatheque.ircam.fr/articles/textes/Jot97b/
- Scheirer & al.: Structured Audio, AudioBIFS

  Multimedia Systems, Vol. 7, no. 1, 1999.
  IEEE Trans. Multimedia. http://sound.media.mit.edu/~eds/
- Signes: BIFS

  http://drogo.cselt.it/mpeg/documents/koenen/signes.zip