

MPEG-4 and 3D-Audio

Andreas Dantele
dantele@web.de

ftw. Telecommunications Forum, 14.02.05
Forschungszentrum Telekommunikation Wien

Content

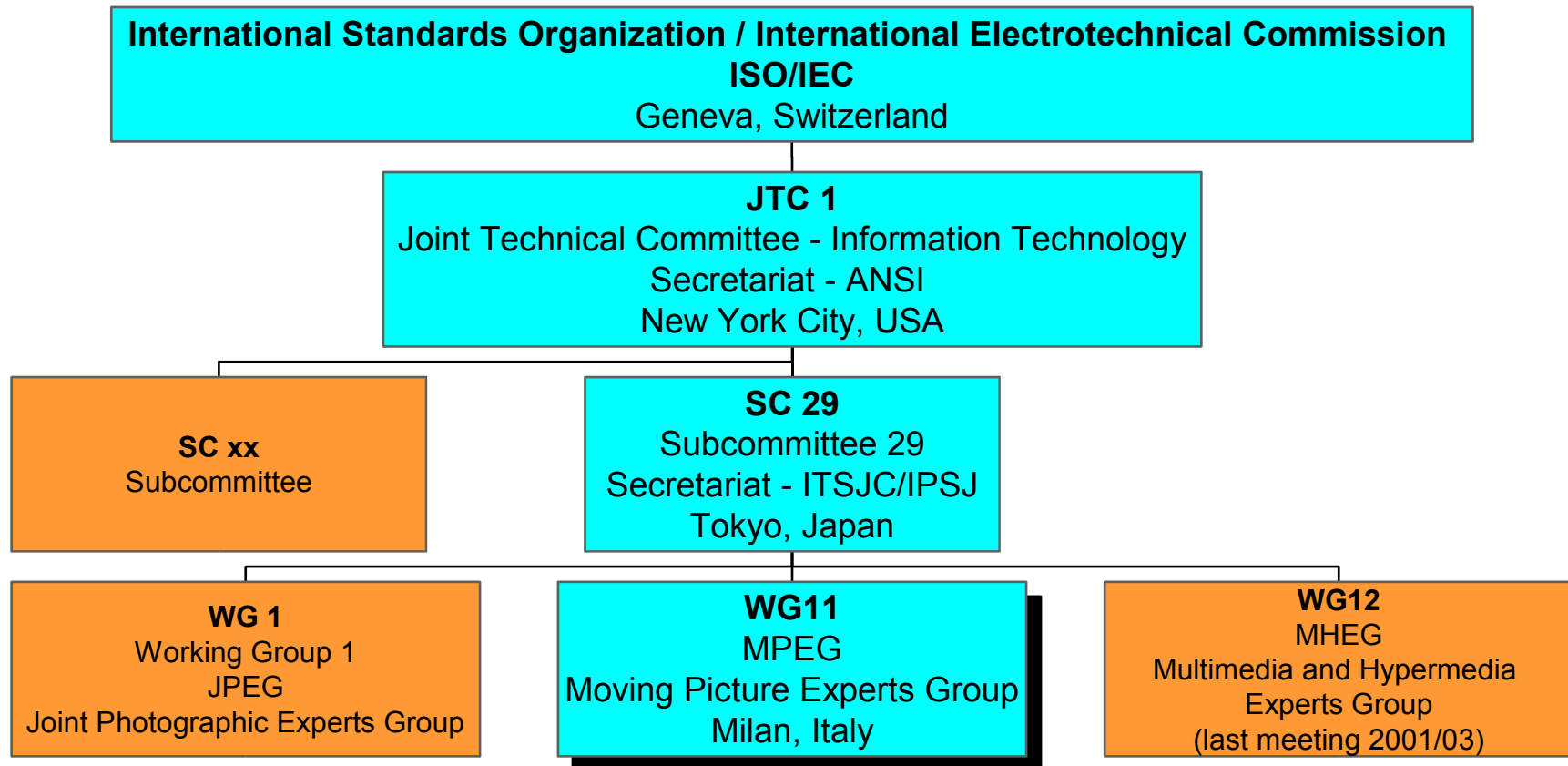
- Moving Picture Experts Group (MPEG)
- Object-based content - What for?
- The MPEG-4 Standard
- 3D Audio – Scene Description
- 3D Audio – Rendering
- Multimodal Perception
- Demonstartions

The Moving Picture Experts Group



- formal:
ISO/IEC JTC 1, SC 29, WG 11
- worldwide consortium of experts
(from industry, universities, research institutes, etc.)
- 4 meetings/year, 300-400 delegates
- area of work (from Terms of Reference):
 - *“Development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio, and their combination, in order to satisfy a wide variety of applications.”*

MPEG (2)



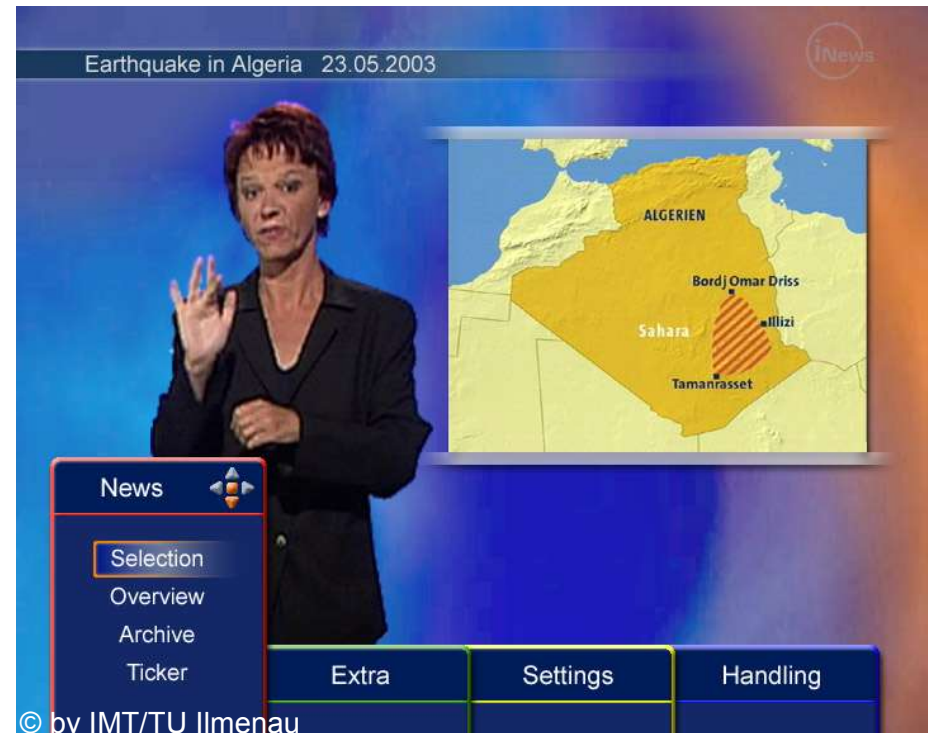
MPEG Standards

- MPEG-1 (1992), MPEG-2 (1994)
 - widely adopted in multimedia industry
 - Digital TV, CD-i, Video-on-Demand, MP3, ...
- MPEG-4 (1998/2000)
 - Object-based audiovisual representation
- MPEG-7 (2001)
 - Multimedia Content Description Interface
 - Metadata
- MPEG-21 (being developed)
 - Multimedia Framework

Object based content - What for?

Interactive News:

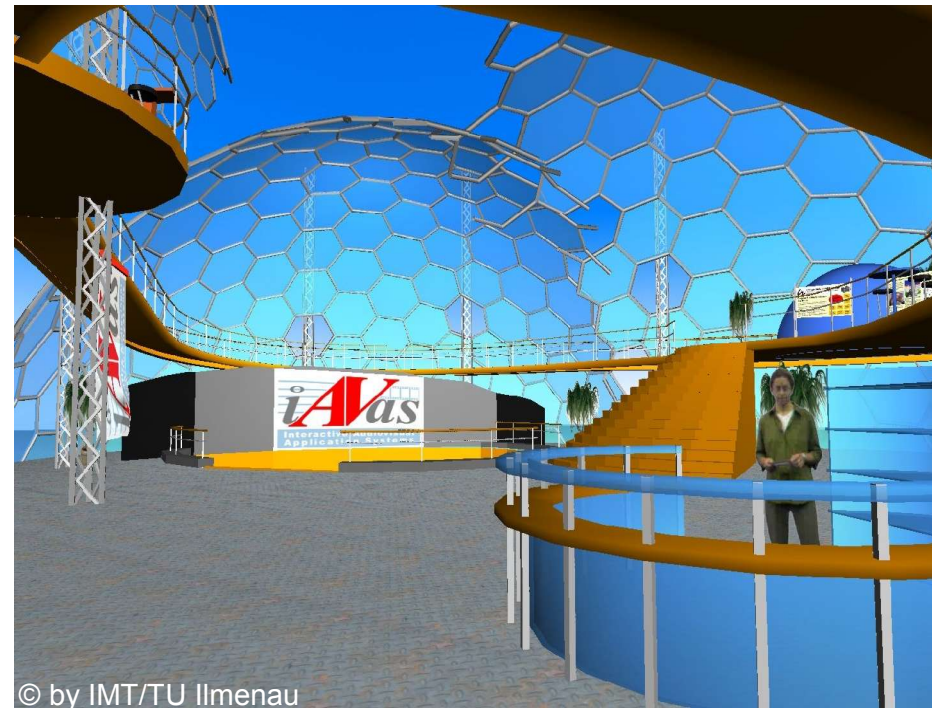
- news selection
 - access to background elements
 - customizable ticker
 - selection of presenter
 - selection of language
- etc.



Object based content (2)

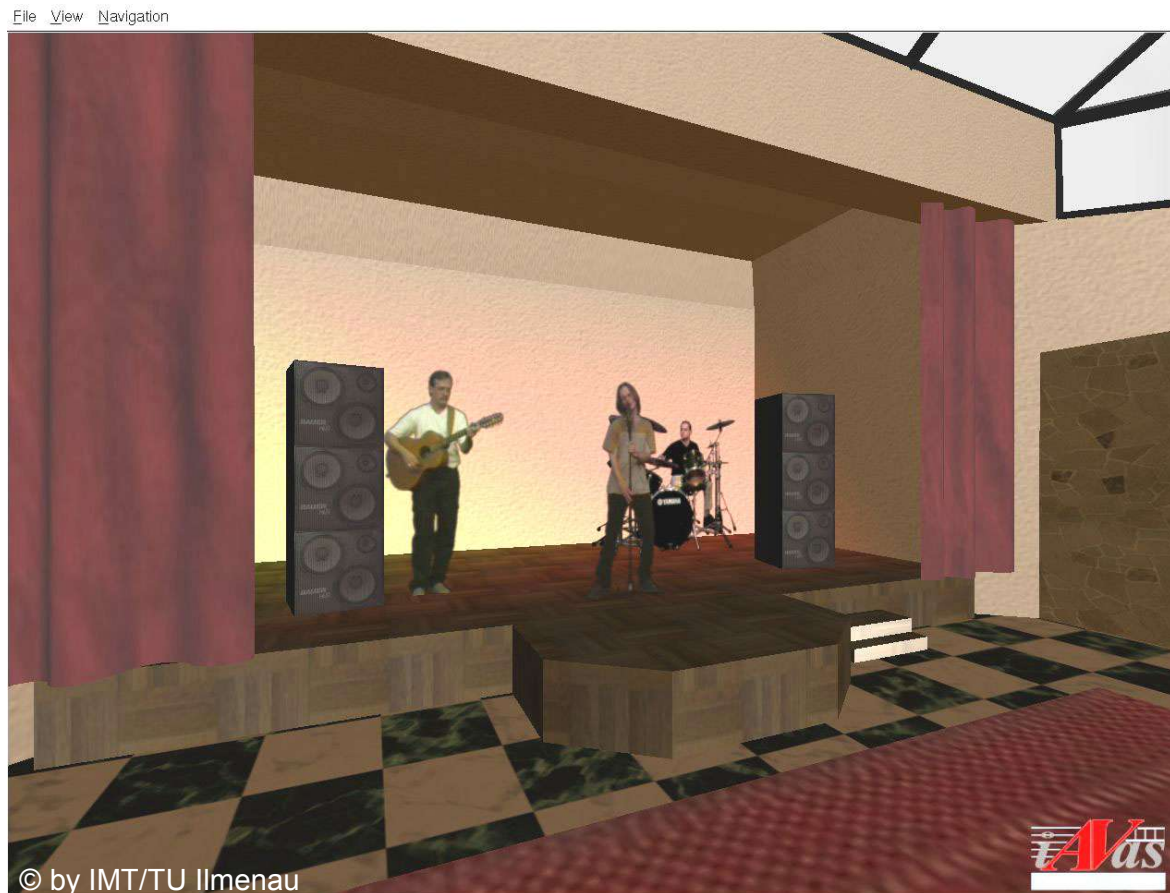
Interactive Fare

- virtual conference center
- Interactive access
- conventional (speaker, poster, etc.) and multimedia content
- rather no limitations for exhibitors (time, space, what to show, etc.)



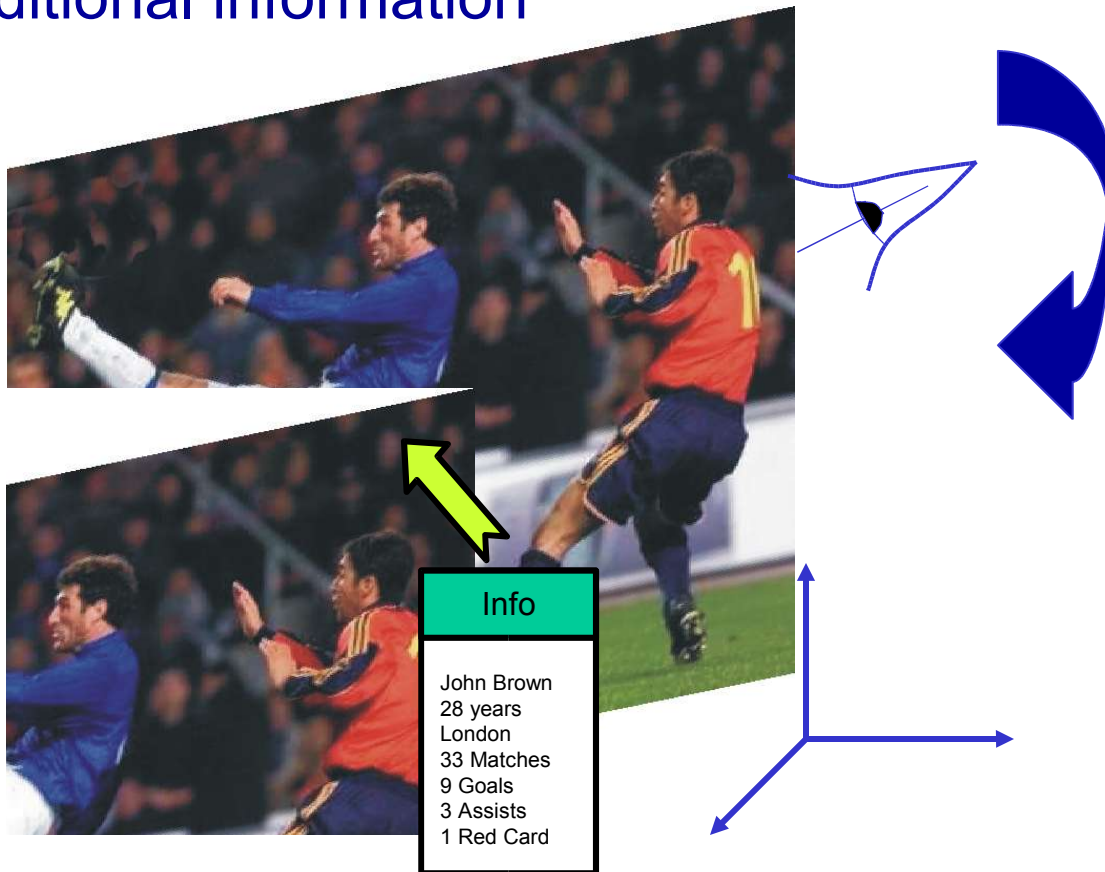
Object based content (3)

- Interactive music show: manage sound impression, selective listening, move around freely, etc.



Object based content (4)

- Sports: free choice of perspective, object picking, additional information



Advantages

- enables interactivity for the consumer
passive user → active user
 - scalability:
adapt to bandwidth, processing power, etc.
 - use on different user devices:
Mobile Phone, PDA, TV-Set, PC, etc.
 - delivery methods:
Internet, DVB, UMTS, DVD, etc.
 - reuse of content
- *new generation of multimedia experience*

The MPEG-4 Standard

- not only an enhancement of MPEG-1/2
- object-oriented approach:
 - A/V-content subdivided into objects
 - object- and scene-description tools
 - separated coding and transmission
 - composition in 2D/3D scenes
 - reuse of content
- ➔ interactivity provided for the user

The MPEG-4 Standard (2)

- version 1 established 1998
- version 2 (backward compatible) est. 2000
- extension is still going on
- subdivided in parts:
 - pt. 1: Systems (scene composition, technical concept incl. synchronization and multiplexing of media objects)
 - pt. 2: compression codecs for visual data
 - pt. 3: compression codecs for audio signals
 - other parts: IPMP, conformance testing, reference software, file format, advanced codecs, etc.

Basic principles

- auditory and visual objects are spatially and temporally assembled in a scene
- sophisticated scene description concept
- custom encoding for every type of object
- scalable encoding
- description by Object Descriptors
- complexity measured by profiles
- terminal/player reproduction not standardized

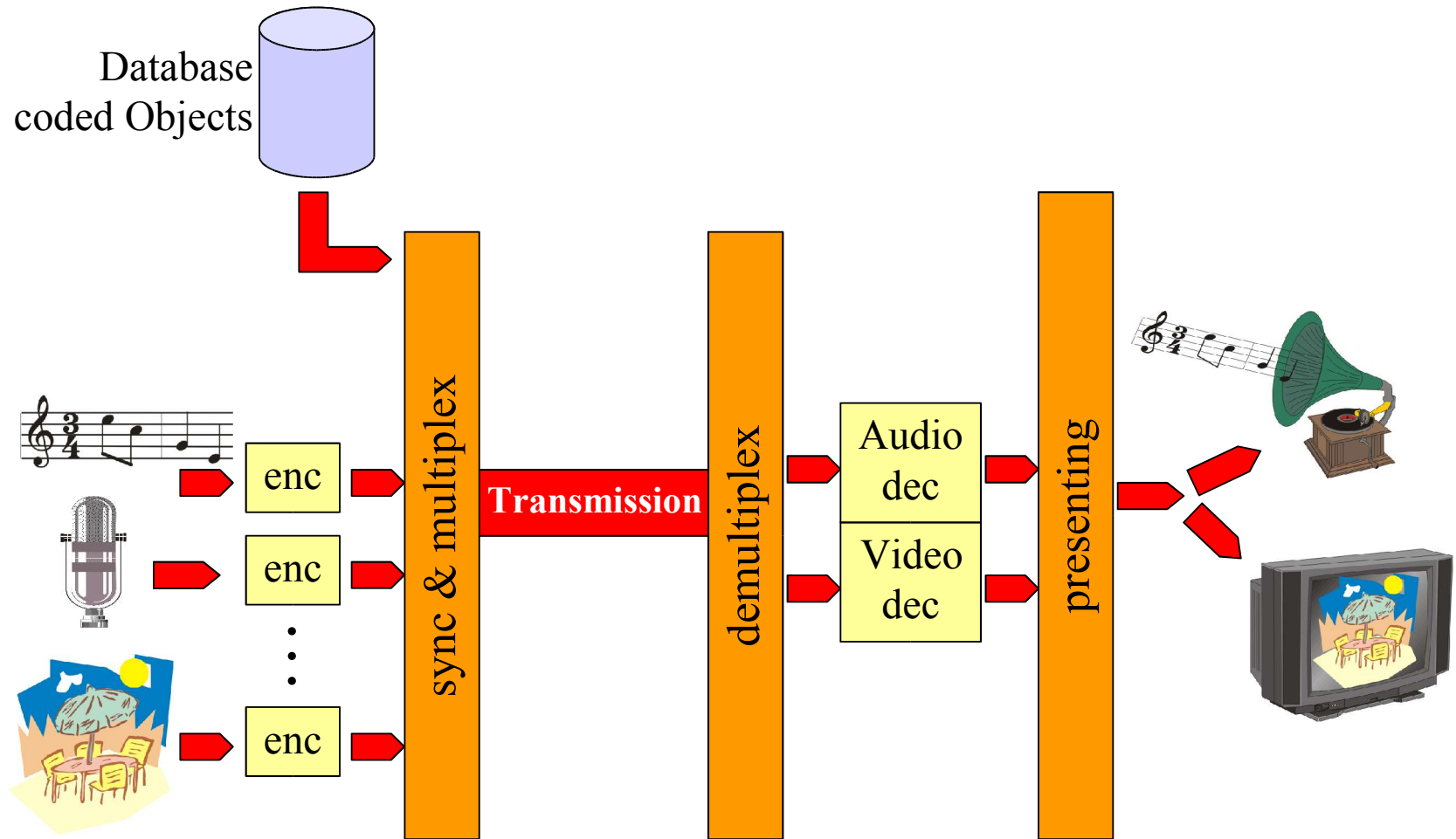
The Codecs

- efficient coding schemes for
 - music and speech
 - video (rectangular and arbitrary shape)
 - text and graphics
 - specific 3D objects (as human face & body animation)
 - synthetically generated speech and music
- scalable from low bit rates to high quality conditions in order to adapt to
 - transmission capacity
 - terminal (player) constraints
- provide error resilience

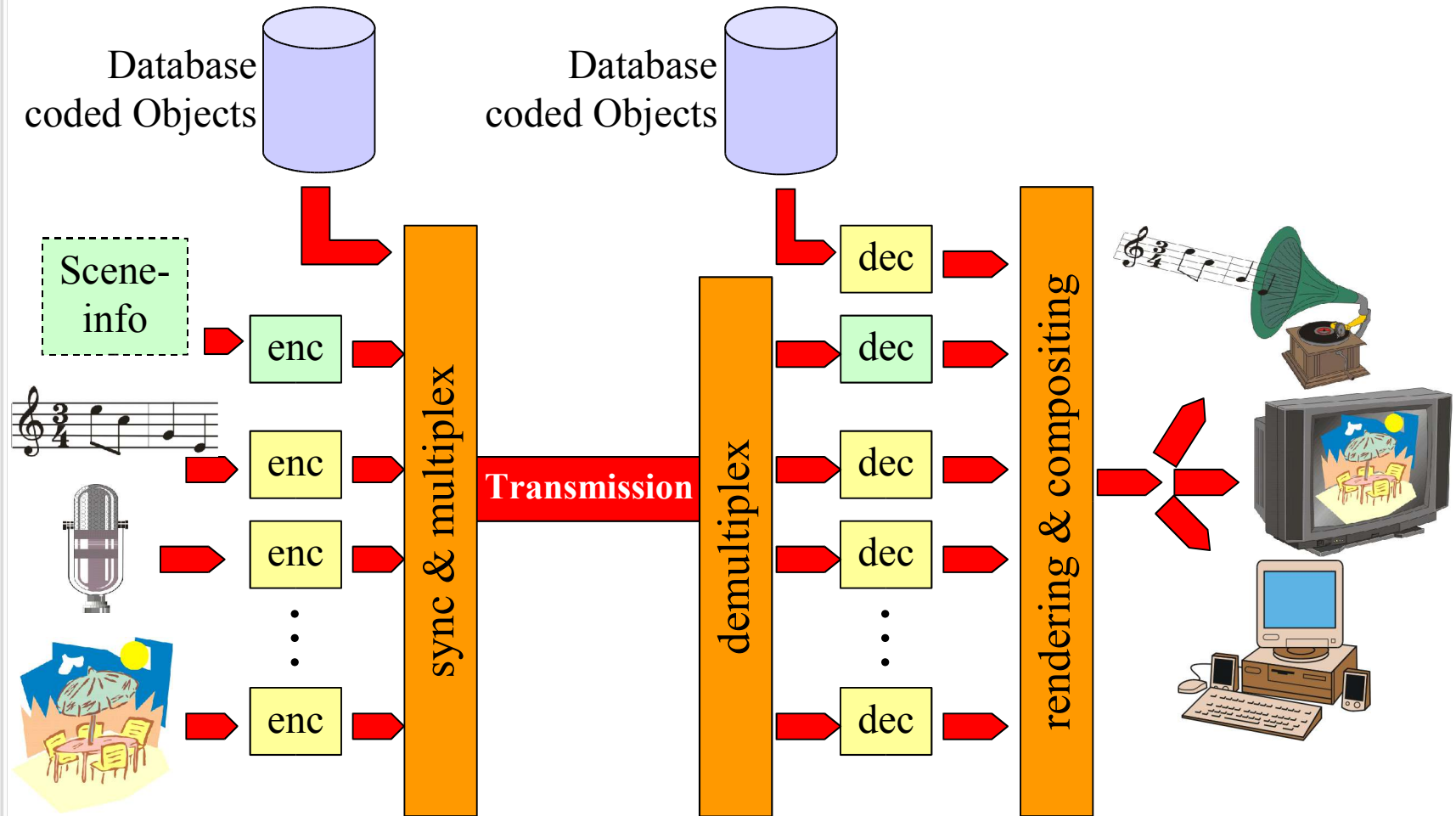
Systems Tools

- object description
syntax of the bit stream for one single object
- scene description toolset:
Binary **F**ormat for **S**cene Description (BIFS)
based on VRML (tree concept)
- profiling concept to manage complexity
e.g. a simple player cannot handle complex streams
→ low complexity profiles contain only basic tools/codecs
- profiles define the set of tools that can be used in a certain MPEG-4 terminal/player
- profiles are introduced upon needs

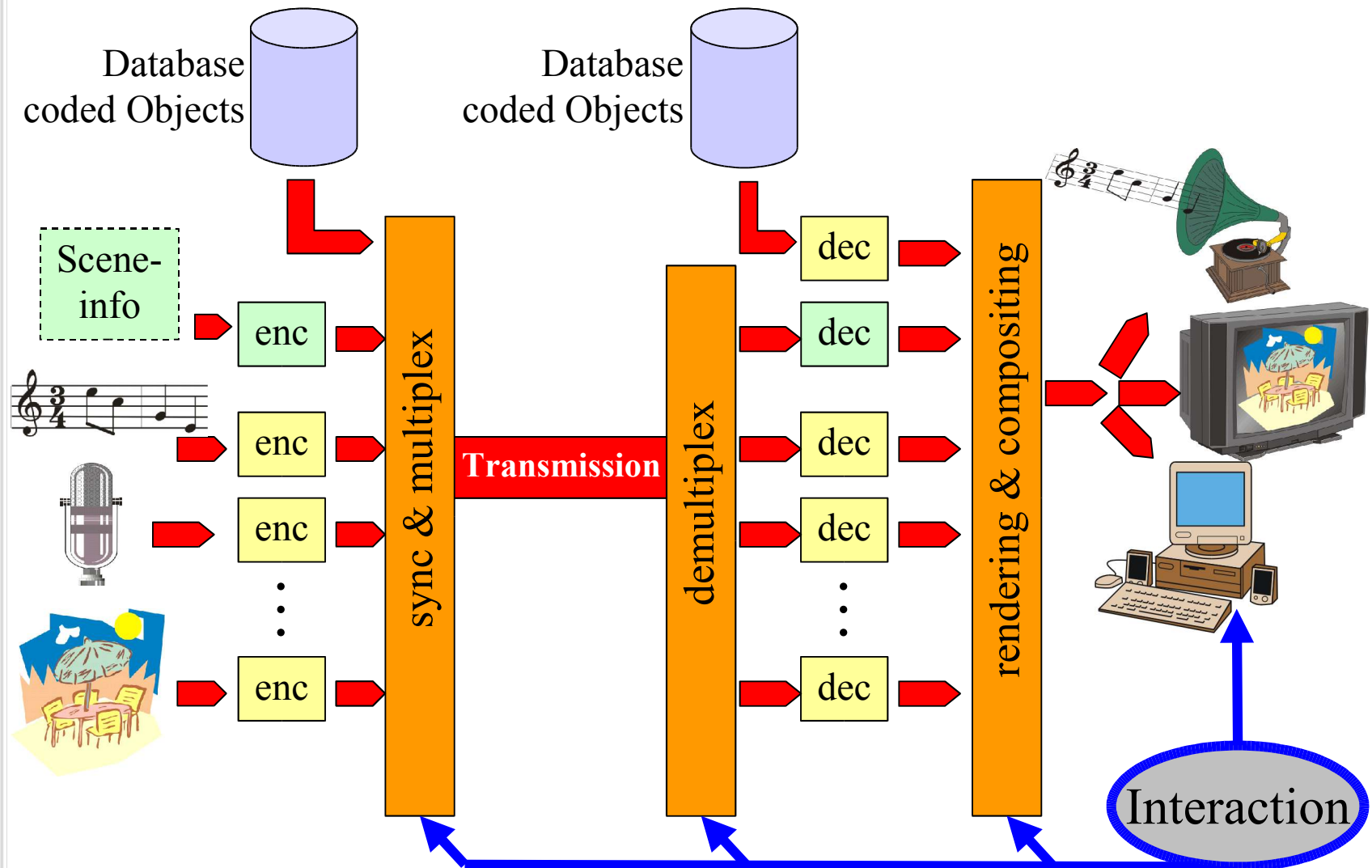
Common A/V-System



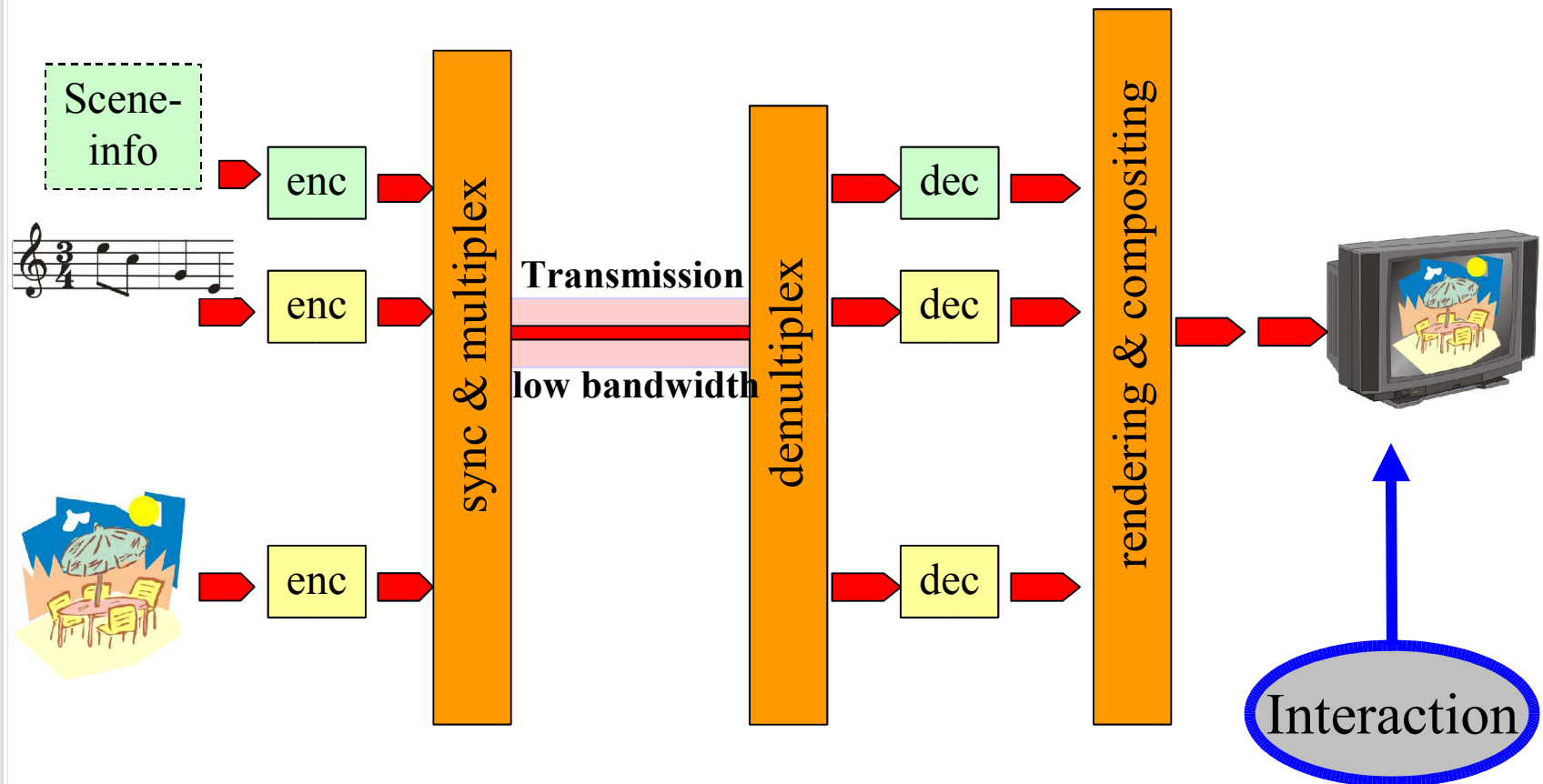
Object-based A/V-System



...and interactive



e.g. low complexity profile



Summary: Goals of MPEG-4

- for authors:
 - flexible scene description tools
 - reuse of content
 - IPMP system to maintain owners' rights
- for network service providers:
 - transparent technical information in standards
 - licensing under fair and reasonable terms
- for end users:
 - high levels of interaction with content
 - multimedia experience within new networks (mobile...)
- for all parties: avoiding the emergence of a multitude of proprietary, non-compatible formats and players

MPEG-4 in Ilmenau

- research project IAVAS (10/2001 – 9/2004)
Interactive AudioVisual Application Systems
Head: Prof. Karlheinz Brandenburg
- conformity with MPEG-4 standard
- virtual 3D scenery
- arbitrarily shaped video objects
- 3D audio
- interactivity
 - freely move around in 3D scene
 - object picking, play back, etc.
- development of a platform independent 3D player



Institute of Media Technology
Technische Universität Ilmenau



www.iavas.de

3D Audio

- more than 5.1
- creation of surrounding acoustics: *auralization*
- loudspeakers (many)
- or headphones
 - use of Head Related Transfer Functions (HRTFs)
 - head-tracker required for interactive moves
- realistic acoustic sensation
 - direction and distance of sound sources
 - reverberation
 - acoustic effects (directivity, obstruction, doppler, etc.)
- dry (unreverberated) sound files as sources

MPEG-4 AudioBIFS

- scene description tool-set for audio features
 - for A/V scenes or audio only scenes
 - AudioBIFS version 1:
behavior of sound emitting objects
 - position, level, directivity, etc.
 - AudioBIFS version 2:
acoustic behavior of environment
 - reverberation time, absorption coefficients, etc.
- thorough description of acoustic sensation possible
- AudioBIFS version 3 (being developed):
 - multi-channel coding schemes, plane waves, etc.

AudioBIFS (2)

- physical approach
 - based on physical properties
e.g. frequency dependent directivity
 - ***measurable, objective***

VS.

- perceptual approach
 - based on human sensation
 - psycho-experimental research necessary
 - focus shifted to the user (non-expert)
 - ***perceptual, subjective***

Why a perceptual approach ?

- drawback: subjective parameters
 - but:
 - scope is shifted towards the user/human perception
 - higher-ranking language
 - easy to understand for non-experts
 - communication between experts and non-experts
 - evaluation and comparison of test data
- ***lots of advantages !!!***


MPEG-4 perceptual approach

- demands on perceptual parameters:
 - satisfy human needs
 - easy to explain
 - ideally non-ambiguous and orthogonal
- unfortunately:
no wide spread 'language' available yet
- in MPEG-4: one system installed
 - developed at IRCAM, Paris/France
 - available in SW packages (Max/MSP)

Perceptual parameters

- Source related attributes
 - impression of the direct sound (distance, directivity)
 - *SourcePresence, SourceWarmth, SourceBrilliance*
- Room related attributes
 - surrounding acoustic space
 - relative damping of low/high frequency bands
 - *LateReverberance, Heaviness, Liveness*
- Source/room interaction
 - behavior of a sound source within a room
 - distribution of energy over time
 - *RoomPresence, RunningReverberance, Envelopment*

Evaluation

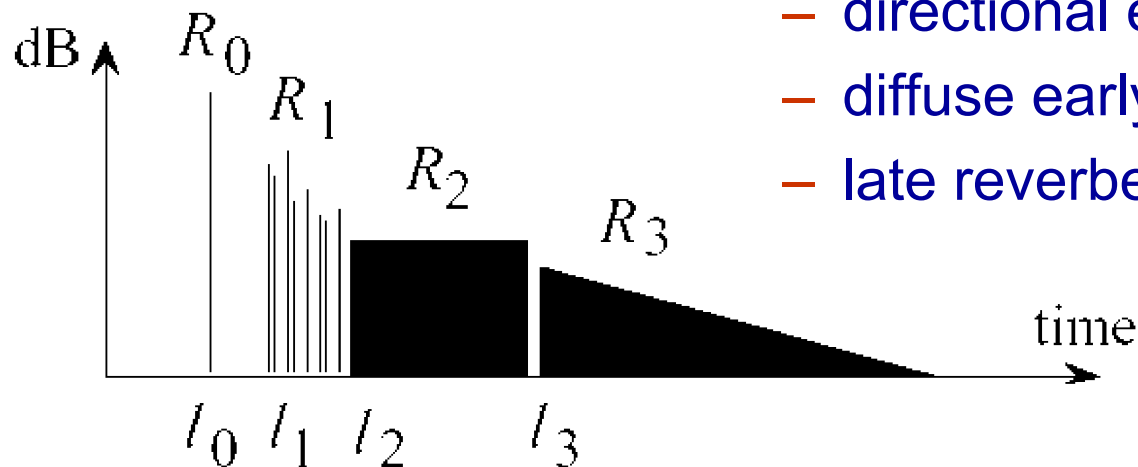
- coherent and easy to understand: yes 😊
 - independent: ?
 - mutual influence of *Envelopment* and *SourcePresence*: both yield changes in the perceived distance 😞
 - non-ambiguous: ?
 - David Griesinger says there are two main parameters:
 - “**apparent source width**”:
 - impression of spatial extent of a sound source
 - related to early energy of direct sound
 - in MPEG-4 called: ***Envelopment***
 - and “**envelopment**”:
 - reverberated sound
 - surrounding the listener
- 
 same name,
different meaning !!
- meaning of parameters depends on context 😞

3D Audio Rendering

- positioning/filtering of sound sources (direct sound)
- auralization according to reverberation model
 - often derived from impulse response
- rendering of early reflections
 - with directional behavior when applicable
- rendering of late reverberation: diffuse
- mapping process:
high level parameters → low level
- depends on reproduction system
 - e.g. ambisonics: decoder, many loudspeakers
 - or headphones with HRTF filtering

Example implementation

- reverberation model:
- **energy** distribution over **time** into 4 segments:
 - direct sound
 - directional early reflections
 - diffuse early reflections
 - late reverberation



- filtering: 3 **frequency** bands

A/V Applications

- acoustic and visual impression have to match to improve level of immersion
- audio rendering has to consider:
 - position and directivity of sound source(s)
 - room size and geometry (affects reverberation)
 - objects in the room (acoustic obstruction)
 - acoustic properties of walls, floor, ceiling and objects (acoustic reflection/transmission)
- library of acoustic properties in authoring-tools needed

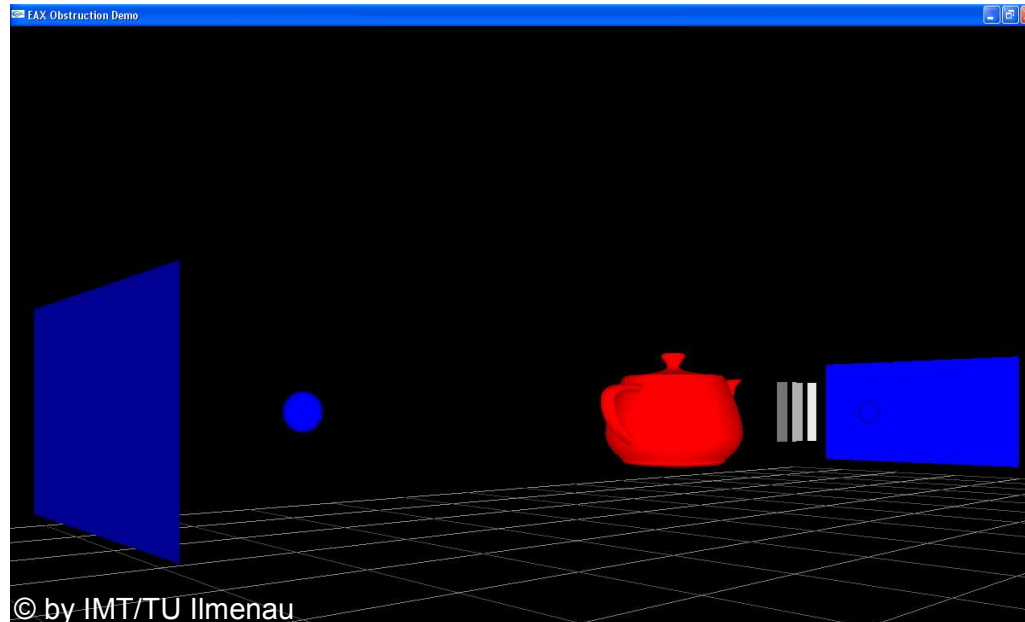
A/V Applications (2)

- early reflections are critical for acoustic impression
- calculation based on scene description
 - Ray-tracing/image source model
 - high demand on processing power
 - new approach: usage of GPU routines
- alternative: model based
 - cheap, not very accurate
- The question is:
How exact do we have to render the auralization when visual cues are present?

Multimodal Perception

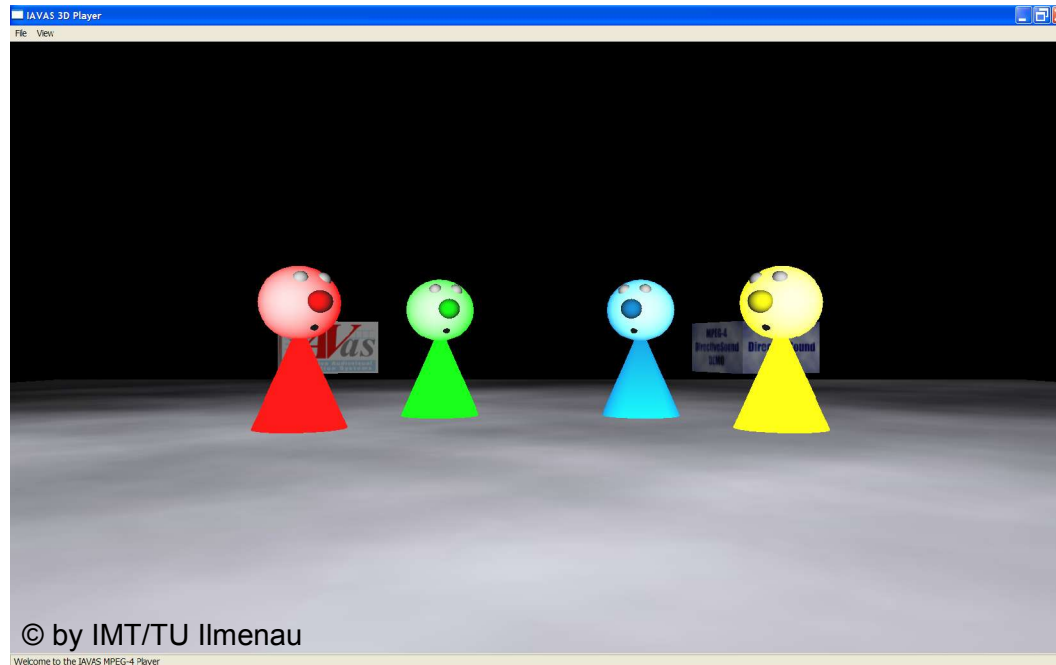
- many perceptive effects known for
 - auditory only reproduction
 - visual only displays
- combination A+V:
 - good when impression matches
 - bad when there is sensorial contradiction
- higher correlation between modalities
 - higher degree of immersion
 - stronger feeling of „being there“ (VR)
 - relief in stressful situations (jet cockpit)
- further research required
 - MPEG-4 provides a good basis for tests

Demonstration: Acoustic Obstruction



- sound sources in virtual scene (OpenGL application)
- distance dependent attenuation
- frequency dependent attenuation behind screens
- implemented with HW-extensions EAX (Creative Audigy®)

Demonstration: A cappella ensemble



- 3D MPEG-4 player, mp4-file (incl. graphics and sound)
- 3D audio rendering in real-time: PureData (Pd)
- frequency dependent source directivity can be experienced when walking through the scene

Thank you for your attention!