

MPEG-4 Audio

“A Preview into the Technology of the Future”



Outline

- MPEG-4 Workplan
- MPEG-4 Audio Objectives
- MPEG-4 Audio Components
- MPEG-4 Audio Components Features
- MPEG-4 Audio Implementation Complexity
- MPEG-4 Audio Profiles
- MPEG-4 Audio / MPEG-4 Systems issues
- MPEG-4 Audio Applications



MPEG-4 Workplan

Version 1

- Final Draft Standard: 12/1998
- Final Standard: 02/2000 ?
- More information on subsequent transparencies

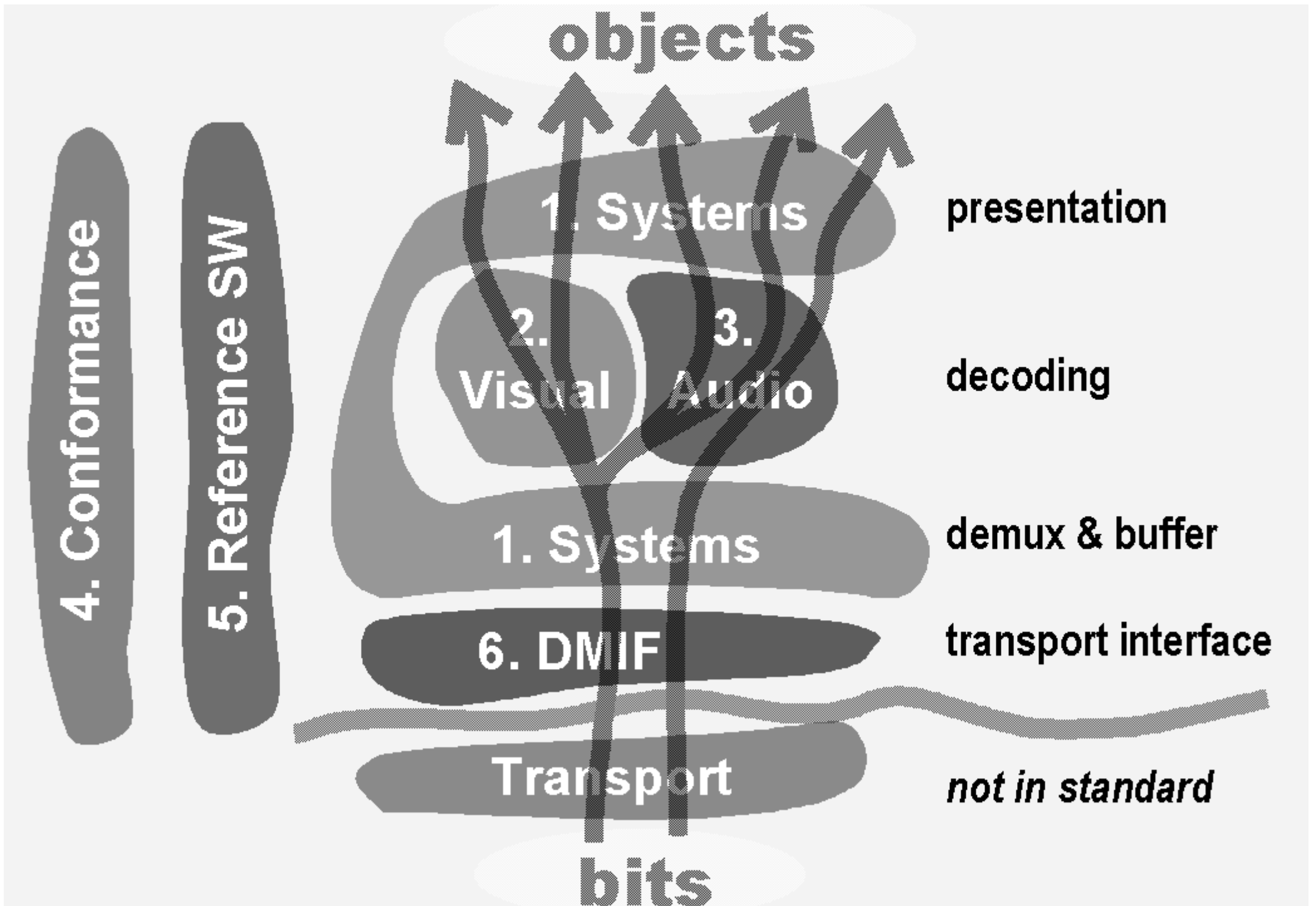
Version 2

- Final Draft Standard 12/1999
- No delta improvements to techniques already in Version 1
- Additional tools for functionalities not yet covered by Version 1

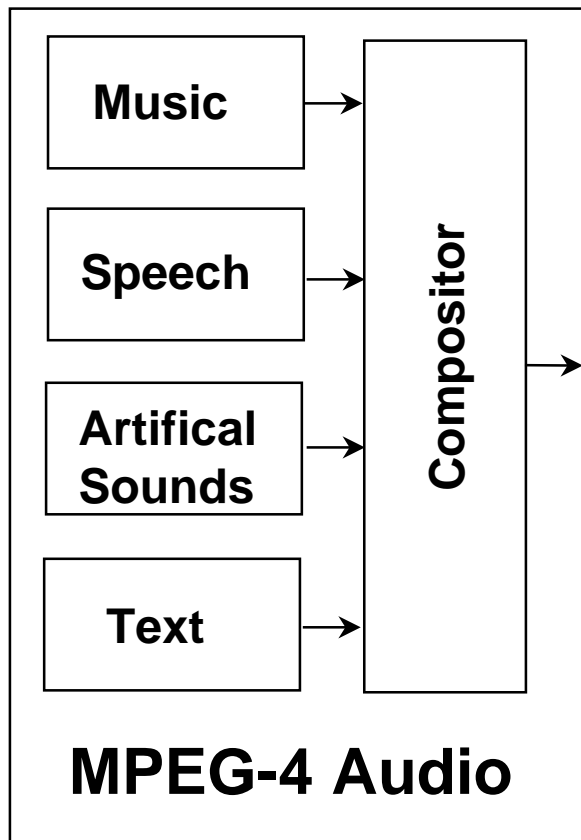
Future:

- Merge of Version 1 and Version 2 ?



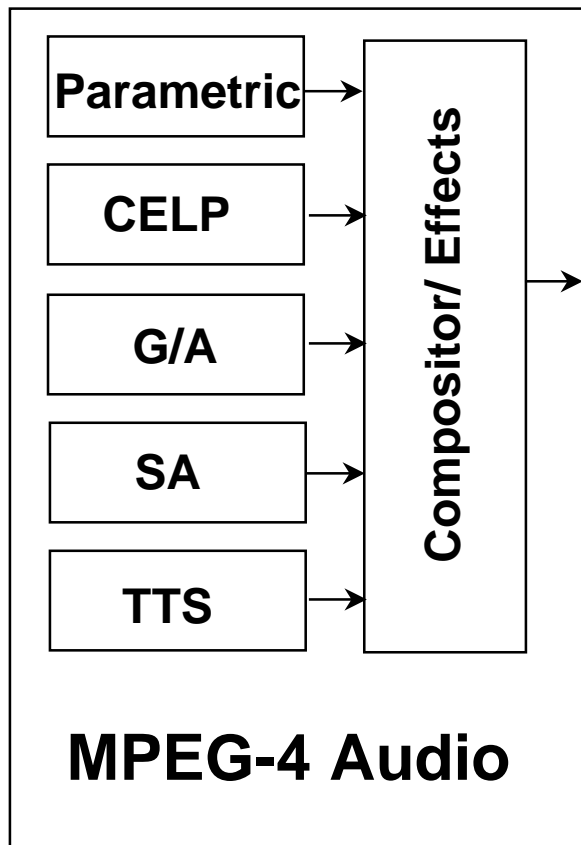


Main Objectives of MPEG-4 Audio



- Continuation of MPEG-1 and MPEG-2 Audio Coding
 - Further improved coding methods for natural audio source material
- New Areas
 - Speech Coding
 - Coding of synthetic audio material
 - Text to Speech
- Additional Functionalities
 - Bit Rate Scalability (Layered Coding)
 - Speed- and Pitch-Change
 - 2-d and 3-d Composition
- Interoperability
 - MPEG-4 Audio for all audio coding applications

MPEG-4 Audio Components

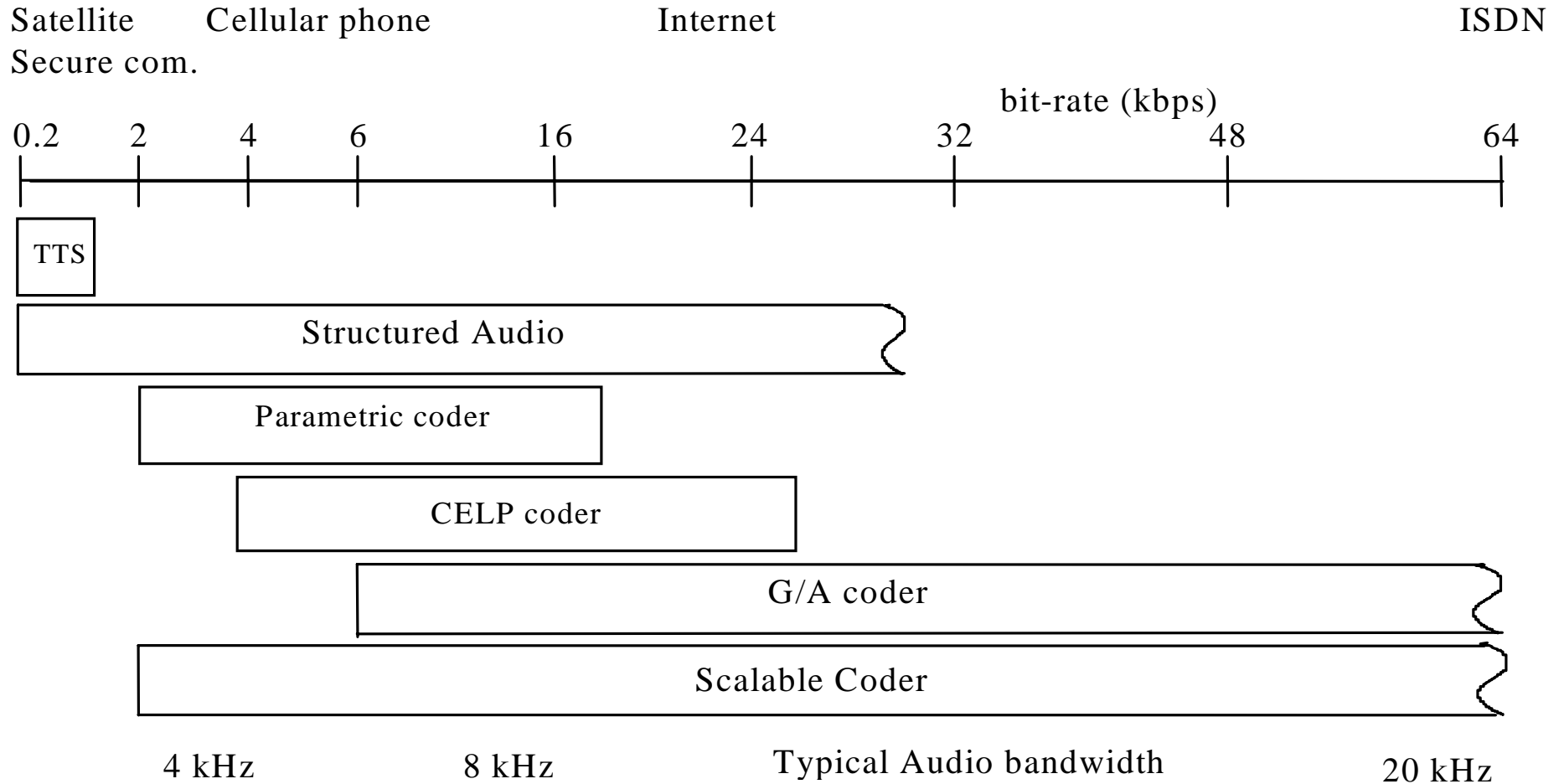


- Contains a Set of different Coders for different Signal Classes and Bit Rate Ranges
 - Parametric Speech and Audio Coder
 - CELP Speech Coder
 - General Audio (G/A) Coder
 - Combined scalable Coder
- Methods for Synthetic Sound Material
 - Structured Audio System
 - Text to Speech Interface
- Sound Effects and Composition
 - Speed Change Tool
 - Audio Compositor
 - Sound Effects

gri 6/98



MPEG-4 Audio Overview "A Preview into the Technology of the Future"



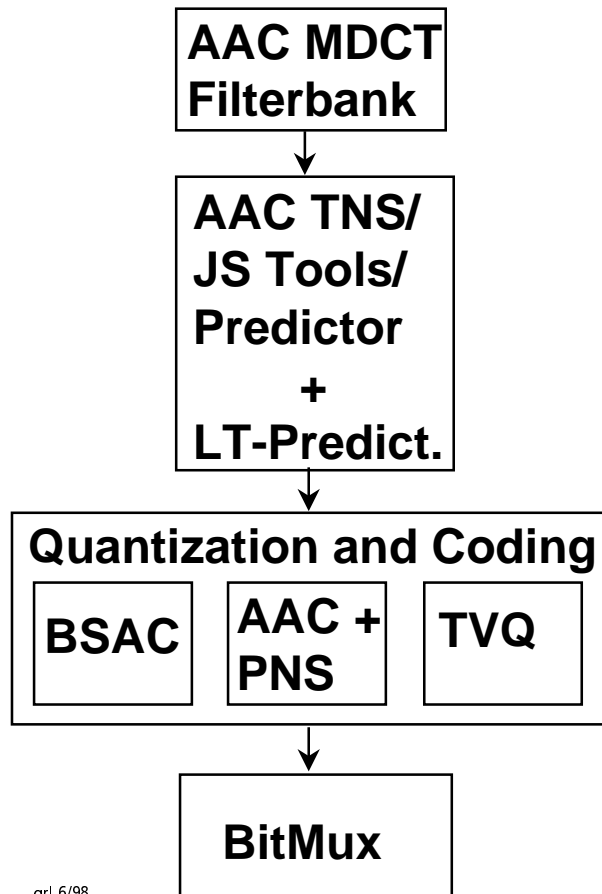
MPEG-4 Audio Components: Parametric Speech Coder

- Bit Rate Range: 2 - 4 kbit/s
- Sampling Rate: 8 kHz
- Coding/Decoding Delay: < 40 ms
- Layered scalable coding option
- Speed- and Pitch-Change functionality is an inherent functionality of this coder
- Excellent quality for speech only applications and speech with background noise
- Not recommended for general audio signals

MPEG-4 Audio Components: CELP Speech Coder

- Narrow-band and Wide-band mode
- Sampling Rates: 8 and 16 kHz (44.1/6 = 7.35 kHz possible)
- Bit Rate Range: 4 - 24 kbit/s
- Coding/Decoding Delay depending on Bit Rate:
 - max. for lowest bit rates < 40 ms (4kbit/s)
 - min. for higher bit rates ~ 15 ms
- Layered scalable coding option
- Best for speech only applications and speech with background noise at low bit rates
- Some stability for general audio signals

MPEG-4 Audio Components: G/A Coder



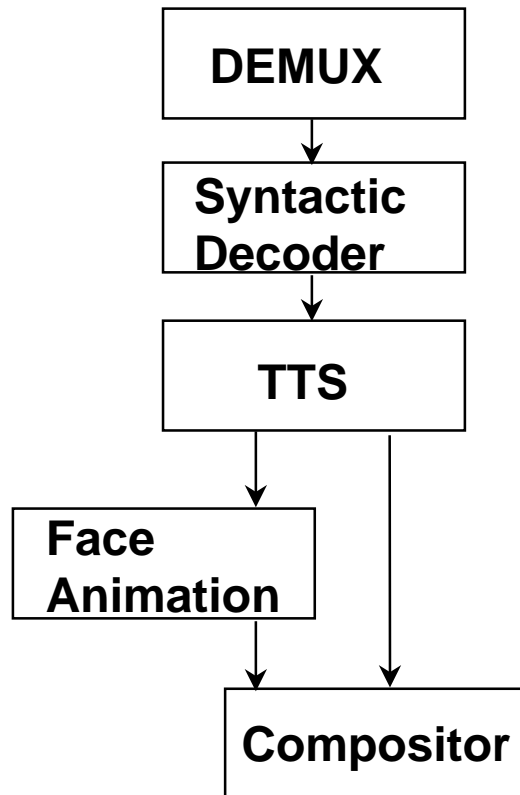
- "Classic" Type of an MPEG Coder
- Includes MPEG-2 AAC with additional MPEG-4 features
 - Perceptual Noise Substitution
 - Low-Complexity Long Term Predictor
 - Bit rate scalability
- Twin-VQ low bit rate coder (6 and 8 kbit/s)
 - combined scalable modes with AAC
- bit rates: 6 - 300 kbit/s
- sampling rates: 7.35 - 96 kHz
- At bit rates > 16 kbit/s best audio quality for all types of audio material
- Twin-VQ best for complex audio signals at 6 or 8 kbit/s

Structured Audio

The MPEG-4 Component for synthetic Sounds

- **SAOL** **Structured Audio Orchestra Language**
 - Description of sound generation algorithms
- **SASBF** **Structured Audio Sample Bank Format**
 - Download and usage information for wave table synthesis samples and control data
- **SASL** **Structured Audio Score Language**
 - Used to describe the manner in which the synthesis algorithms described by SAOL are used to produce sounds
- **MIDI**
 - Less powerful alternative to SASL which is integrated to provide compatibility with existing content

Text To Speech (TTS) Interface



- Only the interfaces are standardized
 - Actual text to speech conversion is not defined in MPEG-4
 - Only a normative way of how to implement interfaces to a TTS system is defined
- Along with the text parameters can be transmitted in a normative way which can control the synthesis if the actually used TTS system supports this
 - speaker age
 - speaker gender
 - ...
- Interface to the MPEG-4 face animation technology

Implementation Complexity

Complexity in Processing Complexity Units (PCU) and RAM Complexity Units (RCU)	Coder	Parameters		
	RCU			PCU
	GA / AAC LC	fs = 48 kHz	3	3
	G/A LC Scalable	fs = 48 kHz	5	3
	TwinVQ	fs = 24 kHz	2	3
	CELP	fs = 8 kHz	1	1
	CELP	fs = 16 kHz	2	1
	HVXC	fs = 8 kHz	2	1

MPEG-4 Audio Profiles

Definition:

- Sub-sets of MPEG-4 Tools
 - Conformance specified for profiles
 - Profiles may have levels which differ in the number of supported audio channels and number of audio objects and in the maximum allowed sampling rate

4 Profiles

- Main
 - All MPEG-4 Audio components including full SA
- Scalable
 - As main profile, but the MPEG-2 “AAC Main” and “AAC SSR” profiles are excluded
- Speech
 - CELP, Parametric Speech, TTSI
- Wavetable synthesis

MPEG-4 Audio Multiplex

- No bit stream like in MPEG-1 or MPEG-2
- ==> No multiplex, no syncword, ... in MPEG-4 Audio
- MPEG-4 audio only defines setup information packets and payload packets for each coder
- MPEG-4 Systems specifies “Flex-Mux” to cover multiplex aspects of special MPEG-4 functionalities (e.g. scalability)
- However, no actual multiplex defined for the packets in MPEG-4 (Only a file format will be included in MPEG-4 version 2)
- TransMux Options
 - MPEG-2 Systems
 - RTP
 - MPEG-4 File Format (.mp4)

MPEG-4 Audio Compositor

- Audio Compositor is part of the MPEG-4 Systems standard
- Structured Audio Components in MPEG-4 Systems
- Functionality
 - 2-d and 3-d composition/mixing of a number of audio channels
 - An audio scene can be composed with several audio sources by specifying their 2-d or 3-d position
 - simple sampling rate conversion
 - simple effects (e.g. reverb)
 - pitch/speed change

MPEG-4 Applications

Everything!!

Internet

Broadcast

**2-Way
Communication**

Games

- Most of the applications which require low bit rate audio (speech) coding should be possible with MPEG-4 Audio
- Audio transmissions on packet-oriented computer networks (Internet)
- Broadcast
 - Scalable coding for graceful degradation by using unequal error protection for the different layers
- Two-way communication (Proposed for Mobile Phones)
- Computer games
 - Artificial environments can be created with the compositor
 - Coding methods for artificial sounds
 - Speed-/pitch-change to adapt audio sequences to video sequences