

MPEG-4

Scalable Audio Coding

Karlheinz Brandenburg
Department Head, Multimedia Department
Fraunhofer Institute Integrated Circuits
Erlangen, Germany



Overview

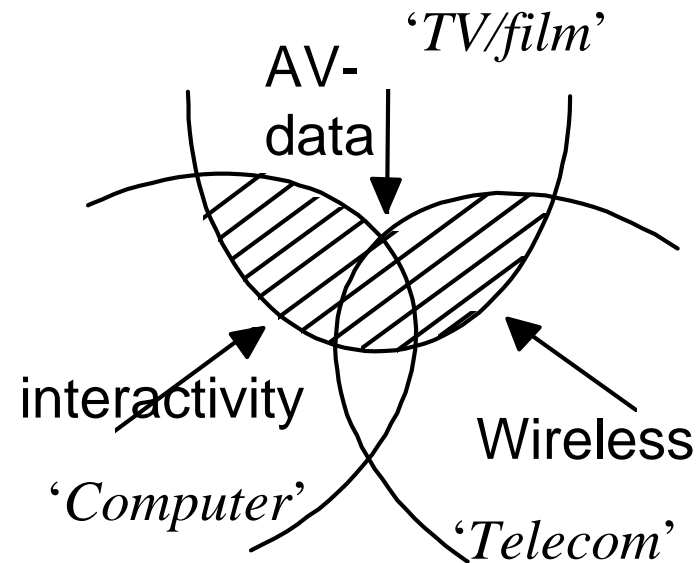
- Introduction: MPEG-4 Audio
- MPEG-4 Scalable Profile
- Speech Coders
- General Audio Coding
- Scalability
- Application Scenarios
- Demos



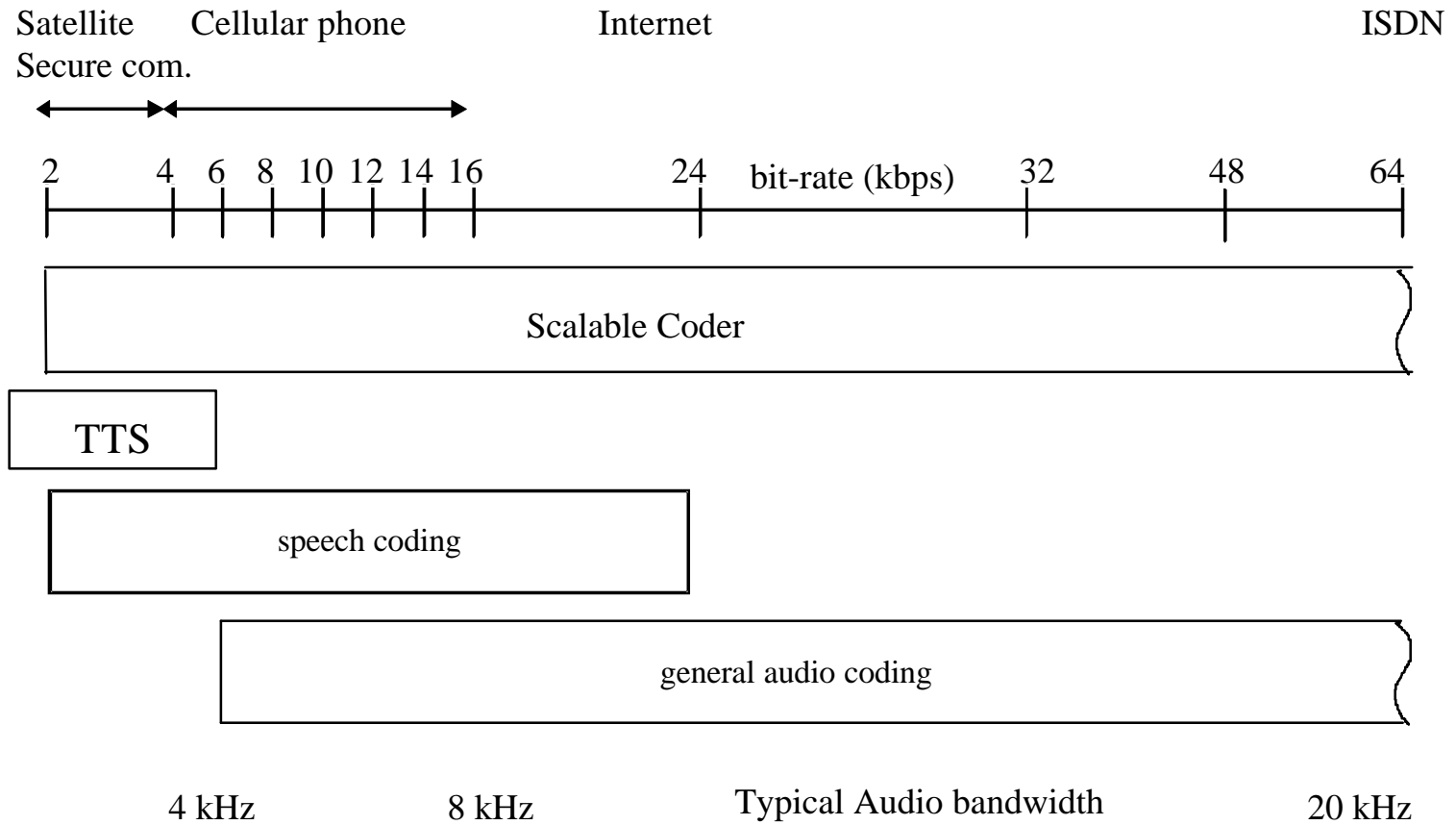
MPEG-4 Audio

- Interactivity
- High compression
- Universal accessibility

MPEG-4 addresses applications in the shaded area.



MPEG-4 Audio Algorithms



MPEG-4 Audio Profiles

- Speech Audio Profile
 - Parametric speech coder (HVXC)
 - CELP speech coder
- Synthesis Audio Profile
 - generate speech and sound
- Scalable Audio Profile
 - contains the speech audio profile
 - General audio coding (AAC)
 - TwinVQ tools
 - Scalable coding of speech and music
- Main Audio Profile
 - contains all other profiles



The Scalable Audio Profile

- Four levels defined according to
 - sampling frequency
 - number of channels / objects
- Objects in the scalable audio profile
 - AAC LC
 - AAC LTP
 - AAC Scalable
 - TwinVQ
 - CELP
 - HVXC
 - TTSI



Speech Coding

- HVXC
 - Bit-rates typically 2 kbps - 4 kbps
 - Parametric speech coding:
High quality for coding of clean speech
 - Speed change / pitch change capability
- CELP
 - Bit-rates typically 6 kbps - 24 kbps
 - Very flexible configuration possibilities
 - Support for 8 kHz and 16 kHz sampling



General Audio Coding

- AAC
 - MPEG-2 AAC contained in MPEG-4
 - State of the art high quality coding
 - additional tools to give higher quality at low bit-rates
 - up to five channels
 - scalable version can be combined with AAC, CELP, Twin-VQ
- TwinVQ
 - best for very low bit-rates
 - integrated with AAC stereo coding tools



Scalable Audio Coding

- Embed lower quality (e.g. lower bandwidth) bitstream in higher bandwidth bitstream
- Key functionality for MPEG-4 audio
- Main types of scalability:
 - Small step scalability
Enhancement layers of ~ 1 kbit/s
 - Large step scalability
Enhancement layers of 8 kbit/s and more
- All natural audio coding in MPEG-4 supports scalability



Application Scenarios

- Internet audio streaming
 - single instance encoding
 - get best quality decoding dependent on
 - type of connection (modem / ISDN ...)
 - network congestion
- Digital broadcasting (e.g. DRM, digital broadcast for the short and medium wave bands)
- Data-base access via different bandwidth connections
 - single encode is good for
 - prelistening
 - high quality audio delivery



Demo 1: Scalable Audio Coding

- One bitstream, different decoding
- Highest quality either ISDN or modem bandwidth
- Mono / stereo scalability



Demo 2: Can you hear it (version 3, 1998) ?

Each ? is either

O (original, 1536 kbit/s per stereo signal) or

C (coded, 64 kbps per stereo signal)

Trumpet solo	?	?	?
Speech	?	?	?
Tracy Chapman	?	?	?

