What is MPEG-4 audio
and what can I do with it?

Schuyler Quackenbush
AT&T Labs - Research
Florham Park, NJ 07932 USA
srq@research.att.com
Workshop Outline

- Overview
- Speech Coding
  - Schuyler Quackenbush, AT&T Labs
- General Audio Coding
  - Jürgen Herre, FhG IIS
- Scalable Audio Coding
  - Bernhard Grill, FhG IIS
- Structured Audio Coding
  - Lee Ray, Creative Labs
  - Eric Scheirer, MIT Media Lab
  - Schuyler Quackenbush, AT&T
- Audio Composition
  - Jyri Huopaniemi, Nokia Research Center
The MPEG Family

• Latest in family of ISO/MPEG standards
  – MPEG-1 1992 (ISO 11172)  Digital Audio/Video
  – MPEG-2 AAC 1997  Multi-channel Audio
  – MPEG-4 1999 (ISO 14496)  Low-rate coding  
    Object-based coding

• MPEG-2: bit stream and decoder
• MPEG-4: object and associated operations
Media Objects

- Objects
  - Natural audio
  - Synthetic audio
  - Control
- Operations on objects
  - Synchronize
  - Decode
  - Compose into compound objects
  - Present
  - Interact
Advantages of Object Framework

• Each signal coded with most efficient coding system
  – Natural
  – Synthetic

• Composition of objects into audio scene
  – Rate conversion
  – Mix and Eq
  – Effects

• Final mix is done in the terminal
System Overview

Demux → Synch → Decode → Compose → Present

- Obj 1 → Dec 1
- Obj 2 → Dec 2
- Description
- IPR Mgmt Interf.
- IPR Control

Audio Scene Graph → Compound Object → Listener

Interaction
Audio Object Functionalities

• Signal compression
• Scalability
  – bit rate
  – signal bandwidth
  – presentation rate
  – encoder or decoder complexity
• Extraction and re-use
• Robustness to channel errors
Scalability

Satellite  Cellular phone  Internet  ISDN
Secure com.

2  4  6  8  10  12  14  16  24  bit-rate (kbps)  32  48  64

Scalable Coder

TTS

Speech coding

General audio coding

4 kHz  8 kHz  Typical Audio bandwidth  20 kHz

May 10, 1999  106th AES Convention
Application Domains: Profiles

• Speech
  – low rate speech coders and TTS

• Synthesis
  – wavetable synthesis
  – score driven synthesis
  – TTS

• Scalable
  – speech coders
  – general audio coders
  – all coders in scalable configuration

• Main
  – all of the above
Summary

- MPEG-4 is a powerful, flexible, object-based coding standard
  - multiple natural audio coders
  - structured audio coding
    - wavetable synthesis
    - score driven algorithmic synthesis
  - audio composition
    - user interaction
- Enhanced functionality
  - state of the art compression
  - bit rate scalability
- Re-usability of audio objects
Future

- MPEG-4 will be the standard for complex multi-media presentations
- Content composed in MPEG-4
- Powerful authoring tools