

Natural Audio Tools in MPEG-4

Schuyler Quackenbush

AT&T Labs - Research

Florham Park, NJ 07932 USA

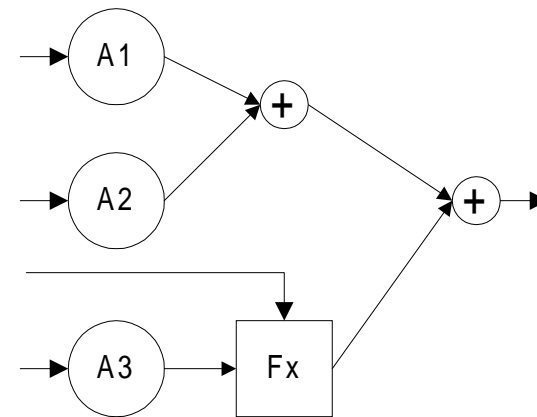
srq@research.att.com

Outline

- Introduction and Overview
- Speech Coding
- General Audio Coding
 - Jurgen Herre, FhG-IIS
- Scalable Audio Coding
 - Bernhard Grill, FhG IIS

Media Objects and Associated Operations

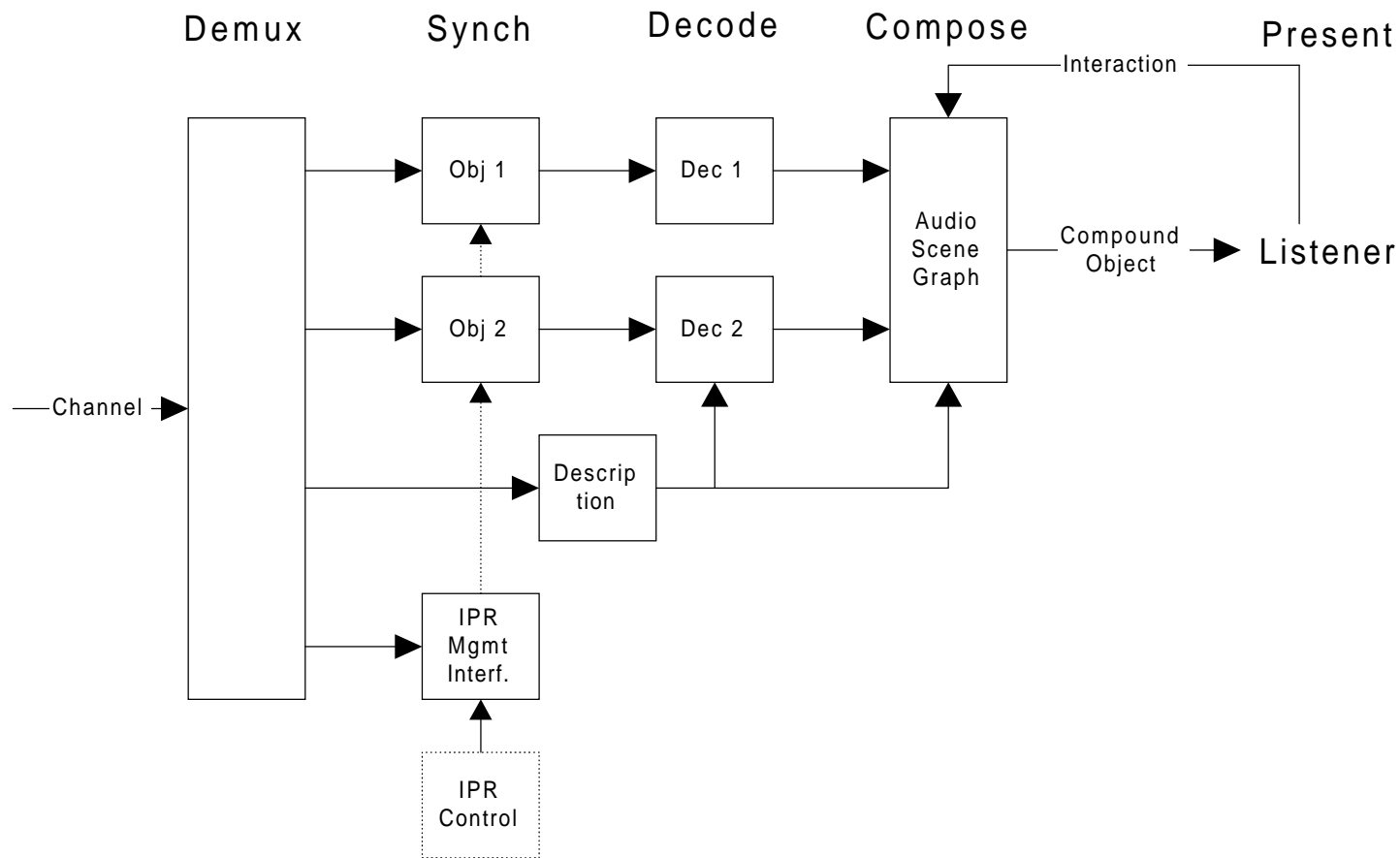
- Objects
 - Natural audio
 - Synthetic audio
 - Control
- Operations on objects
 - Synchronize
 - Decode
 - Compose into compound objects
 - Present
 - Interact



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

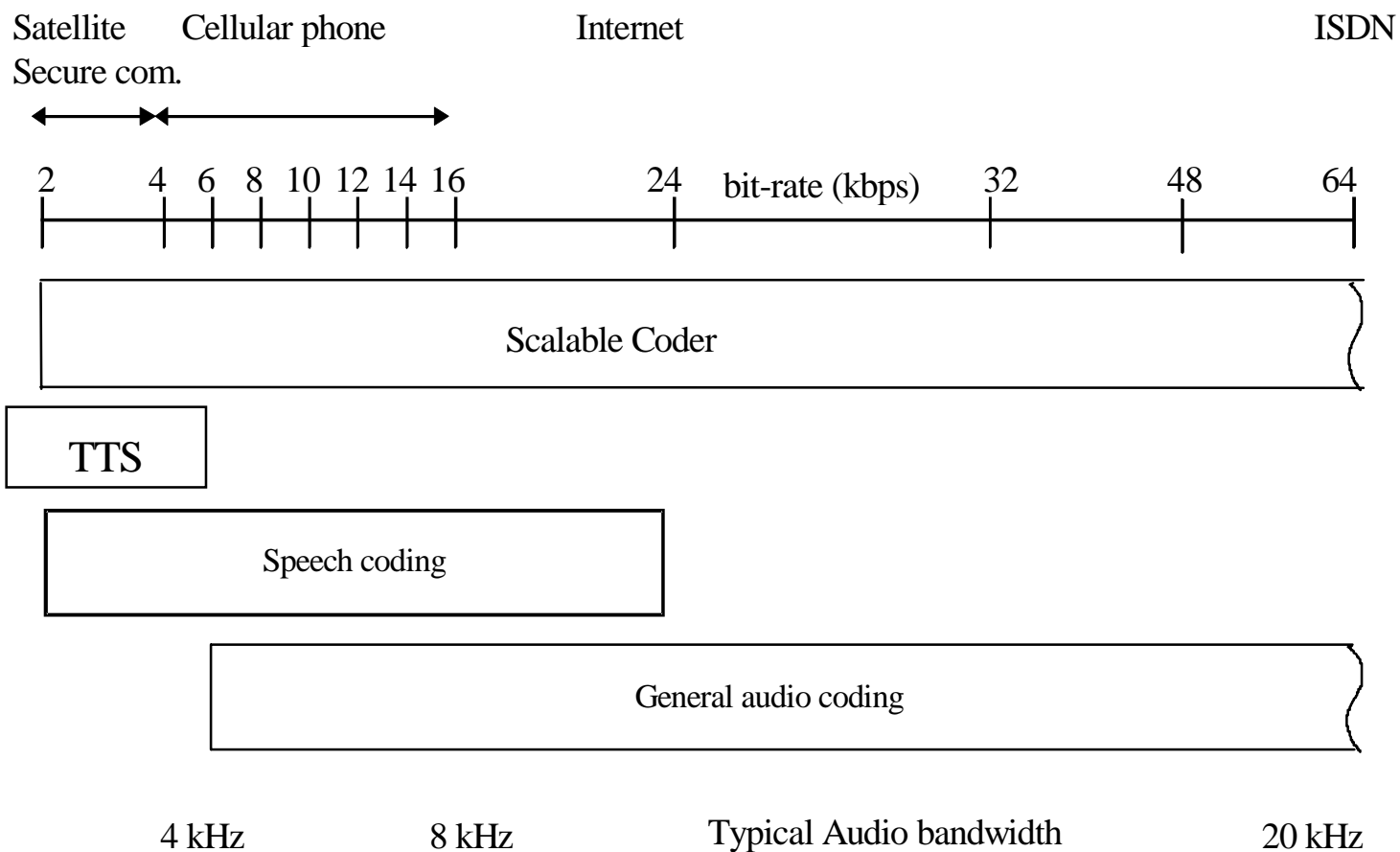
System Overview



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



Application Domains Version 1 Profiles

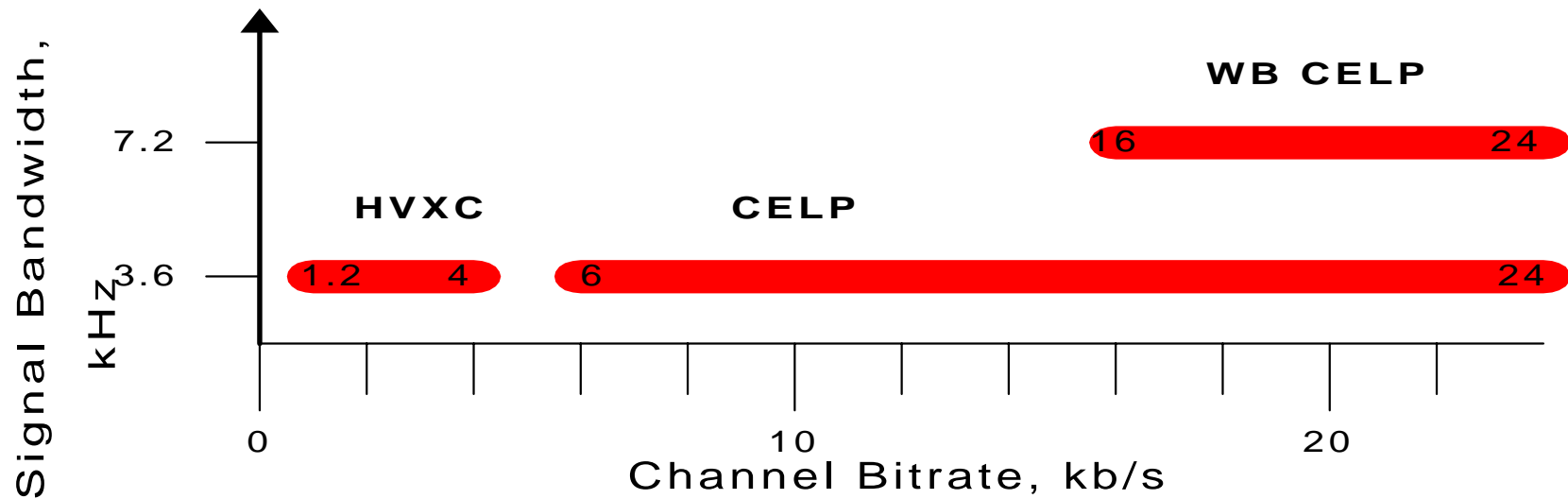
- Speech
 - low rate speech coders and TTS
- Scalable
 - speech coders
 - general audio coders
 - all coders in scalable configuration
- Synthetic
 - wavetable synthesis
 - score driven synthesis
 - TTS
- Main

MPEG-4 Speech Coding: Overview

- Excellent compression by using source model
 - Linear Predictive Coding (LPC)
 - Pitch or noise excitation
- Better compression than “general audio” coders
 - only for “clean speech” from single talker

Speech Coders

- Harmonic Vector Excitation Coder (HVXC)
- Code Excitation Linear Prediction (CELP)
- Wideband CELP



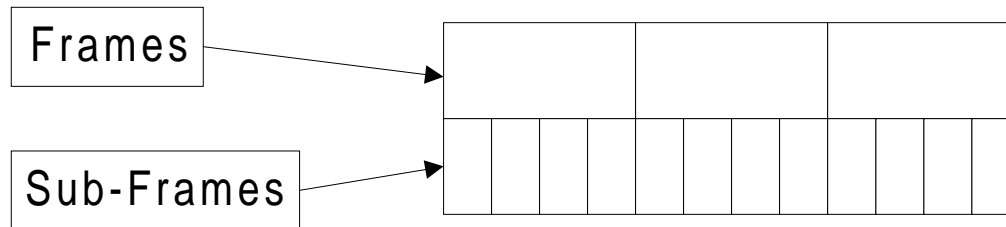
Communication Characteristics

- Low bit rate
 - HVXC 1.2 kb/s to 1.7 kb/s var. rate
2.0 kb/s to 4.0 kb/s const. rate
 - CELP 4.0 kb/s to 24 kb/s const. Rate
- Low one-way delay
 - HVXC 33.5 ms to 56 ms
 - CELP 15 ms to 45 ms
- Not compromised for modem signals

Bit Rate Scalability



- Parameters coded using multi-stage VQ
 - base plus enhancement layer
- Enhancement layers can be stripped in
 - server
 - channel
 - decoder

Parameter Update Scalability



- Linear Prediction Model
 - updated every frame
 - interpolated every sub-frame
- Excitation
 - gain updated every subframe

Demonstration

- 1.7 kb/s variable rate HVXC 
- 12 kb/s CELP 
- 24 kb/s Wideband CELP 