MPEG-4 Synthetic Audio Tools

Riitta Väänänen
Helsinki University of Technology
Laboratory of Acoustics and
Audio Signal Processing
http://www.acoustics.hut.fi/
riitta.vaananen@hut.fi

Outline

- Synthetic Audio Tools in MPEG-4 version 1 and version 2
- Tools in MPEG-4 Audio
- Audio Tools in MPEG-4 Systems (AudioBIFS)
- 3-D Audio Spatialization in v2 Systems (AudioBIFS)

Synthetic Audio in MPEG-4 Version 1

- MPEG-4 Audio:
  - Structured Audio (SA)
  - Text-To-Speech interface (TTS)
- MPEG-4 Systems:
  - Audio scene description (AudioBIFS)
  - VRML 3-D Sound source model

Synthetic Audio in MPEG-4 Version 2

- Advanced audio scene description in BIFS (Advanced AudioBIFS)
  - 3-D spatialization of sound according to the geometrical description of sources and the acoustic response of the visual environment
  - 3-D spatialization of sound according to perceptual description of acoustic response of the environment

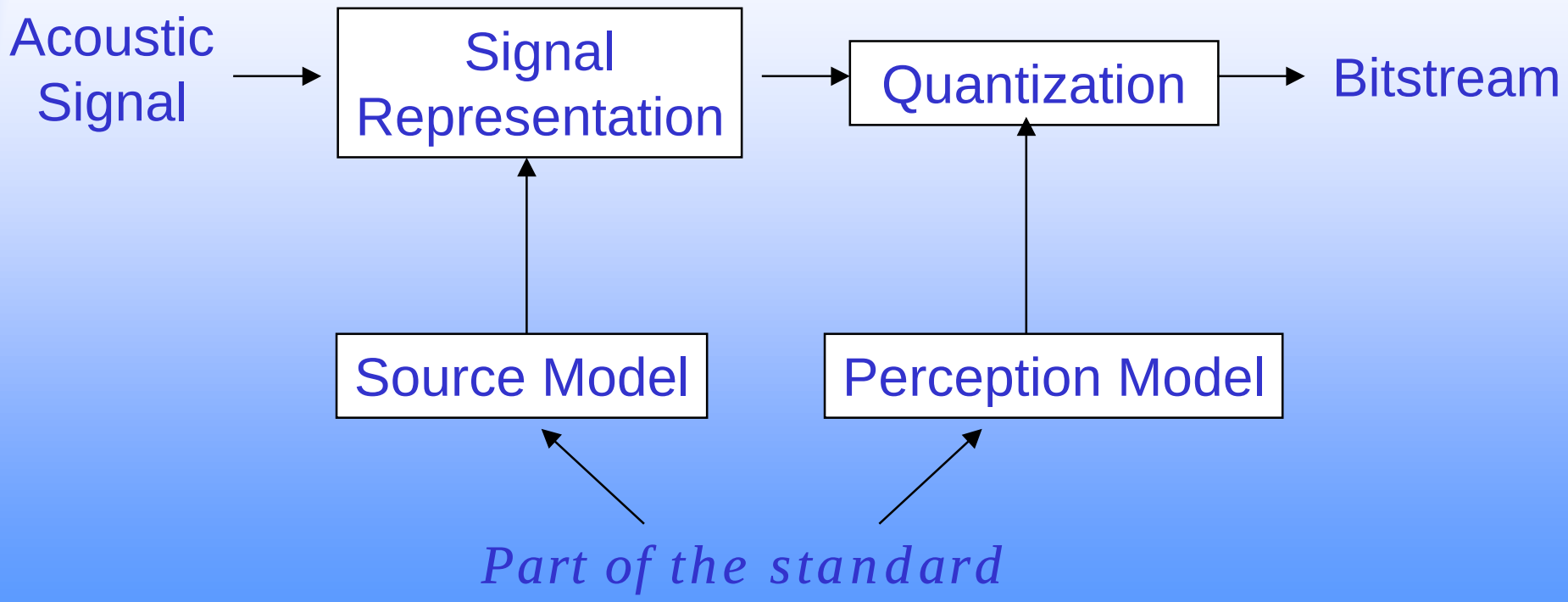
Goals for Synthetic Audio in MPEG-4

- SA: Delivery of audio at very low bit-rate
- AudioBIFS: Composition of Audio at MPEG-4 terminal, 3-D presentation of sound sources
- TTS: Low bitrate delivery of synthetic speech synchronization with face animation
- Advanced AudioBIFS (v.2): Modeling of acoustic environment

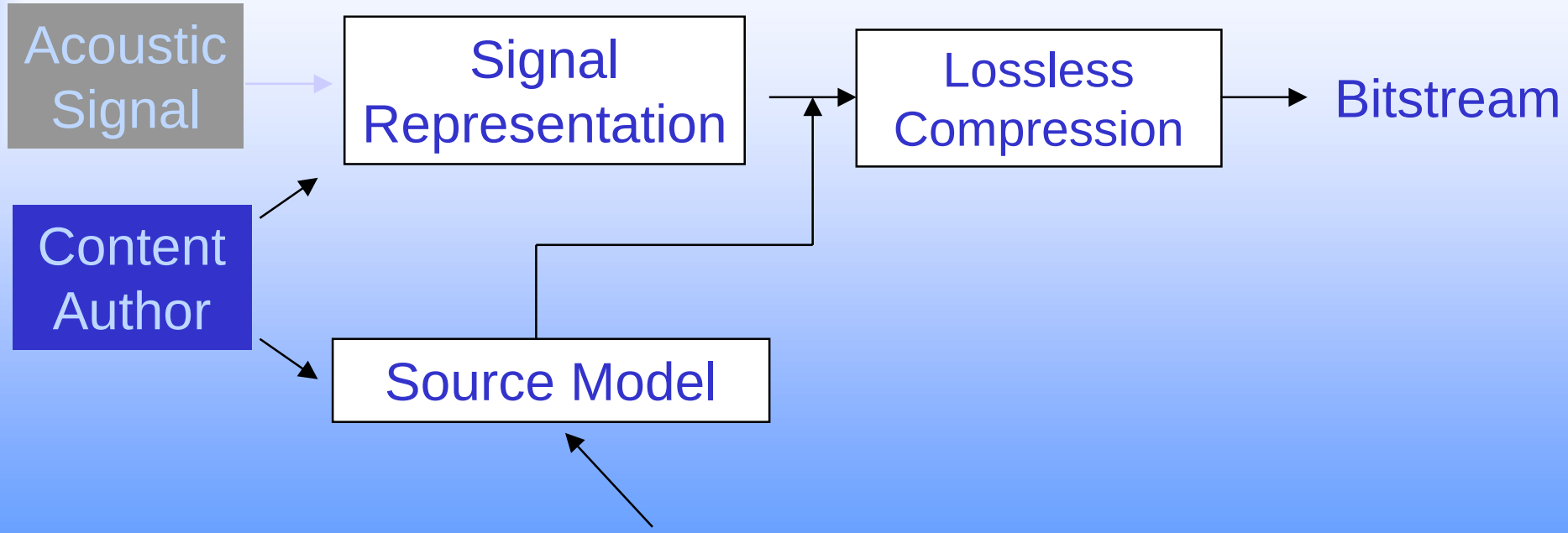
Structured Audio in MPEG-4

- Structured Audio =
  Sound coding using structured descriptions
- Structured Audio decoder:
  music and sound-effect synthesis
- Tools:
  - SAOL : description of synthesis methods
  - SASBF : description of wavetable data
  - SASL, MIDI: description of control parameters

Basic Audio Coding



Structured Audio Coding



SAOL: Structured Audio Orchestra Language

- New format for describing audio synthesis algorithms
- No “synthesis method” is standardized
  - Method for describing synthesis methods
- Based on years of computer-music research (new concept in audio coding)
- Also used in Audio BIFS

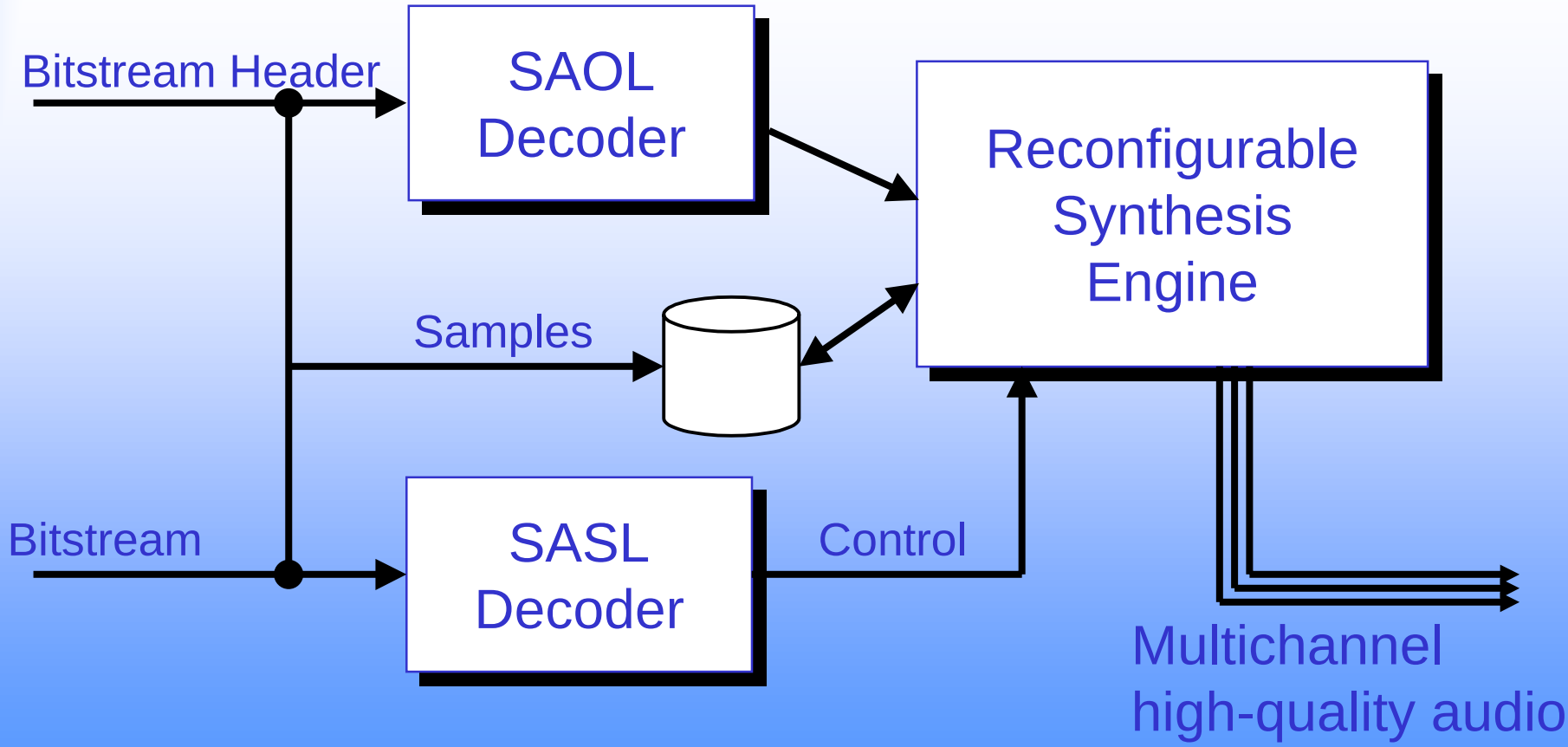
SASBF: Structured Audio Sample Bank Format

- New format for efficiently describing banks of wavetable data
- Wavetable (sampling) synthesis is simple, commonly used today
- Intended for limited-function terminals (e.g. karaoke systems) in MPEG-4

SASL: Structured Audio Score Language

- New format for describing control parameters
- Designed to interface well with SAOL
- MIDI (Musical Instrument Digital Interface)
  - Simpler format for describing control
  - Included as alternate control method
  - Leverages existing authoring tools

Decoding Process



Text-To-Speech in MPEG-4

- Phonemic (language-independent) syntax
- Prosody, timing cues
- Language, dialect, gender, age parameters
- Automatic synchronization with FBA
- Exact TTS synthesis non-normative;
  only interface is specified

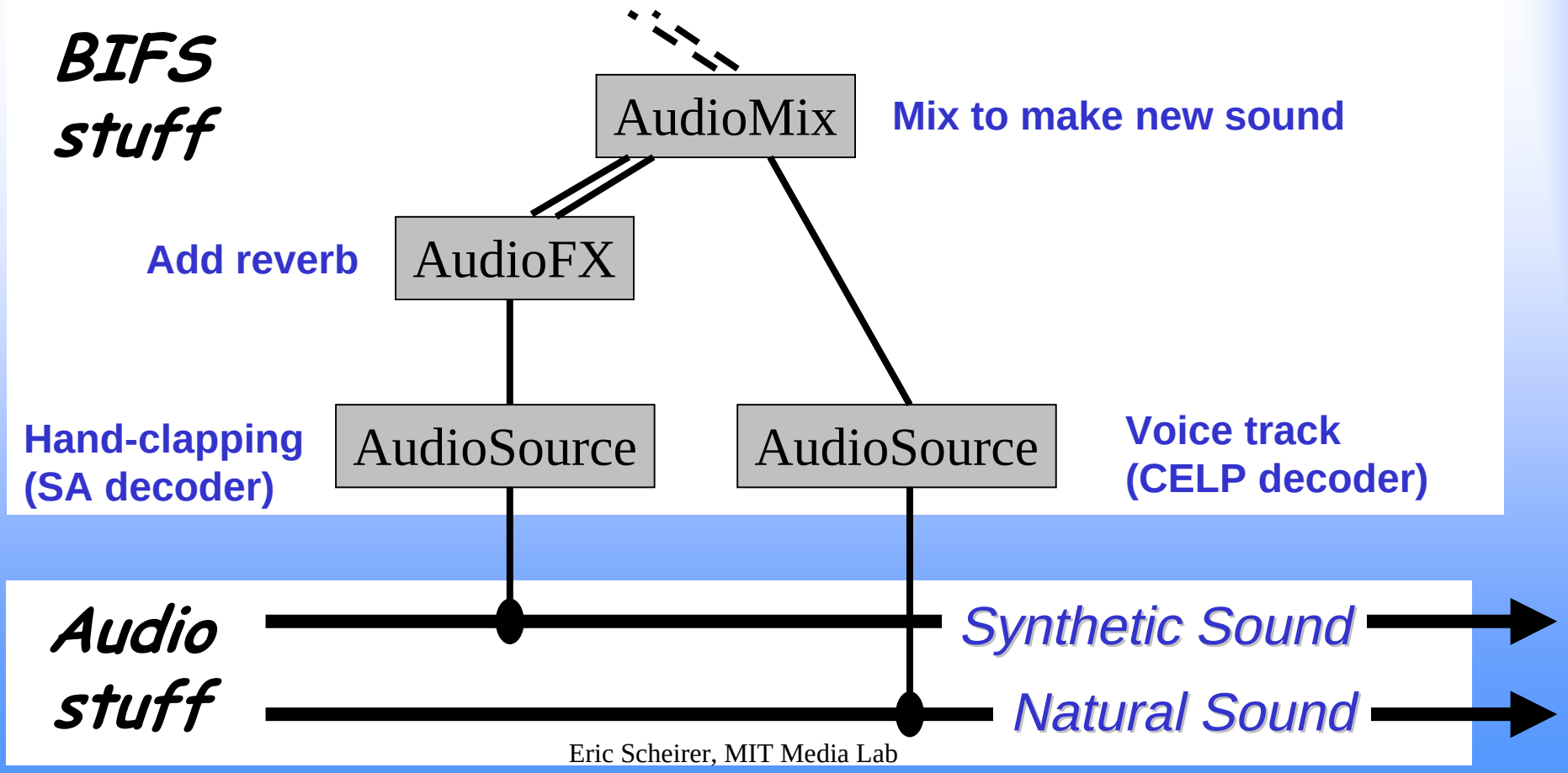
Audio Scene Description

- Code different parts of a soundtrack in different ways
- Mix, produce, spatialize sounds at terminal
- AudioBIFS: description of mixing, effects, post-production

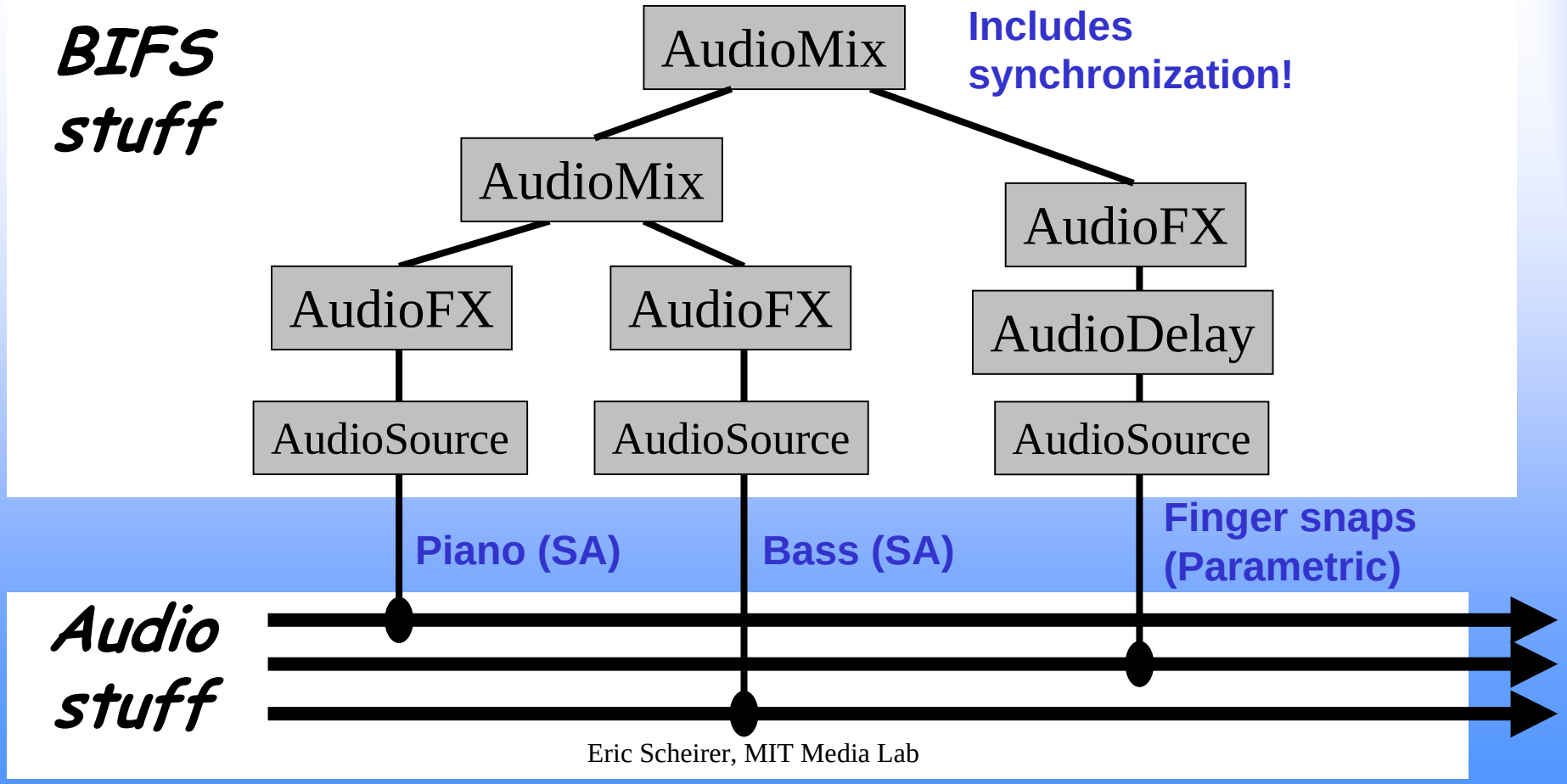
AudioBIFS Capabilities

- Mix and synchronize multichannel sound
- Use SAOL to describe effects (AudioFX)
- Position sounds in virtual 3-D space

Structured Audio Scene



Structured Audio Scene



Demo of Structured Audio

- Manipulator by Michael Casey
- Header size = 16 KB + 100 KB of samples
- Bitrate = 0.13 Kbps
- Sounds the same on every terminal
- 90 seconds of CD-quality stereo sound

Demo on Soundtrack Composed of Different Streams

- Demo: 16 kbps MPEG-4 soundtrack
  - natural and synthetic speech
  - synthetic music and sound effects
  - mixing, reverb, etc in BIFS with SAOL

Sound Source Model in BIFS (v1)

- Same as in VRML97:
- Enables:
  - Positioning of sound sources in a 3-D space
  - Simple sound soure directivity and attenuation modeling
  - Two nested ellipsoids form the region of audibility in the scene
- Demo: Single VRML sound source

Advanced AudioBIFS (v2)

- Adds more advanced 3-D sound sourcce and acoustic environment modeling
- New features include:
  - Directive sound sources
  - Propagation of sound in the medium
  - Sound reflections or transmission of sound through objects
  - Reverberation
  - Room acoustic response based on perceptual parameters

Sound Source Radiation

- Frequency dependent source directivity
  - For modeling sound source directivity patterns of e.g., a human speaker, musical instruments, loudspeakers
  - Scene author can affect the complexity (detailness) of directivity modeling
- **Demo**: rotating directive sound source

3-D Direction of Arrival

- Reproduction not specified, therefore 3-D spatialization method not part of the standard:
  - Spatialization enabled by a flag in the sound source node
  - Decoder uses the best possible quality of spatialization and suitable to the loudspeaker configuration

Sound Propagation in the Medium

- Distance dependent attenuation
- Distance dependent lowpass filtering caused by air absorption
- Taking into account the propagation delay between the source and the listener:
  - Depends on the speed of sound in the medium and the distance between the source and the listener
  - Used to create Doppler effect which depends on the relative motion between the soure and the listener
  - **Demos**: Influence of speed of sound in the Doppler effect

Sound reflections and Transmission Through Objects

- Parameters can be attached to flat polygonal (visual) surfaces that give reflective and transmission properties
- Enables detailed modeling of (specular) sound reflections and echoes taking into account their delays and directions when they arrive at the listener
- Enables inclusion of sound obstructive (attenuating) objects in a BIFS scene
- **Demo**: Single reflective and sound passing surface

Room Acoustics Modeling in Advanced AudioBIFS

- Physical approach:
  - Early room reflections according to geometrical description of acoustic surfaces
  - Late reverberation added according to reverb. time definition
  - All spatial audio processing in a specified 3D region in the scene -> allows creation of several acoustic "rooms"
- Perceptual approach:
  - Room acoustic effect created by definition of perceptual room acoustic parameters

- Demos on room acoustics modeling:
  - Small room with reflective & sound passing walls
  - Same room with small speed of sound
  - Large room with long reverberation time
  - Scene with 2 rooms & 3 sound sources
  - Scene with 2 rooms and one source

Conclusions

- Synthetic audio tools in MPEG-4:
  - Structured Audio: For low bitrate delivery of sound
  - AudioBIFS: For composition and spatial positioning of sound sources
  - Advanced AudioBIFS: For immersive sound environment modeling
- Thanx to: Eric Scheirer (MIT Media Laboratory) for the SA demos, slides on SA and AudioBIFS; More info at:
  http://sound.media.mit.edu/~eds/mpeg4/
  email:eds@media.mit.edu