Audio Coding Standards

Kari Pihkala

13.2.2002

Tik-111.590 Multimedia Communications

Outline

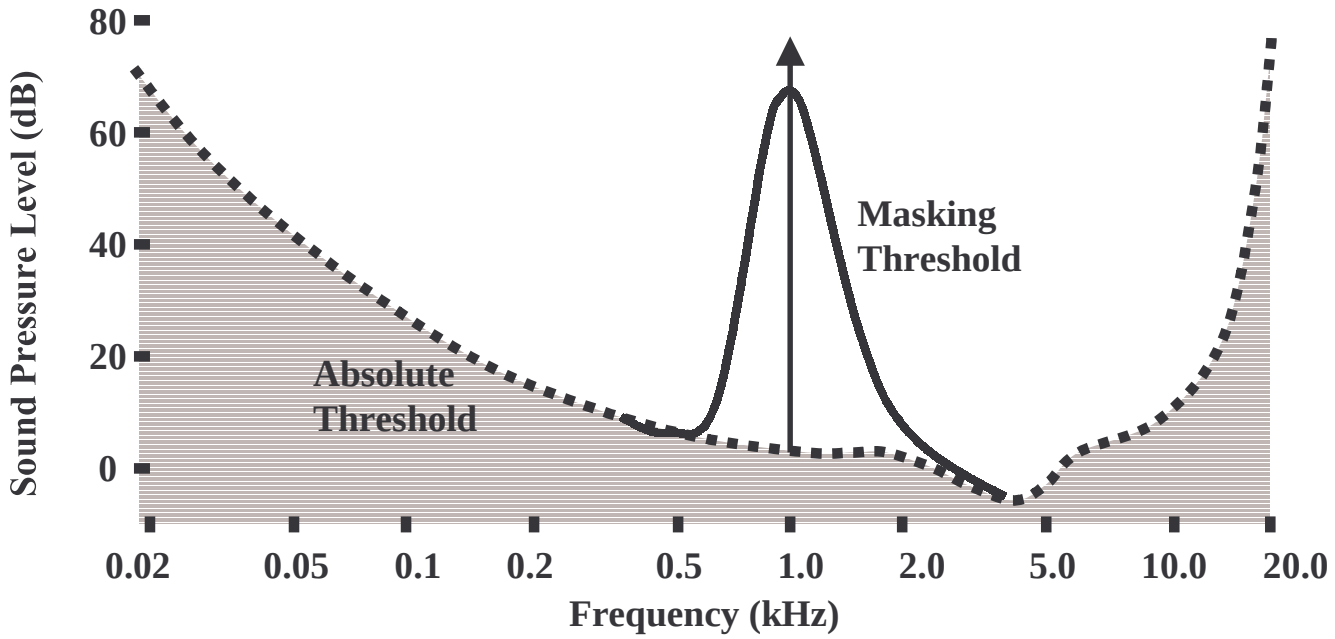
- **Architectural Overview**
- **MPEG-1**
- **MPEG-2**
- **MPEG-4**
- **Philips PASC (DCC cassette)**
- **Sony ATRAC (MiniDisc)**
- **Dolby AC-3**
- **Conclusions**

Architectural Overview

- **Psychoacoustic Modeling**
- **Time-Frequency Mapping**
- **Quantization**
- **Variable-Length Coding**
- **Multichannel Correlation and Irrelevancy**
- **Long-Term Correlation**
- **Pre-echo Control**
- **Bit Allocation**

Psychoacoustic Modeling

§ **Perceptually unimportant info discarded**



Time-Frequency Mapping

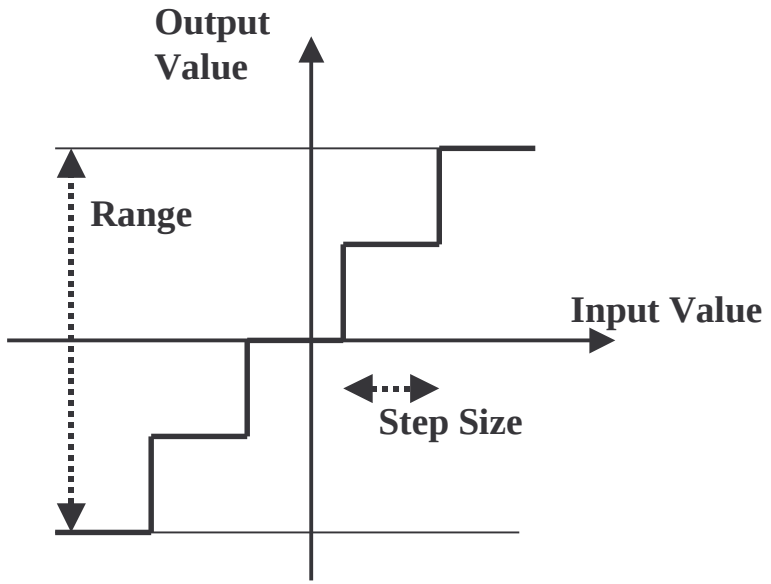
§ **PQMF, TDAC filters, or modified discrete cosine transform**

1. Window-and-overlapping addition (WOA)
2. Modulated cosine transform (MCT)
   - Time-domain aliasing cancellation
   - Variants of the TDAC filterbank
   - Polyphase filterbank

$$X_k = \sum_{i=0}^{N-1} x_i \cos\left(\frac{\pi}{2N} (2i + 1 + \frac{N}{2})(2k + 1) \right)$$

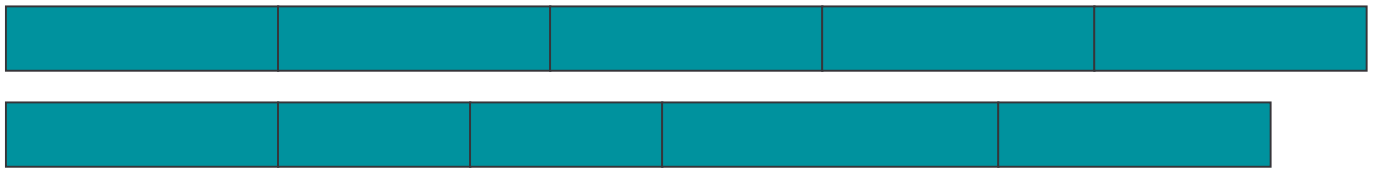
Quantization

- **Represent outputs of filterbank by a finite number of levels**
  - Uniform quantizer
  - Non-uniform quantizer



Variable-Length Coding

- **Takes advantage of different probabilities of quantizer outputs**
- **Implementation problems:**
  - Decoder complexity: bit-by-bit decoding
  - Bit allocation complexity

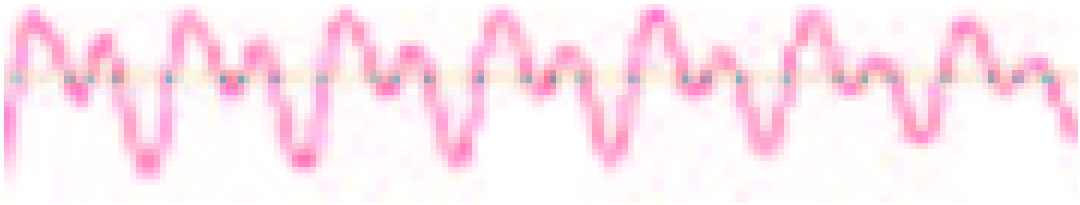


Multichannel Correlation and Irrelevancy

- **Correlation**
  - All channels from same source
  - middle/side coding: left & right into sum & diff.
- **Irrelevancy**
  - Freq. above 2 kHz have temporal envelope
  - intensity coding: high-freq. parts summed
- **Multichannels coding**
  - Matrix or coupling technique

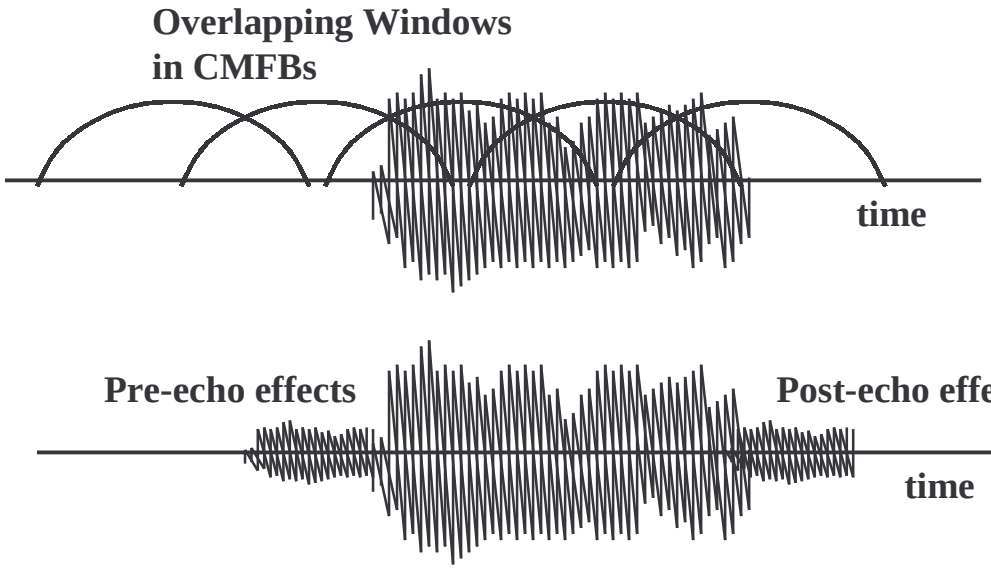
Long-Term Correlation

- **Stationary signals have correlation between adjacent frames**



Pre-echo Control

- **Inverse CMFB will spread quantization noise to the whole window**

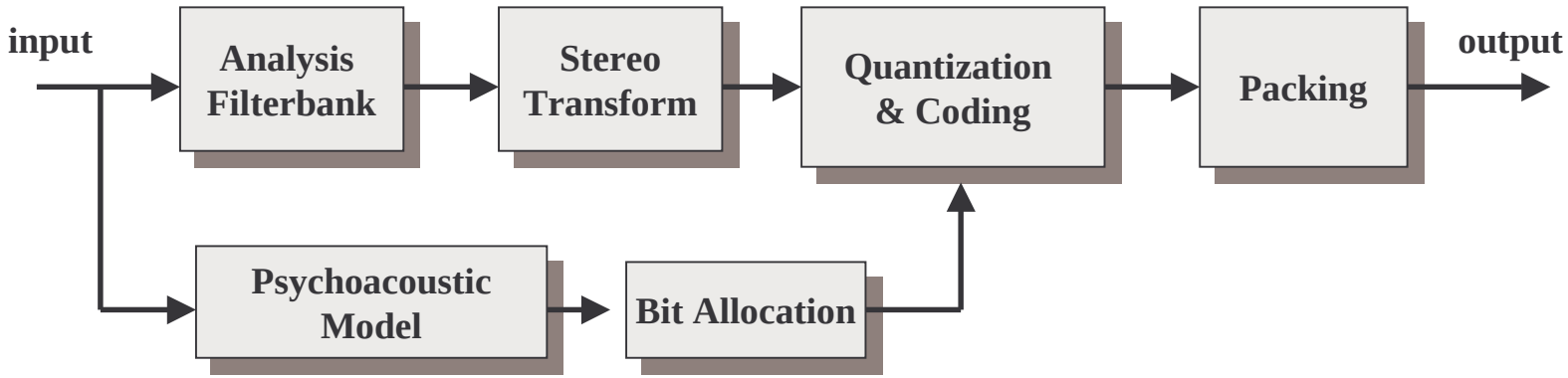


Bit Allocation

- **Allocates the total number of bits available for the quantization**
  - uniform quantizer
    - Simply allocate the bits for subband signals
  - Non-uniform quantizer
    - Quantization noise varies with respect to the input values – quantizer not easy to control
  - Variable-length
    - Relies on quantizer outputs – not easy to control

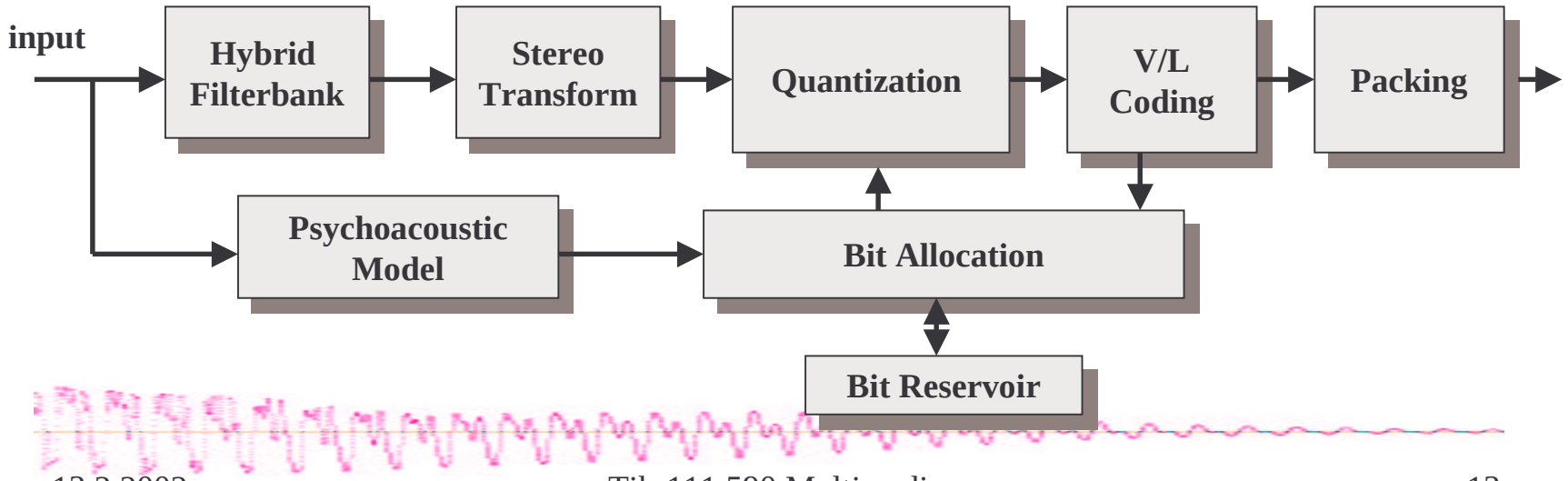
MPEG-1 Layers I & II

- **Filterbank divides into 32 subbands**
- **Psychoacoustic 512/1024 point FFT**



MPEG-1 Layer III

- **Filterbank a cascade of two filterbanks**
- **Non-uniform transf. with V/L coding**
- **Bit reservoir to donate / borrow bits**

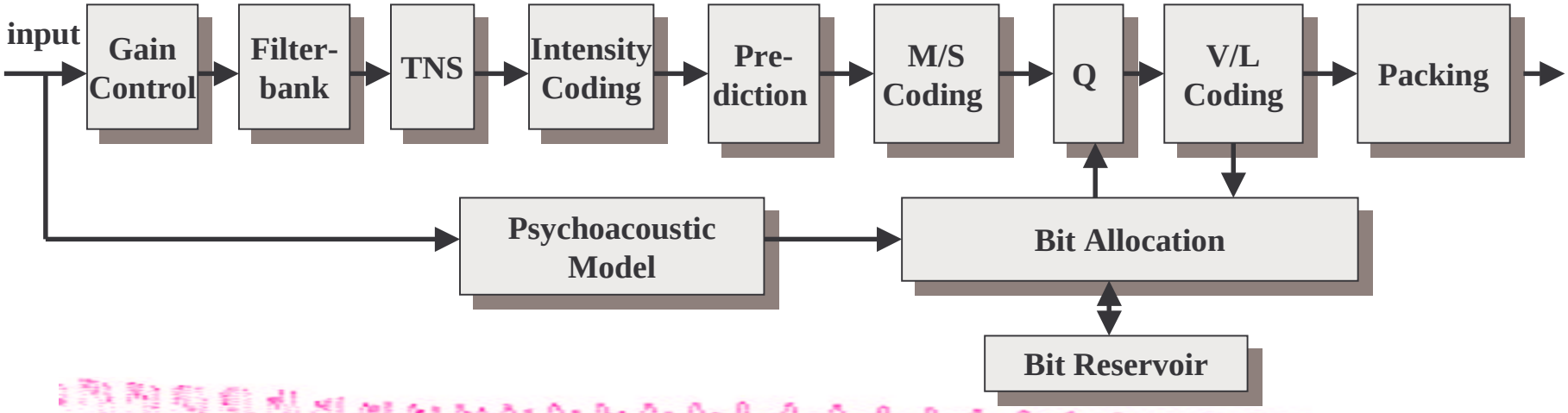


MPEG-2 BC

- **Supports up to 5.1 channels**
- **Lower sampling rates available (16, 22.05 & 24 kHz)**
- **Backward compatible with MPEG-1**

MPEG-2 AAC

- **Not compatible with MPEG-1**
- **Similarities to MPEG-2 BC**
- **Modularity: three profiles**



MPEG-4

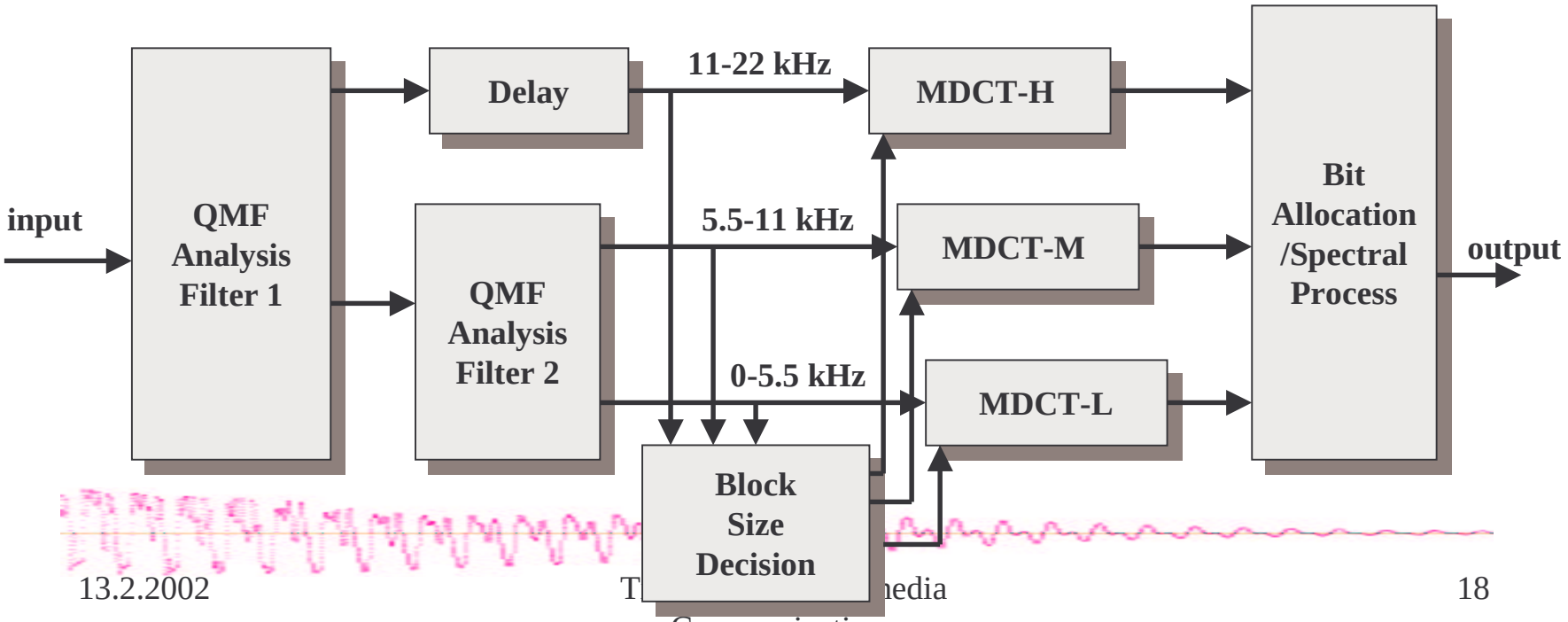
- **Parametric coding**
  - Speech signals sampled at 8 kHz
- **Code-excited linear predictive (CELP) coding**
  - Audio sampled at 8 and 16 kHz
- **Time/frequency (T/F) coding**
  - Wideband audio sampled at 16 kHz or above
  - Similar to MPEG-2 AAC

Philips PASC

- **Used in Philips' DCC cassette**
- **Precision Adaptive Subband Coding**
- **Simplified version of MPEG-1 Layer I**
  - No FFT – uses filterbank for psychoacoustics

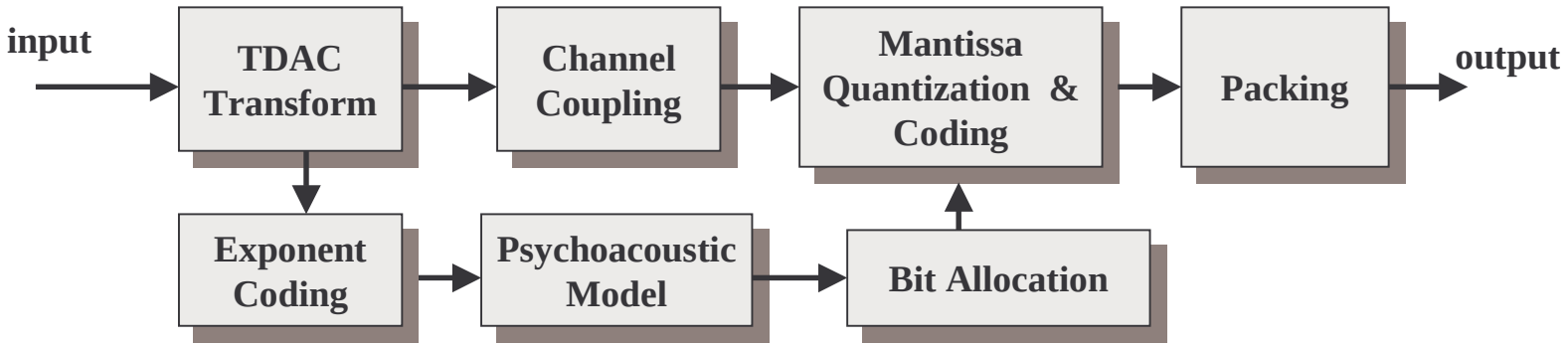
Sony ATRAC

- **Used in Sony MiniDisc**
- **Adaptive TRansform Acoustic Coding**



Dolby AC-3

- **Used in U.S. in HDTV and some DVDs**
- **Up to 5.1 channels**



Conclusions

- **Still room for bit rate and quality improvements**
- **Future work: scalability and editability**

My comment:

- **Open free standards**
  - Ogg Vorbis – rival for MPEG?