

Perceptual Audio Coding: An Overview

Marina Bosi



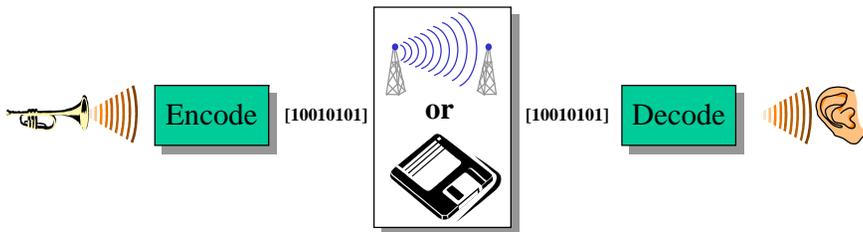
T18 Tutorial
Audio Compression
121st AES Convention October 7, 2006

What Will We Be Talking About?

- Overview of Perceptual Audio Coding
- Examples of Audio Coder Designs
- Sound Examples

Audio Coding

- In general, an audio coder (or codec) is an apparatus whose input is an audio signal and whose output is an audio signal which is perceptually identical (or at least very close) to the (somewhat delayed) input signal



Some Familiar Coders

- Portable Devices, MP3 files, AAC: MPEG Layer III, AAC
- DVDs: Dolby Digital (AC-3) or DTS
- Digital Radio (DAB): MPEG Layer II (MUSICAM), MPEG AAC
- Digital Television (HDTV, DVB): Dolby Digital (AC-3), MPEG Layer II, HE AAC
- Electronic Distribution of Music (EMD): MPEG Layer III (MP3), AAC, WMA
- 3rd Generation Mobile (3GPP): MPEG HE AAC

Evolution of Data Rates for Good Sound Quality for Stereo Signals

- 1992 256 kb/s MPEG Layer II
- 1993 192 kb/s MPEG Layer III
- 1994 128-192 kb/s MPEG MP3
- 1995 384-448 kb/s per 5.1 signal AC-3
- 1997 96-128 kb/s MPEG-2 AAC
- 2000 64-96 kb/s MPEG-4 AAC
- 2001 48-64 kb/s AAC+ (HE AAC)
- 2004 24-48 AAC+ PS
- 2006 64 kb/s per 5.1 signal MPEG Surround

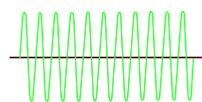
Two Key Ideas

- In perceptual audio coding, two key ideas in the audio signals representation are:
 - removal of Redundancy
 - removal of Irrelevancy

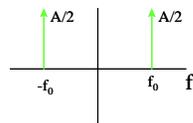
Redundancy

“Redundant *adj* 1. Exceeding what is necessary or normal. 2. Characterized by or containing an excess: *specif* more words than necessary....” [Websters Dictionary]

- In audio coding redundant means that the same information can be represented with fewer bits
- For example, consider a sine wave signal:
 - Redundant: sample the waveform 44,100 times per second and describe each sample with 16 bits
 - Concise: Describe the amplitude, frequency, phase, and duration



$x(t)$



$X(f)$

- Notice that the concise representation of the sine wave is basically equivalent to the information in its Fourier Transform.
- Since music and many other audio signals are very tonal, most coders work in the frequency domain to reduce redundancy

Time to Frequency Mapping Stage

- Designed to provide a compact representation of the audio signals
- Maximize the ability to separate frequency components
- Minimize audibility of blocking artifacts
- Critical sampling
- Perfect reconstruction
- Time delay
- Computational complexity

Examples of Filter-Banks in Audio Coding

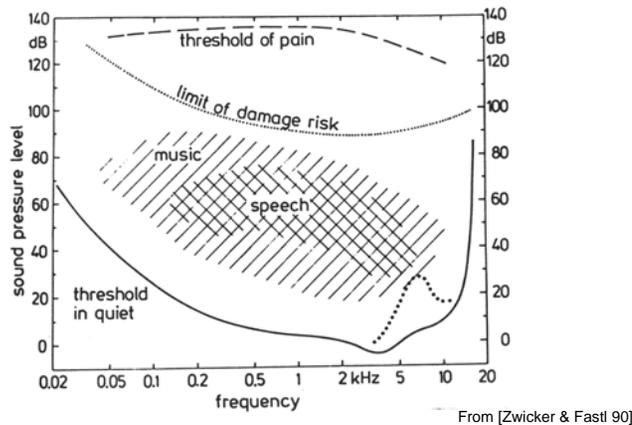
- **PQMF**
 - MPEG Layers I and II: 32-band, 511 PQMF
- **DCT**
 - OCF: 512 frequency lines; 576 impulse response (early version)
- **MDCT/MDST**
 - AC-2A: 256/64 frequency lines; 512/128 impulse response
- **MDCT**
 - AC-3: 256/128 frequency lines; 512/256 impulse response
 - MPEG AAC, PAC: 1024/128 frequency lines; 2048/256 impulse response
- **Hybrid**
 - MPEG Layer III : 576/192 frequency lines; 1664/896 impulse response
 - ATRAC : 512/64 frequency lines; 1072/304 impulse response
- **Wavelets (EPAC)**
 - Tree structure with higher frequency resolution at low frequencies and higher temporal resolution at high frequencies, utilized during transients only
- **Int MDCT (Lossless Coding)**
 - MPEG-4 SLS : same as AAC with 4x over sampling also enabled

Irrelevancy

•“Irrelevant *adj* 1. Not having significant and demonstrable bearing on the matter at hand.” [Websters Dictionary]

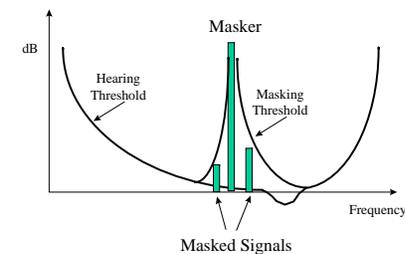
- In audio coding irrelevant data means that you can't hear any difference in the audio signal if those data are omitted
- Main causes of irrelevancy:
 - Hearing Threshold
 - Masking
- Hearing Threshold
 - We can't hear sounds below a certain frequency-dependant level
- Masking
 - Loud sounds can prevent us from hearing softer sounds nearby in time or frequency
- Exploiting irrelevancy
 - Don't code signal components you can't hear
 - Only quantize audible signal components with enough bits to keep quantization noise below the level it can be heard

Hearing Threshold

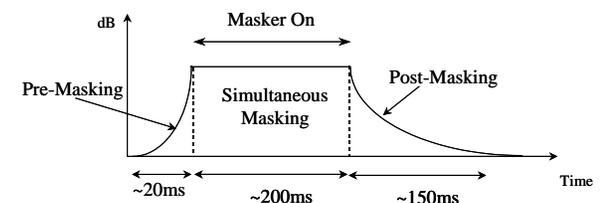


Masking

Frequency (or Simultaneous) Masking

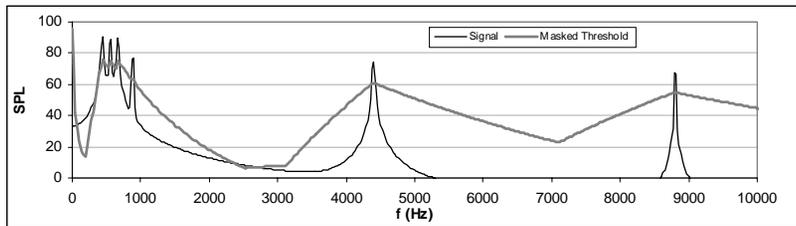


Temporal Masking



Masked Threshold

- The hearing threshold can be combined with the effects of masking from the signal to create the Masked Threshold
- The Masked Threshold represents the level below which noise added to the signal should be inaudible



© 2004-6 Marina Bosi-All rights reserved

13

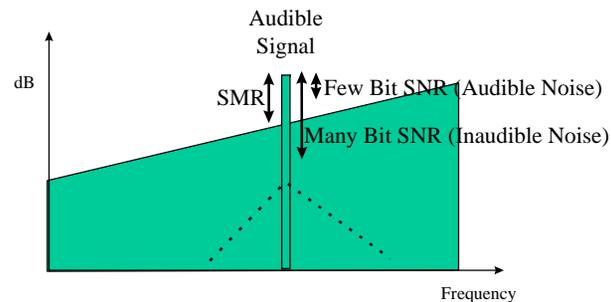
Quantization

- Quantization is the representation of a continuous signal amplitude (time or frequency sample) with a finite number of bits
- Quantization is a lossy process and is the main source of signal degradation in a digital audio coder
- Each additional bit buys you about 6 dB more of signal to noise
 - uniform quantization: count down from overload level of quantizer
 - floating point quantization (linearized A-law): count down from nearest 6 dB point above signal level
- If you know where the Masked Threshold is, you know how many bits are needed to get quantization noise
- Psychoacoustic-based bit allocation is the secret to Perceptual Audio Coders!

© 2004-6 Marina Bosi-All rights reserved

14

Bit/Noise Allocation Using Masked Threshold



© 2004-6 Marina Bosi-All rights reserved

15

Demo: 13 dB Miracle

- The “13 dB miracle” paradox (Johnston and Brandenburg ‘90), where the original signal 🗣️ was injected with noise that was either
 - a) shaped according to psychoacoustic masking models 🗣️
 - b) white 🗣️
- shows that two systems with identical SNR = 13 dB have very different perceived audio quality
- In case a) the quantization noise is shaped so that it is contained below masked thresholds 🗣️
- In case b) the quantization noise is shaped so that it is uniformly distributed in frequency (in general above masked thresholds) 🗣️

© 2004-6 Marina Bosi-All rights reserved

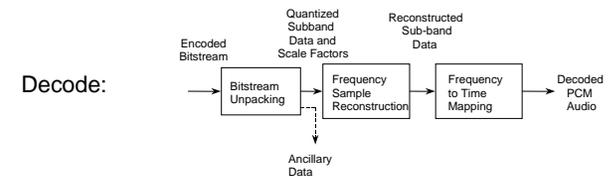
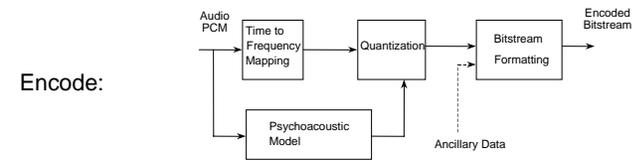
16

Rate Constraints

- The psychoacoustic model provides the SMR values that are needed to achieve transparency (at least according to the psychoacoustic model) – implying a corresponding bit allocation to transparently encode the signal
- However, data rate constraints often limit the “allowed” bit rate of the encoded signal below that needed to achieve transparency so methods are needed to allocate bits subject to both a data rate constraint and the calculated SMR values
- Bit allocation algorithms allocate a greater number of bits to spectral regions with higher than average SMR values at the cost of lower allocations to lower-than-average SMR regions

$$R_b \approx R + \frac{1 \text{ bit}}{6.02 \text{ dB}} \left(SMR_b - \frac{1}{K} \sum_{c=0}^{B-1} N_c SMR_c \right)$$

Basic Building Blocks for a Perceptual Audio Coder



New Trends...

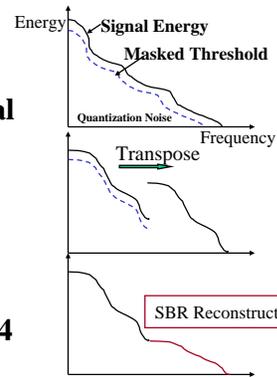
- **Lossless Coding**
- **Parametric Coding**
 - Full Synthesis
 - Spectral Band Replication
 - Spatial Audio Coding (Stereo/Multichannel)
- **Scalable Coding**
- **Each of these are examples of ways to increase the coding efficiency of the system and/or to better encode the signal to specific target application requirements**

Lossless Coding

- **Pure lossless coding allows for average compression ratios of about 2:1 which are much lower than compression ratios achieved in perceptual coding (10-30)**
- **MPEG-4 lossless is based on the following technology :**
 - ADPCM and noiseless coding
 - Int MDCT and noiseless coding
 - 1-bit lossless based on LPC and entropy coding (SACD)

Spectral Band Replication (SBR)

- Only the low part of the signal spectrum is waveform coded
- The high frequency components of the signal are reconstructed from the low frequency components of the signal through a small amount of side information
- Compression efficiency can be significantly improved by using SBR (mp3PRO, MPEG-4 HE AAC)
- Similar principles applied in Enhanced AC-3



Stereo/Multichannel Coding

- Exploit correlations/spatial irrelevancies between stereo/multichannel signals
- Two common approaches:
 - M/S coding
 - Intensity coding
 - M/S coding
 - Change basis from L,R channels to sum (M) and difference (S) channels
 - Intensity coding
 - Approximate signal with a mono signal plus a phase angle to define how signal splits between L,R channels
- Parametric stereo coding
 - The stereo signal is coded as a monaural signal plus a small amount of stereo parameters
- Similar matrix basis changes can be applied to multichannel coding

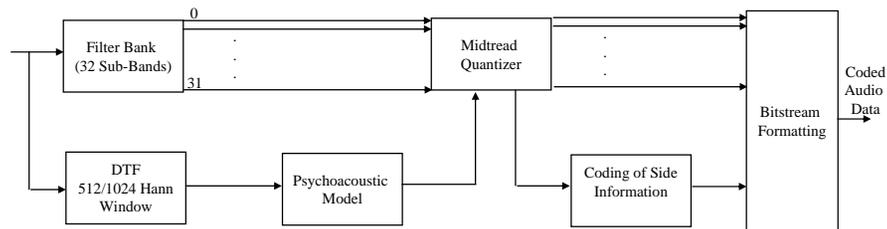
Scalable Audio Coding

- Embed lower bandwidth bitstream in higher bandwidth bitstream
- Key functionality for MPEG-4 audio
- Main types of scalability:
 - Small step scalability
Enhancement layers of ~ 1 k/s (BSAC)
 - Large step scalability
Enhancement layers of 8 k/s and more
 - General audio coding in MPEG-4 supports scalability

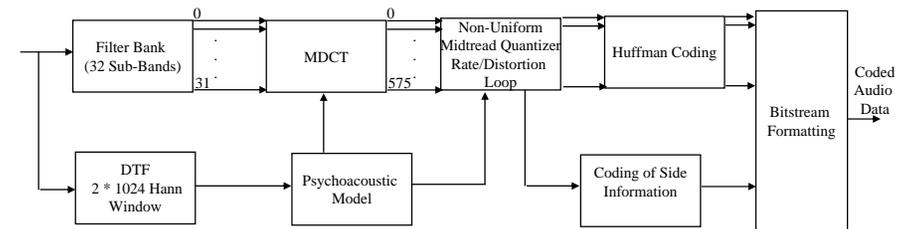
Examples of Audio Coder Designs

- MPEG-1/2 Layers I, II, and III
- MPEG-2/4 AAC
- Dolby AC-3

Layers I and II (Single Channel Mode)



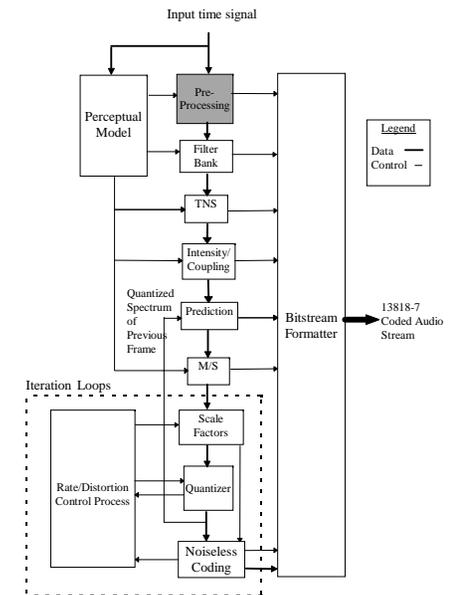
Layer III (Single Channel Mode)



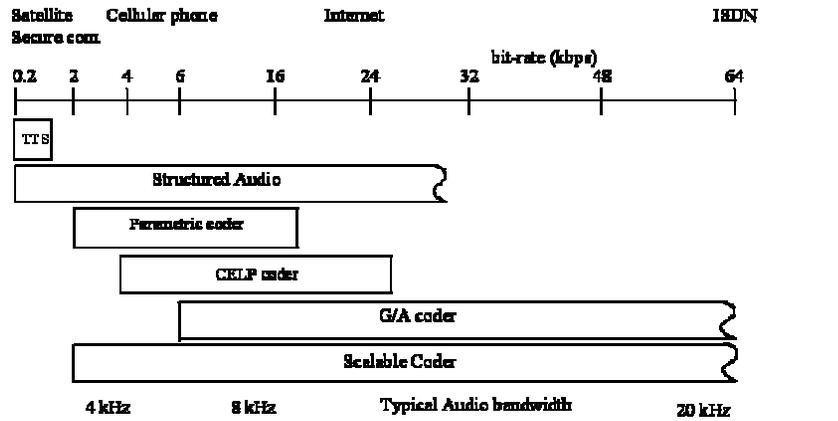
Bitstreams for MPEG-1 Audio Layers

Layer	Header (32)	CRC (0,16)	Bit Allocation (128-256)	Scale Factors (0-384)	Samples	Ancillary Data	
LAYER I							
LAYER II			Bit Allocation (26-188)	SCFSI (0-120)	Scale Factors (0-1080)	Samples	Ancillary Data
LAYER III			Side Information (130-246)	Main Data (May start at a previous frame)			

AAC Encoder Configuration

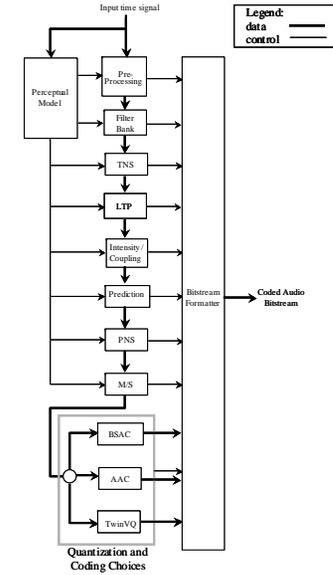


MPEG-4 Audio

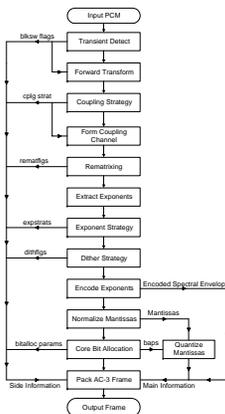


(from ISO/IEC 14496-3)

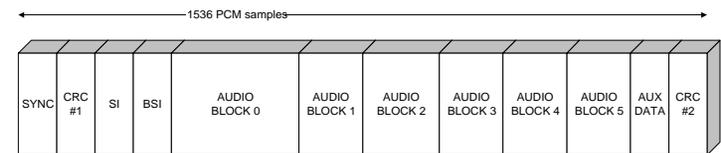
Block Diagram of the MPEG-4 GA Encoder



AC-3 Encoder Flow Diagram



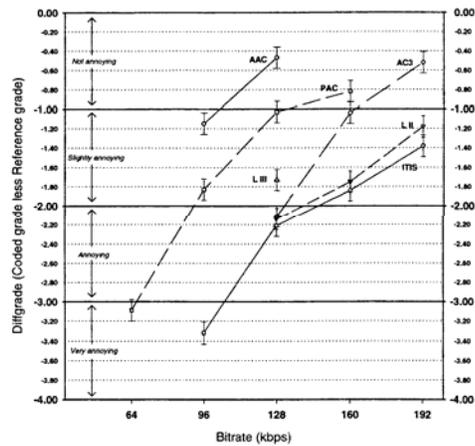
AC-3 Bitstream Overview



- The AC-3 bitstream is composed of independent frames
- Each frame represents a fixed amount of time, equal to 1536 PCM samples:

1 frame = 32 ms for 48 kHz sample rate

Comparison of AC-3, MPEG-2 AAC, MPEG LII and LIII



© 2004-6 Marina Bosi-All rights reserved

(from CRC AES publication)

33

Sound Examples

34

To Learn More:

- M. Bosi and R. E. Goldberg, "Introduction to Digital Audio Coding and Standards", Kluwer /Springer 2003
- "Collected Papers on Digital Audio Bit-Rate Reduction" Neil Gilchrist and Christer Grewin, Editors, Audio Engineering Society 1996
- E. Zwicker and H. Fastl, "Psychoacoustics", Springer-Verlag 1990
- Proceedings of the AES 17th International Conference on "High-Quality Audio Coding", K. Brandenburg and M. Bosi Co-chairs, Florence September 1999
- AES CD-ROM On Perceptual Audio Coders 2001: "Perceptual Audio Coders: What to Listen For", AES 2001.

© 2004-6 Marina Bosi-All rights reserved

35