A TUTORIAL ON
MPEG/AUDIO COMPRESSION

Davis Pan, IEEE Multimedia, 1995

Presented by: Geetha Ramachandran
Outline

• Introduction to MPEG/Audio Compression
• MPEG Encoder Architecture
• Polyphase Filter Bank
• Psychoacoustic Model
• Bit Allocation
• Compression Layers
• Future Developments
About MPEG/Audio Compression

• First international standard for digital compression of high-fidelity audio (1992)
• Three part compression standard - synchronized video and audio at 1.5 Mbps
• Provides compression upto a factor of 6 without affecting sound quality
• Specifies the syntax of the coded bitstream, decoding process and compliance tests
About MPEG/Audio Compression

• “Transparent“, perceptually lossless
• Exploitation of auditory masking
• Audible spectrum can be partitioned in to critical bands
• Threshold for noise masking at a given frequency solely dependant on signal activity within critical band
Features

• Sampling rates: 32, 44.1 or 48 kHz
• Optional CRC error detection codes and ancillary data
• One or two audio channels in four modes
  – Monophonic
  – Dual-monophonic
  – Stereo
  – Joint-stereo mode
• Compression Layers – I, II and III
Encoder Architecture

PCM Audio Input

Time to frequency mapping filter bank

Bit/noise allocation, quantizer, coding

Psychoacoustic model

Bitstream formatting

Ancillary data (optional)

Encoded bitstream
The Polyphase Filter Bank

PCM Audio Input

Time to frequency mapping filter bank

Psychoacoustic model

Bit/noise allocation, quantizer, coding

Ancillary data (optional)

Bitstream formatting

Encoded bitstream
The Polyphase Filter Bank

- Divides the audio signal into 32 equal-width frequency subbands

\[ M[i][k] = \cos\left(\frac{(2i + 1) \times (k - 16) \times \pi}{64}\right) \]

\[ s[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] \times (C[k + 64j] \times x[k + 64j]) \]
The Polyphase Filter Bank

• Equal widths of subbands do not reflect human ear’s frequency dependant behaviour – critical bands with least noise masking determine quantization bits
• Filter bank and its inverse are not lossless transformations
• Adjacent filter banks have a frequency overlap – single signal can affect adjacent outputs
Psychoacoustic Model

PCM Audio Input

Time to frequency mapping filter bank

Bit/noise allocation, quantizer, coding

Bitstream formatting

Psychoacoustic model

Ancillary data (optional)

Encoded bitstream
Psychoacoustic Model

• Psychoacoustics – study of human perception of sound
• Determines the maximum allowable quantization noise energy in each critical band such that quantization noise remains inaudible
• Time align audio data so that the audio data is centered within the analysis data from the filter bank
• Convert audio to frequency domain representation using Fast Fourier transform
Psychoacoustic Model

• Partition spectral values into critical bands
• Separate spectral values into tonal and nontonal components – masking abilities differ
• Set a lower bound for threshold values – threshold in quiet
• Find the masking threshold for each subband
• Calculate the signal-to-mask ratio and pass the value to the bit allocation section
Bit Allocation

1. PCM Audio Input
2. Time to frequency mapping filter bank
3. Psychoacoustic model
4. Bit/noise allocation, quantizer, coding
5. Ancillary data (optional)
6. Bitstream formatting
7. Encoded bitstream
Bit Allocation

• Determines the number of code bits allocated to each subband
• Layer I and II - 

\[ MNR_{dB} = SNR_{dB} - SMR_{dB} \]

• MNR computed for all subbands
• Subband with lowest mask-to-noise ratio is allocated code bits
• Layer III computes the maximum noise allowed for each scale-factor band – increases the scale factor values until quantization noise is below the threshold
Compression Layers

• Layer I – above 128 Kbps per channel
  – Codes audio in frames of 384 audio samples
  – Groups 12 samples from each of the 32 subbands
  – A group of 12 samples gets a bit allocated
  – If not zero, a scale factor is used to size the samples to the full range of the quantizer
  – At the decoder, quantizer output is multiplied with the scale factor to get quantized subband value
Compression Layers

• Layer II – around 128 Kbps per channel
  – Forms frames of 1152 samples per audio channel
  – Codes data in 3 groups of 12 samples per subband
  – Different scale factors for each group used only to avoid distortion
  – Shares scale factors in case
    • The values of the scale factors are sufficiently close
    • Temporal noise masking hides distortion caused by using one scale factor
Compression Layers

• Layer III – around 64 Kbps per channel
  – Compensates for filter bank deficiencies by processing with MDCT
  – Subdivides the subbands outputs in frequency to provide better spectral resolution
  – Alias reduction
  – Variable length Huffman codes for encoding
  – Variable length frame - can borrow or add bits to a reservoir
Future Developments

- MPEG-1 audio laid the foundation for all modern audio compression techniques
- MPEG-2 (1996) extends MPEG audio compression to support 5.1 channel audio
- MPEG-4 (1998) codes audio-visual objects in the stream
- MPEG-7 (2001) – Multimedia Content Description Interface – enables searching, filtering and browsing of multimedia content
References


THANK YOU!