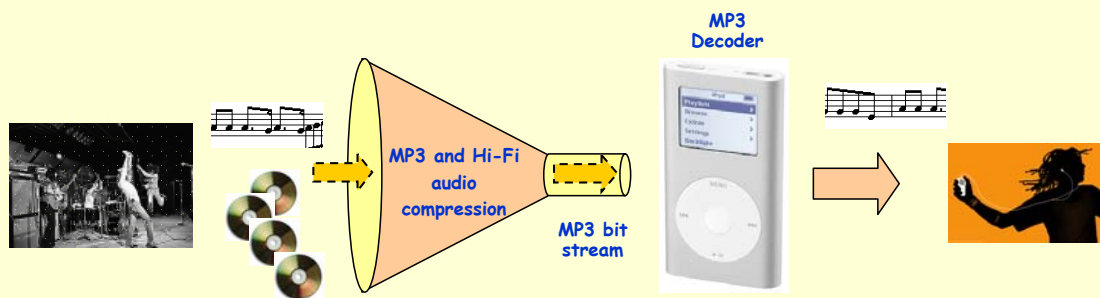


# AUDIO SIGNAL PROCESSING AND CODING

A. Spanias, T. Painter, and V. Atti



Perceptual audio coding, Psychoacoustics, Quantization, Bit-allocation, Huffman coding, Linear prediction/Subband/Transform/Sinusoidal coding, MP3, MP4, Lossless audio coding, Dolby AC3, DTS, Sony SDDS, Perceptual quality measures, and Watermarking

Wiley-Interscience, John Wiley & Sons.

# List of Contents

---

1. Introduction
  - 1.1. Historical perspective
  - 1.2. A General perceptual audio coding architecture
  - 1.3. Audio Coder Attributes
    - 1.3.1. Audio quality
    - 1.3.2. Bitrates
    - 1.3.3. Complexity
    - 1.3.4. Codec delay
    - 1.3.5. Error robustness
  - 1.4. Types of audio coders – An overview
  - 1.5. Organization of the book
  - 1.6. Notational conventions
  
2. Signal Processing Essentials
  - 2.1. Introduction
  - 2.2. Spectra of analog signals
  - 2.3. Review of convolution and filtering
  - 2.4. Uniform sampling
  - 2.5. Discrete-time signal processing
    - 2.5.1. Transforms for discrete-time signals
    - 2.5.2. The discrete and the fast Fourier transform
    - 2.5.3. The discrete cosine transform
    - 2.5.4. The short-time Fourier transform
  - 2.6. Difference equations and digital filters
  - 2.7. The transfer and the frequency response functions
    - 2.7.1. Poles, zeros, and frequency response
    - 2.7.2. Examples of digital filters for audio applications
  - 2.8. Review of multirate signal processing
    - 2.8.1. Down-sampling by an integer
    - 2.8.2. Up-sampling by an integer
    - 2.8.3. Sampling rate changes by non-integer factors
    - 2.8.4. Quadrature mirror filter banks
  - 2.9. Discrete-time random signals
    - 2.9.1. Random signals processed by LTI digital filters
    - 2.9.2. Autocorrelation estimation from finite-length data
  - 2.10. Summary
  
3. Quantization and Entropy Coding
  - 3.1. Introduction
    - 3.1.1. The quantization--bit allocation--entropy coding module
  - 3.2. Density functions and quantization
  - 3.3. Scalar quantization
    - 3.3.1. Uniform quantization
    - 3.3.2. Non-uniform quantization
    - 3.3.3. Differential PCM
  - 3.4. Vector quantization
    - 3.4.1. Structured VQ
    - 3.4.2. Split-VQ
    - 3.4.3. Conjugate-structure VQ
  - 3.5. Bit-allocation algorithms

- 3.6. Entropy coding
  - 3.6.1. Huffman coding
  - 3.6.2. Rice coding
  - 3.6.3. Golomb coding
  - 3.6.4. Arithmetic coding
- 3.7. Summary
  
- 4. Linear Prediction in Narrowband and Wideband Coding
  - 4.1. Introduction
  - 4.2. Review of LP-based source-system modeling for speech
  - 4.3. Short-term linear prediction
    - 4.3.1. Long-term prediction
    - 4.3.2. ADPCM using linear prediction
  - 4.4. Open-loop analysis-synthesis linear prediction
  - 4.5. Analysis-by-synthesis linear prediction
    - 4.5.1. Code-excited linear prediction algorithms
  - 4.6. Linear prediction in wideband coding
    - 4.6.1. Wideband speech coding
    - 4.6.2. Wideband audio coding
  - 4.7. Summary
  
- 5. Psychoacoustic Principles
  - 5.1. Introduction
  - 5.2. Absolute threshold of hearing
  - 5.3. Critical bands
  - 5.4. Simultaneous masking, masking asymmetry, and the spread of masking
    - 5.4.1. Noise-masking-tone
    - 5.4.2. Tone-masking-noise
    - 5.4.3. Noise-masking-noise
    - 5.4.4. Asymmetry of masking
    - 5.4.5. The spread of masking
  - 5.5. Non-simultaneous masking
  - 5.6. Perceptual entropy
  - 5.7. Example codec perceptual model: ISO/IEC 11172-3 (MPEG - 1) psychoacoustic model-1
    - 5.7.1. Step 1: Spectral analysis and SPL normalization
    - 5.7.2. Step 2: Identification of tonal and noise maskers
    - 5.7.3. Step 3: Decimation and reorganization of maskers
    - 5.7.4. Step 4: Calculation of individual masking thresholds
    - 5.7.5. Step 5: Calculation of global masking thresholds
  - 5.8. Perceptual bit allocation
  - 5.9. Summary
  
- 6. Time-Frequency Analysis: Filter Banks and Transforms
  - 6.1. Introduction
  - 6.2. Analysis-synthesis framework for M-band filter banks
  - 6.3. Filter banks for audio coding: Design considerations
    - 6.3.1. The role of time-frequency resolution in masking power estimation
    - 6.3.2. The role of frequency resolution in perceptual bit allocation
    - 6.3.3. The role of time resolution in perceptual bit allocation
  - 6.4. Quadrature mirror and conjugate quadrature filters
  - 6.5. Tree-structured QMF and CQF M-band banks
  - 6.6. Cosine modulated “pseudo QMF” M-band banks
  - 6.7. Cosine modulated perfect reconstruction (PR) M-band banks and the modified discrete cosine transform (MDCT)
    - 6.7.1. Forward and inverse MDCT
    - 6.7.2. MDCT window design

- 6.7.3. Example MDCT windows (Prototype FIR filters)
- 6.8. Discrete Fourier and discrete cosine transform
- 6.9. Pre-echo distortion
- 6.10. Pre-echo control strategies
  - 6.10.1. Bit reservoir
  - 6.10.2. Window switching
  - 6.10.3. Hybrid, switched filter banks
  - 6.10.4. Gain modification
  - 6.10.5. Temporal noise shaping
- 6.11. Summary

## 7. Transform Coders

- 7.1. Introduction
- 7.2. Optimum coding in the frequency domain
  - 7.2.1. OCF-1
  - 7.2.2. OCF-2
  - 7.2.3. OCF-3
- 7.3. Perceptual transform coder
  - 7.3.1. PXFM
  - 7.3.2. SEPXFM
- 7.4. Brandenburg-Johnston hybrid coder
- 7.5. CNET coders
  - 7.5.1. CNET DFT coder
  - 7.5.2. CNET MDCT coder-1
  - 7.5.3. CNET MDCT coder-2
- 7.6. Adaptive spectral entropy coding
- 7.7. Differential perceptual audio coder
- 7.8. DFT noise substitution
- 7.9. DCT with vector quantization
- 7.10. MDCT with vector quantization

## 8. Subband Coders

- 8.1. Introduction
  - 8.1.1. Subband Algorithms
- 8.2. DWT and Discrete Wavelet Packet Transform (DWPT)
- 8.3. Adapted WP Algorithms
  - 8.3.1. DWPT Coder with Globally Adapted Daubechies Analysis Wavelet
  - 8.3.2. Scalable DWPT Coder with Adaptive Tree Structure
  - 8.3.3. DWPT Coder with Globally Adapted General Analysis Wavelet
  - 8.3.4. DWPT Coder with Adaptive Tree Structure and Locally Adapted Analysis Wavelet
  - 8.3.5. DWPT Coder with Perceptually Optimized Synthesis Wavelets
- 8.4. Adapted Non-Uniform Filter Banks
  - 8.4.1. Switched Non-Uniform Filter Bank Cascade
  - 8.4.2. Frequency-Varying Modulated Lapped Transforms
- 8.5. Hybrid WP and Adapted WP/Sinusoidal Algorithms
  - 8.5.1. Hybrid Sinusoidal/Classical DWPT Coder
  - 8.5.2. Hybrid Sinusoidal/M-Band DWPT Coder
  - 8.5.3. Hybrid Sinusoidal/DWPT Coder with WP Tree Structure Adaptation (ARCO)
- 8.6. Subband Coding with Hybrid Filter Bank/CELP Algorithms
  - 8.6.1. Hybrid Subband/CELP Algorithm for Low-Delay Applications
  - 8.6.2. Hybrid Subband/CELP Algorithm for Low Complexity Applications
- 8.7. Subband Coding with IIR Filter Banks

## 9. Sinusoidal Coders

- 9.1. Introduction
- 9.2. The Sinusoidal Model

- 9.2.1. Sinusoidal Analysis and Parameter Tracking
- 9.2.2. Sinusoidal Synthesis and Parameter Interpolation
- 9.3. Analysis/Synthesis Audio Codec (ASAC)
  - 9.3.1. ASAC Segmentation
  - 9.3.2. ASAC Sinusoidal Analysis-by-Synthesis
  - 9.3.3. ASAC Bit Allocation, Quantization, Encoding, and Scalability
- 9.4. Harmonic and Individual Lines Plus Noise Coder (HILN)
  - 9.4.1. HILN Sinusoidal Analysis-by-Synthesis
  - 9.4.2. HILN Bit Allocation, Quantization, Encoding, and Decoding
- 9.5. FM Synthesis
  - 9.5.1. Principles of FM Synthesis
  - 9.5.2. Perceptual Audio Coding Using an FM Synthesis Model
- 9.6. The Sines + Transients + Noise (STN) Model
- 9.7. Hybrid Sinusoidal Coders
  - 9.7.1. Hybrid Sinusoidal-MDCT Algorithm
  - 9.7.2. Hybrid Sinusoidal-Vocoder Algorithm
- 9.8. Summary

## 10. Audio coding standards and algorithms

- 10.1. Introduction
- 10.2. MIDI Versus Digital Audio
  - 10.2.1. MIDI Synthesizer
  - 10.2.2. General MIDI (GM)
  - 10.2.3. MIDI Applications
- 10.3. Multi-channel Surround Sound
  - 10.3.1. The evolution of surround sound
  - 10.3.2. The Mono, the stereo, and the surround sound formats
  - 10.3.3. The ITU-R BS.775 5.1-channel configuration
- 10.4. MPEG Audio Standards
  - 10.4.1. MPEG-1 Audio (ISO/IEC 11172-3)
  - 10.4.2. MPEG-2 BC/LSF (ISO/IEC-13818-3)
  - 10.4.3. MPEG-2 NBC/AAC (ISO/IEC-13818-7)
  - 10.4.4. MPEG-4 Audio (ISO/IEC 14496-3)
  - 10.4.5. MPEG-7 Audio (ISO/IEC 15938-4)
  - 10.4.6. MPEG-21 Framework (ISO/IEC-21000)
- 10.5. Adaptive Transform Acoustic Coding (ATRAC)
- 10.6. Lucent Technologies PAC, EPAC, and MPAC
  - 10.6.1. Perceptual Audio Coder (PAC)
  - 10.6.2. Enhanced PAC (EPAC)
  - 10.6.3. Multi-channel PAC (MPAC)
- 10.7. Dolby Audio Coding Standards
  - 10.7.1. Dolby AC-2, AC-2A
  - 10.7.2. Dolby AC-3 / Dolby Digital / Dolby SR- D
- 10.8. Audio Processing Technology APT-x100
- 10.9. DTS – Coherent Acoustics

## 11. Lossless Audio Coding and Digital Watermarking

- 11.1. Introduction
- 11.2. Lossless audio coding ( $L^2AC$ )
  - 11.2.1.  $L^2AC$  Principles
  - 11.2.2.  $L^2AC$  Algorithms
- 11.3. DVD-Audio
  - 11.3.1. Meridian Lossless Packing (MLP)
- 11.4. Super Audio CD (SACD)
  - 11.4.1. SACD Storage Format
  - 11.4.2. Sigma-Delta Modulators (SDM)

- 11.4.3. Direct Stream Digital (DSD) Encoding
  - 11.5. Digital audio watermarking
    - 11.5.1. Background
    - 11.5.2. A generic architecture for DAW
    - 11.5.3. DAW schemes – Attributes
  - 11.6. Summary of Commercial Applications and Advancements in the Audio Coding Standardization
- 
- 12. Quality Measures for Perceptual Audio Coding
    - 12.1. Introduction
    - 12.2. Subjective Quality Measures
    - 12.3. Confounding Factors in Subjective Evaluations
    - 12.4. Subjective Evaluations of Two-Channel Standardized Codecs
    - 12.5. Subjective Evaluations of 5.1-Channel Standardized Codecs
    - 12.6. Subjective Evaluations Using Perceptual Measurement Systems
      - 12.6.1. CIR Perceptual Measurement Schemes
      - 12.6.2. NSE Perceptual Measurement Schemes
    - 12.7. Proposed Algorithms for Perceptual Measurement
      - 12.7.1. Example 1: Perceptual Audio Quality Measure (PAQM)
      - 12.7.2. Example 2: Noise-to-Mask Ratio (NMR)
      - 12.7.3. Example 3: Objective Audio Signal Evaluation (OASE)
    - 12.8. ITU-R BS.1387 and ITU-T P.861: Recent Standards for Perceptual Quality Measurement
    - 12.9. Future Directions for Perceptual Codec Quality Measures