FFT-Based Digital Audio Compression

Scott Levine and Julius O. Smith III (jos@ccrma.stanford.edu)
Center for Computer Research in Music and Acoustics (CCRMA)
Department of Music, Stanford University
Stanford, California 94305

April 10, 2018

• Subband Coding
• Transform Coding
• Princen-Bradley Filter Bank
• Dolby AC-2 and AC-3
• MPEG Audio Compression (MUSICAM)
• JPEG Image Compression
References


Subband Coding

Quantize outputs of critically sampled filter bank

- If filter bank mimics hearing (e.g., constant-Q), quantization can be based on *auditory masking*
- Quantized filterbanks outputs can be *entropy coded*, e.g., Huffman
- FFT efficiently implements uniform filter bank (cf. Portnoff on implementing the phase vocoder using the FFT)
Quantize outputs of critically sampled STFT

- Need $R = M = N$ for critical sampling (Hop size = Window length = DFT length) $\implies$ rectangular window
- Quantization noise causes discontinuities in reconstruction due to rectangular window
- Need smooth post-window (synthesis filter) to hide frame-to-frame discontinuities, e.g., weighted overlap-add with $w(n) = \sqrt{\text{Hanning}(n)}$
- Smooth windows require at least 50% overlap $\implies$ 200% initial data expansion
- Is there a better way?
Princen-Bradley Filter Bank

- Alternate DCT and DST using 50% OLA, constant-OLA window, and quarter-frame rotation

\[ \text{DCT}(x) \approx \text{re} \{ \text{FFT}(x) \} = \frac{X(\omega_k) + X^*(\omega_k)}{2} \quad \leftrightarrow \quad \frac{x + \text{FLIP}(x)}{2} \]

- Thus, DCT data is \textit{time aliased} with its flip

\[ \text{DST}(x) \approx \text{im} \{ \text{FFT}(x) \} = \frac{X(\omega_k) - X^*(\omega_k)}{2j} \quad \leftrightarrow \quad \frac{x - \text{FLIP}(x)}{2} \]

- Thus, DST data is \textit{time aliased} with \textit{minus} its flip

- Alternating DCT and DST in this way \textit{cancels} aliasing

- This is “time-domain aliasing cancellation”

- Princen-Bradley filter bank = special case of “Lapped Orthogonal Transforms (LOT)” (see Malvar)

  - Let number of filter bank channels = \( N \)
  - Let length of each channel analysis filter be \( M \)
  - LOT = Critically sampled FIR filter bank with \( M = 2N \)
Dolby AC-2 and AC-3

- Original AC-2: fixed factor of 6 “transparent” compression for 44.1kHz 16-bit audio
- Now adjustable from 64 to 192 kilobits/sec/channel (ratios from 11 to 3.7 for 44.1kHz 16-bit audio)
- Mono algorithm (no use of stereo correlation)
- Can decode 2 channels in real time on 1 Motorola DSP5600x at 25MHz
- Uses Princen-Bradley Filterbank (DCT,DST)
- FFT can be used to compute DCT and DST for speed
- Nominal frame size = 512 samples at 44.1kHz (12ms)
- Second frame size (128) chosen for transients
- 256 FFT bins partitioned into 40 critical bands
- Masking pattern estimated
- One exponent per critical band (K. Brandenburg)
- Mantissa bit allocation based on signal to masking ratio
MUSICAM

MUSICAM = “Masking-pattern Universal Subband Integrated Coding and Multiplexing”

• Commonly referred to as “MPEG Audio”
• Compresses 44.1kHz 16-bit audio from 706 Kbits/sec down to around 128 Kbits/sec (ratio = 5.5)
• Quality is “transparent”
• Subband coder
  – 32-band uniform FIR filter bank
  – Uniformly spaced filters allow use of fast transform
  – Less delay than a dyadic constant-Q filter bank
  – Analysis filters are length 512 $\Rightarrow$ length $512/32 = 16$ polyphase channel filters
• FFT used in parallel with filter bank
  – Masking pattern based on spectral power estimate
• No entropy coding
JPEG Image Compression

• Compresses individual images (no motion prediction as in MPEG)
• Baseline JPEG quantizes 2D DCT of $8 \times 8$ block of pixels
• Specialized, optimized FFT-like DCT transforms used
• Colors processed separately
• DCT blocks ordered in fixed “raster” pattern
• DCT approximates the Karhunen-Loeve transform (equal in the limit as transform size $\rightarrow \infty$)
• Compression ratio variable
• Progressive coding supported
  – Low-frequency DCT coefficients sent first
  – Higher frequency DCT coefficients sent later
• Hierarchical (“pyramidal”) resolution coding supported (HF coding differential wrt LF)
• Lossless predictive coding also supported (no DCT)
• “Blocking” artifacts possible due to non-overlapping DCT blocks
• Uses entropy coding (Huffman)