Digital Audio Compression: Why, What, and How

An Absurdly Short Course

Jeff Bier

Berkeley Design Technology, Inc.

Outline

- Why Compress?
- What is Audio Compression?
- How Does it Work?
- Conclusions
Digital Audio Compression: Why, What, and How

Why: Too Much Data!

- “CD quality”
  - Rate: 2 audio channels $\times$ 44,100 samples/sec $\times$ 16 bits = 1.4 Mbit/second audio data
  - Audio CD capacity: 0.8 gigabytes audio data
- “Cinema quality”
  - ~6 audio channels $\times$
    48,000 samples/sec $\times$ 16 bits = 4.6 Mbit/second, ~1 Gbyte/hour
- Typical compression today: few hundred kbit/sec for even 6 channels

Crossing Thresholds

Memory Capacity - Mbytes/chip

<table>
<thead>
<tr>
<th>Year</th>
<th>Mbytes/chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0.008</td>
</tr>
<tr>
<td>1980</td>
<td>0.032</td>
</tr>
<tr>
<td>1985</td>
<td>0.125</td>
</tr>
<tr>
<td>1990</td>
<td>0.5</td>
</tr>
<tr>
<td>1995</td>
<td>2</td>
</tr>
<tr>
<td>2000</td>
<td>32</td>
</tr>
<tr>
<td>2005</td>
<td>128</td>
</tr>
<tr>
<td>2010</td>
<td>512</td>
</tr>
</tbody>
</table>

- Audio CD
- DVD Sound Track
- MP3 Album
- MP3 Song
Commercial Applications

- Cinema (digital film sound)
- Consumer devices
  - Mini-disc (Sony)
  - Handheld players
  - Games
  - DVD
- Distribution
  - Internet distribution
  - Audio over cable (set-top box)
  - Satellite/terrestrial audio broadcast
  - Digital television

What: Compression Goals

- Reduced bandwidth and/or storage
- Make decoded signal as close as possible to original signal
- Lowest implementation complexity
- Reasonable arithmetic requirements
- Applicable to as many signal types as possible
- Robust
- Scalable
- Extensible
Psychoacoustics

- What does it cover?
  - Relationship between what arrives at the ear and what we hear
- Why is it important for compression?
  - Don’t transmit what the ear can’t hear
- How to figure out what ear can’t hear?
  - Range of human hearing
  - Masking

Range of Human Hearing

Zwicker/Fastl p. 17
Digital Audio Compression: Why, What, and How

Auditory Masking

- One signal can make another inaudible

Auditory Masking (cont’d)

- Temporal Masking

After Zwicker/Fastl p. 78, Buser/Imbert p. 47

© 2000 Berkeley Design Technology, Inc.

ICSPAT 2000 Page 5 October 18, 2000
Perceptual Compression

- Window (Snapshot)
- Spectral Analysis
- Calculate Thresholds
- Remove Inaudible Components
- Pack Data into Frames
- Quantize
- Coding Scheme

Spectral Analysis/Synthesis

- What is it?
  - Break a signal into spectrum
  - Recover the signal from its spectrum

Time domain ↔ Frequency domain

Forward (analysis) ↔ Reverse/inverse (synthesis)
Spectral Analysis (cont’d)

- How is it done?
  1. Filter bank
  2. Transform

- Does reconstructed output = input?
  Yes, if:
  - Don’t change transform data (no filtering)
  - Design window correctly
  - Window input often enough
  - Overlap windows in analysis/resynthesis properly

- If yes: Identity System
  - Solid basis for further changes
How: Analysis

- Spectral analysis
  - DCT
  - Wavelet
  - ...

How: Analysis

- Calculate masking thresholds
  - “Perceptual model”
How: Noise Allocation

- Remove inaudible components
- Quantize remaining components
  - Use minimum # bits
- Adds noise
  - Keep below masking threshold

(Insert graph showing noise allocation)

How: More Tricks

- Filter
  - Bandlimit input signal
    - LFE bandlimited to <120 Hz
- Differential coding of spectral values
- Coupling
  - Across time
  - Across channels
How: Coding Scheme

- Direct coding
- Entropy (e.g., Huffman) coding
  - MPEG: “noiseless coding”
  - PAC: “information-theoretic coding”
- Quantization table
- Run-length
- Vector quantization

Artifacts: “Pre-echo”

- Quantization noise spread
- Noise components \( \geq 1-2 \text{ msec} \) before impulsive signal not masked
- Fix: Shorter window; wavelets (ePAC)
Comparing Specifications

- Bit rate ranges: < 8 kbps - 9.6 Mbps
- Bit widths: 16-24 bits
- Sample rates: 8-192 kHz
- Number of channels: 1-many dozen
- Spectral bins: 128-1024
- Time resolution: 4-12 msec
- Compression ratios: 6-12:1 typical
- Audio quality… transparent to annoying

Which Algorithm?

- MPEG
  - MPEG-1 1992
  - MPEG-2 1994
  - AAC 1997
  - MPEG-4 1999...

- ATRAC (Sony) 1992
- AC-3 (Dolby) 1995
- TwinVQ (NTT) 1995
- Coherent Acoustics (DTS) 1996
- MLP (Meridian) 1997

- PAC (Lucent) 1992
- TwinVQ (NTT) 1995
- G2 (RealNetworks) 1998
- WMA (Microsoft) 1999
- Qdesign 1999
**MPEG Family**

- Moving Pictures Experts Group
- Moving pictures + associated audio
- MPEG-1, MPEG-2 (MP3), MPEG-4
- Ongoing standardization effort (MPEG-7)

**MPEG-1 Audio**

- 1992
- Able to work well with CD, DAT
- One or two channels
  - Single channel
  - Two independent channels
  - Stereo
  - Stereo with joint coding
- 32, 44.1, 48 kHz
- Specifies bit stream format, decoder structure, but not encoder (!)
MPEG-1 Audio

Layers

- Layer 1: simplest; Philips DCC
- Layer 2: more efficient coding; DAB, CD-I
- Layer 3: higher frequency and time resolution; ISDN, Internet
- All 3 layers use same header structure
- Decoder for one layer must also decode lower-numbered layers
- Higher-numbered layers have more complex decoder

MPEG-2 Audio

- 1994
- MPEG-2 video for digital TV
- Higher bit rates than MPEG-1
- Backward compatible with MPEG-1
  - Three layers, like MPEG-1
- Add lower sample rates
  - 16, 22.05, 24 kHz
- 5.1 + up to 7 multilingual/commentary channels
- “MP3” = MPEG-1/2 Layer 3 (not “MPEG-3”)
MPEG-2 Advanced Audio Coding (AAC)

- 1997
- Goals:
  - “Indistinguishable” at 384 kbit/sec
  - Higher quality, multi-channel
- Features:
  - “Non-backward-compatible” ("NBC")
  - Up to 48 channels (stereo, 5.1 ...)
  - “Tools” (modules) combined into “profiles”

AAC Profiles

- LC (Low-Complexity)
  - Most commonly used
  - TNS (Temporal Noise Shaping)
- SSR (Scalable Sampling Rate)
  - Features gain control “tool”
- Main
  - LTP (Long-Term Predictor)
  - Delivers the best audio quality of the three profiles
Conclusions

- Entertainment is going digital
  - Audio is a key component
  - Many new market opportunities opening up
    - Internet audio is hot; audio may be the Internet “killer app”

- Audio compression is a key technology
  - Many algorithms, many applications
  - Better algorithms → better quality, more compression
  - Computation requirements are going up

In the Future...

Photo credit: Dr. Richard O. Duda, SJSU
For More Information

http://www.BDTI.com  Collection of BDTI's papers on DSP processors, tools, apps, benchmarking

http://www.eg3.com/dsp  Links to other good DSP sites

comp.dsp  Usenet group

Microprocessor Report  For info on newer DSPs

DSP Processor Fundamentals, BDTI  Textbook on DSP processors

Or, join BDTI...We're Hiring! (See www.BDTI.com)