Digital Audio Compression: Why, What, and How

An Absurdly Short Course

Jeff Bier

Berkeley Design Technology, Inc.

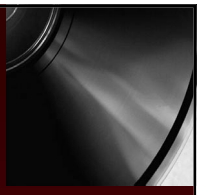
1

Outline

- Why Compress?
- What is Audio Compression?
- How Does it Work?
- Conclusions

2

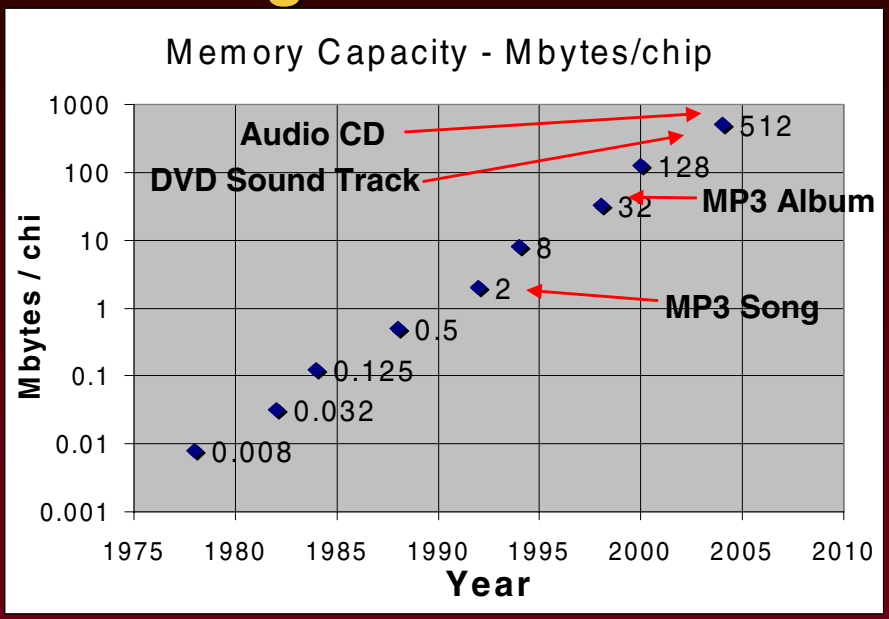
Why: Too Much Data!



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

3

Crossing Thresholds



4

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

5

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

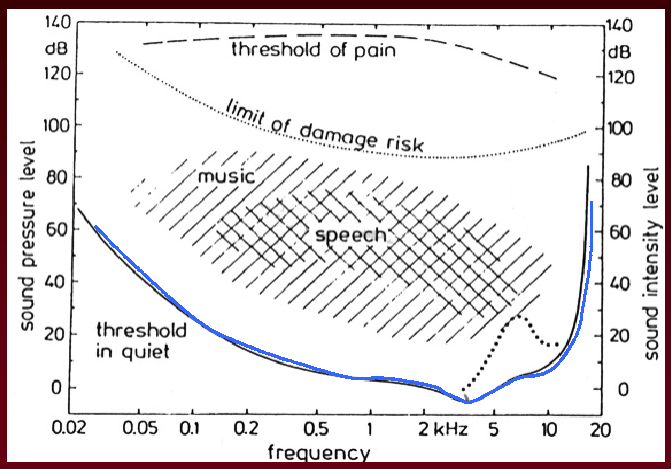
6

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

7

Range of Human Hearing

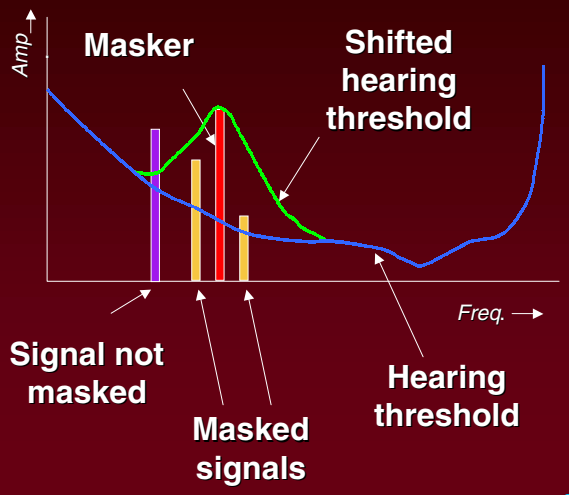


Zwicker/Fastl p. 17

8

Auditory Masking

- One signal can make another inaudible

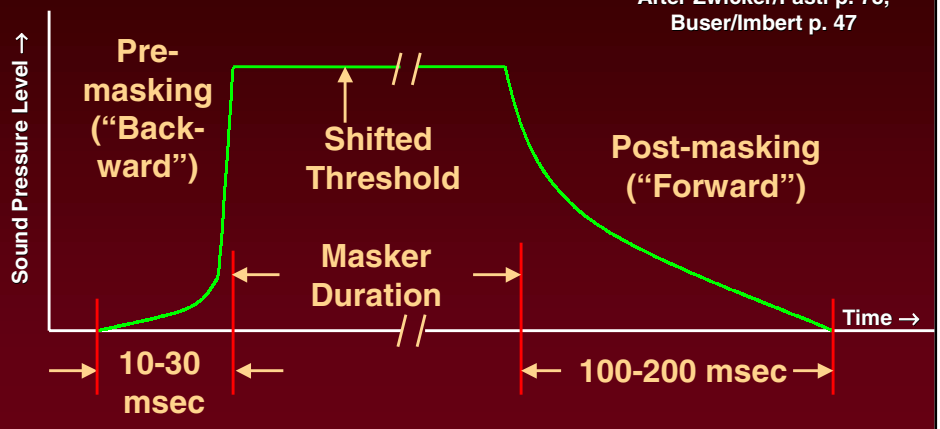


9

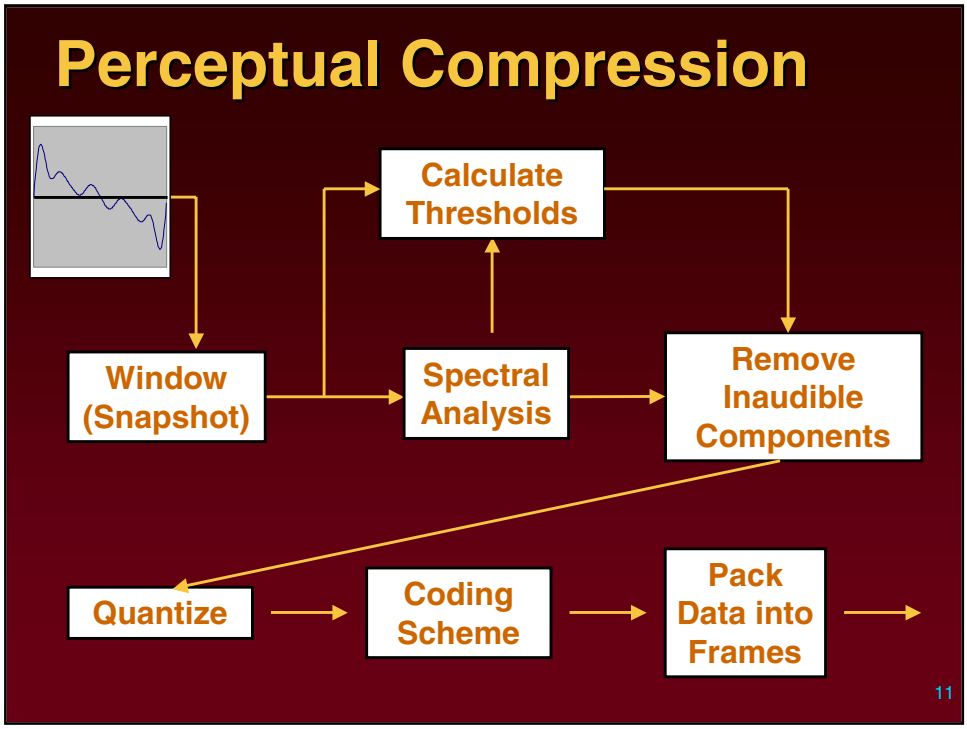
Auditory Masking (cont'd)

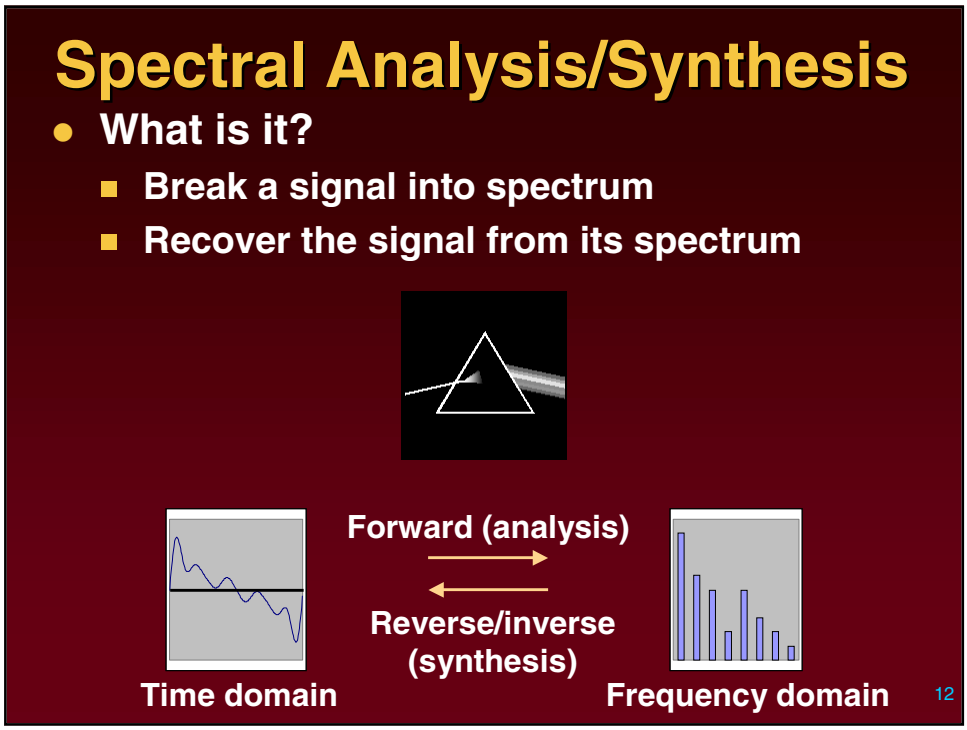
- Temporal Masking

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



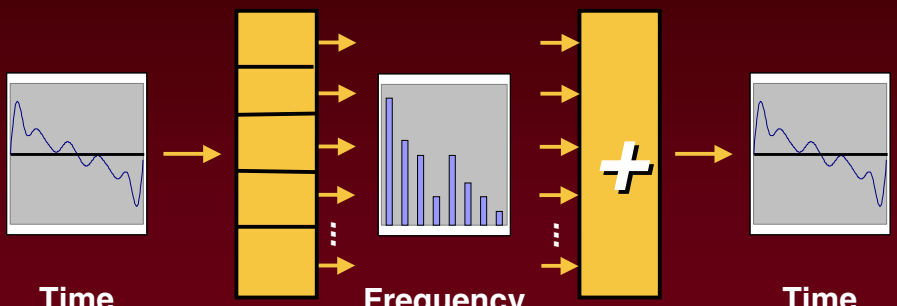
10





Spectral Analysis (cont'd)

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



13

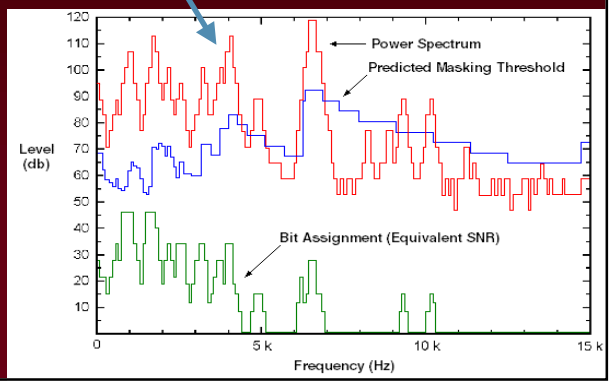
Spectral Analysis (cont'd)

- 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

14

How: Analysis

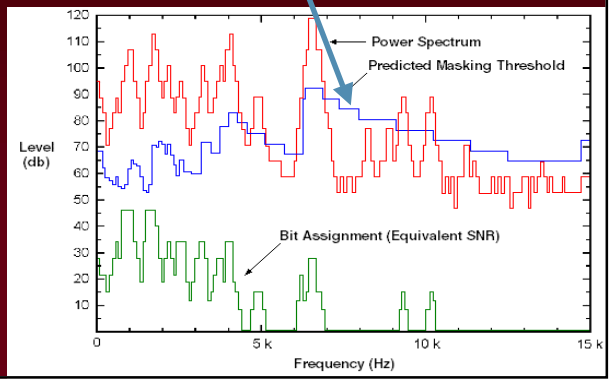
- Spectral analysis
  - DCT
  - Wavelet
  - ...



Davidson et al., 1994

How: Analysis

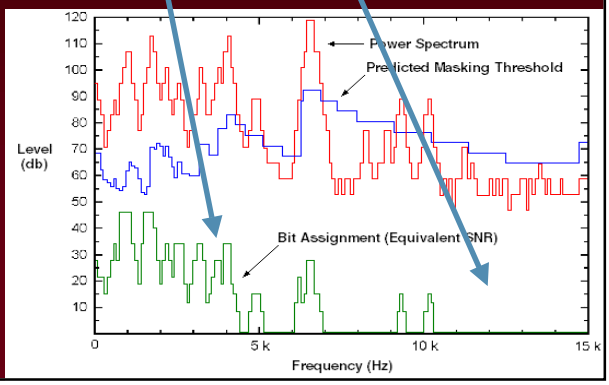
- Calculate masking thresholds
  - "Perceptual model"



Davidson et al., 1994

How: Noise Allocation

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



Davidson et al., 1994

How: More Tricks

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

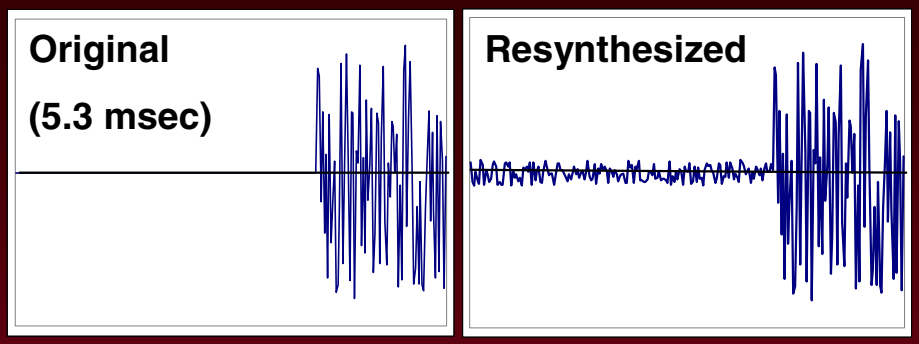
18

How: Coding Scheme

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

19

Artifacts: "Pre-echo"



- Quantization noise spread
- Noise components ≥ 1-2 msec before impulsive signal not masked
- Fix: Shorter window; wavelets (ePAC)

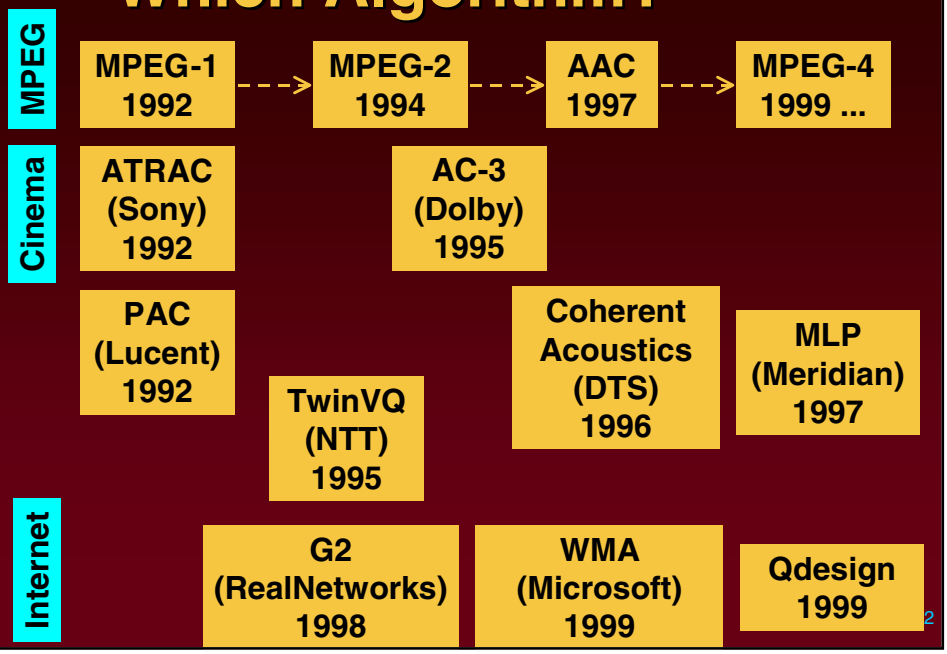
20

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

21

Which Algorithm?



MPEG Family

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

23

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 (!)

24

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

25

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")

26

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"

27

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

28

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

29

In the Future…

Photo credit: Dr. Richard O. Duda, SJSU

30

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)

31