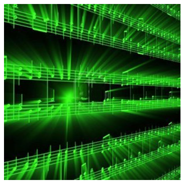
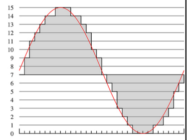
Penn Engineering ESE





Lecture #6 – Psychoacoustic Model/Compression/MP3

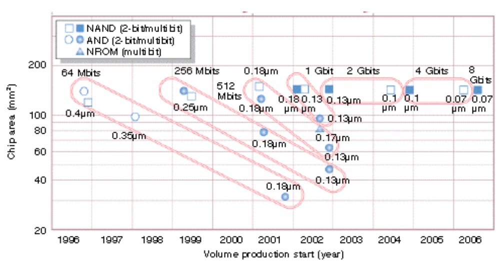
ESE 150 – Digital Audio Basics

ESE150 Spring 2018

First MP3 Player

- MpMan -- 1998
- SaeHan Information Systems
  - South Korea
- 32MB of Flash memory

Flash Scaling



PCM and CD Audio

- **PCM – Pulse Code Modulation**
  - CD-Quality Audio – state-of-the-art 1990s
    - Filtering/Sampling/Quantizing/Encoding using ADC
    - DAC/Reconstruction Filter
  - CD Quality Digital Audio uses PCM (uncompressed, lots of storage!)
    - 44,100 samples per second, each sample 16-bits
    - 1 sec. of music: 44,100 x 16bits = 705,600 bits or 86 kB
    - 60 seconds of music: 705,600 x 60 = 42,336,000 bits = 5167 kB = 5 MB
    - 3 minute song: 42,336,000 x 3 = 127,008,000 bits = 15 MB !
    - You want it in stereo??? 15 MB x 2 = 30 MB! (no compression!)

Observe

- **If we kept the CD Audio encoding format**
  - Could hold **one** song on the 1998 MpMan
  - (maybe 2 on the 64MB version)
- **For solid-state audio to be viable**
  - Needed more compact encoding for music

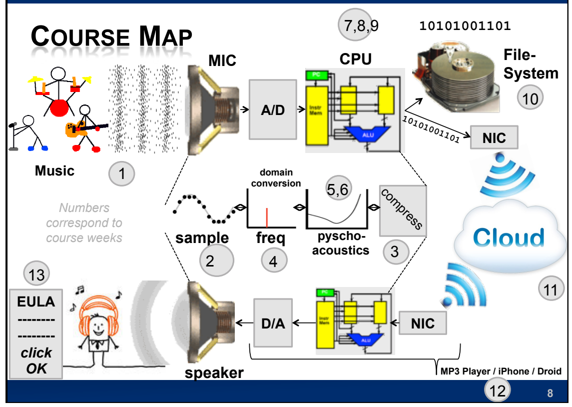
First MP3 Player

- **MpMan -- 1998**
- **SaeHan Information Systems**
  - South Korea
- **32MB of Flash memory**
- **Held 6 songs (MP3)**
- **(12 on 64MB version)**
- **3 years before Apple iPod**
  - October 2001
  - Initially hard disk
- **Diamond Rio later 1998**

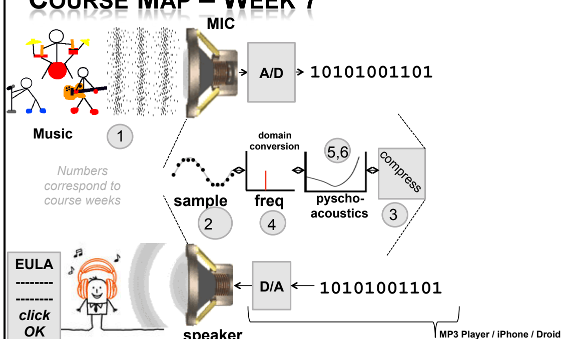
Lecture Topics

- Teaser
- Where are we?
- Preclass
- How do we take advantage of psychoacoustics in MP3
  Achieve this 6—12x reduction from CD Audio
  - Review Tricks
  - Formulate Optimization
  - Adaptation
- Next Lab
- References

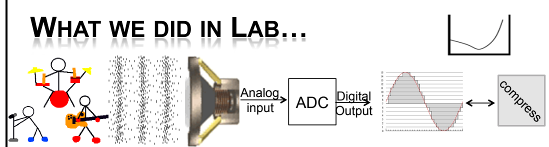
Course Map



Course Map – Week 7



What we did in Lab…



- Week 1: Converted Sound to analog voltage signal
  - a "pressure wave" that changes air molecules w/ respect to time
  - a "voltage wave" that changes amplitude w/ respect to time
- Week 2: Sampled voltage, then quantized it to digital sig.
  - Sample: Break up independent variable, take discrete 'samples'
  - Quantize: Break up dependent variable into n-levels (need 2^n bits to digitize)
- Week 3: Compress digital signal
  - Use even less bits without using sound quality!
- Week 4: Before we compress…
  - Put our 'digital' data into another form…BEFORE we compress…less stuff to compress!
- Week 5: Psychoacoustics
  - Measured limits of human hearing; measured masking

Preclass

Preclass

- 4 critical bands
- 10 frequencies
- 16b amplitude
- Preclass 1: Bits to represent (no further encode)?

PRECLASS

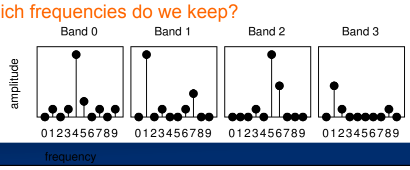
- **4 critical bands**
- **10 frequencies**
- **16b amplitude**
- **80b encoding budget**
- **Preclass 2: amplitude quantization necessary to achieve budget?**
- **Preclass 3: frequencies (reduced sampling rate) can keep to achieve budget?**

PRECLASS

- **4 critical bands**
- **10 frequencies**
- **16b amplitude**
- **80b encoding budget**
- **Preclass 4**
  - Bits to represent which frequency?
  - Bits to encode (frequency, amplitude) pair?
  - Number of (frequency,amplitude) pairs fit within budget?

PRECLASS

- **4 critical bands**
- **10 frequencies**
- **16b amplitude**
- **80b encoding budget**
- **Preclass 5**
  - Which frequencies do we keep?



PRECLASS

- **4 critical bands**
- **10 frequencies**
- **16b amplitude**
- **80b encoding budget**
- **Preclass 6: which likely to sound best?**
  - Amplitude quantization
  - Frequency quantization (reduce sampling rate)
  - Frequency selection based on masking

TRICKS FOR COMPRESSION

- **Quantization**
- **Sampling Rate / Frequency Quantization**
- **Critical Band Masking**
  - Selective frequency dropping
- **Other tricks?**

BACKGROUND

The MPEG-1 Standard

- **ISO (International Standards Organization)**
  - Looking for ways to reduce transmission requirements for digital video and audio (low bandwidth transmission of digital media)
  - 1988 – establishes a sub-committee of ISO:
    - *Moving* Picture Experts Group (MPEG)
    - Goal: Develop common standard for coding/compressing audio/video
      - To reduce size of data to transmit without sacrificing quality
      - Fraunhofer Institute and German University of Erlangen
        - Lots of basic research in Digital Audio Broadcast, tapped to be MPEG
  - Result: 1992: Finalized Standard called: MPEG-1 (Phase I)
    - 3 Parts: Audio/Video/System
    - Audio component: defined 3 layers: 1, 2, 3
      - Increasing levels of compression and complexity
      - MPEG-1, layer 3 achieves 12:1 compression ratio! *(for short MP3)*

The MPEG-1 Standard

- **MPEG-1, (3) Layers for Audio Coding:**

Complexity ↓

Coding	Ratio	Required bitrate
PCM CD Quality	1:1	1.4 Mbps
Layer I	4:1	384 kbps
Layer II	8:1	192 kbps
Layer III (MP3)	12:1	128 kbps

- **Why is PCM CD Quality 1.4 Mbps?**
  - Recall: 1 sec. of music: 44,100 x 16bits = 705,600 bits
  - Don't forget stereo (R/L): 2 x 706,600 = 1,413,200 (1.4Mbs)
    - Defines bandwidth requirements of network
  - Notice: 128 kbps was just about double modem speed in 1992
    - Enables transmission of audio (MP3) via modem! (Napster!)

The MPEG-1 Standard

- **How did MP3 achieve a 12:1 compression ratio?**
  - Psychoacoustics was the key
    - Becomes a scheme for lossy encoding
    - But we don't lose things human's can actually here!
    - Makes human a part of the "decompression" scheme
  - CD Audio PCM is lossless coding of digital audio
    - But human beings can't hear all sounds
  - MP3 Encoding is lossy coding scheme
    - But we don't lose things average human's can actually here!
    - We don't sacrifice high frequencies (sampling rate preserved)
    - The trick: drop quantization level when it doesn't matter!
      - The key is taking advantage of masking (frequency/temporal)
  - MP3 Combines: Psychoacoustics (lossy compression)
    - and Huffman coding (lossless compression)

They Kept on Going: MPEG-1, 2, 4, 7

MPEG is actually family of encoding standards for digital multimedia information
- *They all use Psychoacoustics to achieve high compression*

- **MPEG-1:** **a standard for storage and retrieval of moving pictures and audio on storage media (e.g., CD-ROM).**
  - *MPEG-1 Layer 3: (MP3)*
- **MPEG-2:** **standard for digital television, including high-definition television (HDTV), and for addressing multimedia applications.**
  - Advanced Audio Coding (AAC)
- **MPEG-4:** **a standard for multimedia applications, with very low bit-rate audio-visual compression for those channels with very limited bandwidths (e.g., wireless channels).**
- **MPEG-7:** **a content representation standard for information search**

Some Knowledge of Digital Audio Formats

- **.WAV & .AIFF**
  - Uncompressed PCM (lossless) CD-Audio Quality
- **.MPA**
  - MPEG-1 Layer 2
  - Compressed PCM using Psychoacoustics (lossy)
- **.MP3**
  - MPEG-1 Layer 3
  - Compressed PCM using Psychoacoustics (lossy)
- **.M4A (Audio only Apple's iTunes format)**
  - MPEG-4 Part 14 (MP4)
  - Compressed PCM using Psychoacoustics

Some Terminology

- **Psychoacoustics**
  - Research into how brain interprets sounds from ear
- **Perceptual Coding**
  - An audio encoding technique that takes advantage of psychoacoustics
  - CD delivers all recorded sounds to your ear…
    - …but your brain can't actually interpret them
  - Perceptual coded audio delivers sounds to your ear…
    - …that your brain can actually interpret!
- **CODEC**
  - Coder/Decoder
- **MP3 is a standard for a perceptual audio codec…**
  - That takes advantage of frequency/time masking to encode audio data!

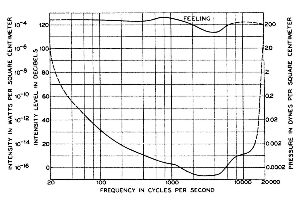
Optimizing Encoding

Knobs we can turn

- **Amplitude quantization**
- **Frequency quantization**
- **Frequencies kept (per critical band)**
  - Too soft
  - Masked
- **…and can perform lossless compression**

Bands vary in importance

- **Not equally sensitive across bands**
- **If quantize bands differently, where want finer resolution?**
- **Where tolerate more quantization?**



Knobs we can turn

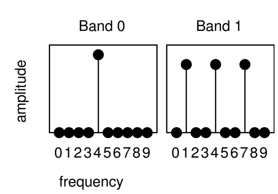
- **Amplitude quantization**
  - Per band
- **Frequency quantization**
  - Per band?
- **Frequencies kept (per critical band)**
  - Per band
- **…and can perform lossless compression**

Opportunities/Challenges

- **Trying to hit fixed bit rate, what challenge does lossless compression impose?**
  - Encounter many common frequencies?
  - Encounter many uncommon frequencies?

Opportunities/Challenges

- **What challenge/opportunities might these band spectra represent?**



Suggest

- May want to do something smarter than
  - Allocating fixed number of frequencies per band
  - Allocating fixed quantization to a band
- Like to adapt our encoding to the data
  - If more Huffman compressible, we get more frequencies
  - If fewer frequencies suffice for one band,
    - Allow more frequencies for another
    - …or allocate less quantization

Optimization Problem

- How fit in the resource constraints (128Kb/s) while maximizing goodness (sound quality)?
- Optimization problems central to engineering

Optimization Problem

- How fit in the resource constraints (128Kb/s) while maximizing goodness (sound quality)?
- Quantify bits used $\sum_{bands}\sum_{freq} Bits(freq)$
- Quantify goodness

$$\sum_{freq} Error(freq) \times W(freq)$$

Quantifying Bit cost

- Simple, fixed sample: Frequencies × Bits/freq
- Fixed frequencies per Band:
  - Bands × (Frequencies/Band) × Bits/freq
- Variable Frequencies per Band:

$$\sum_{bands}\left(frequencies(band) \times bits/freq\right)$$

- Variable Frequencies and quantization per Band:

$$\sum_{bands}\left(frequencies(band) \times bits(band)\right)$$

- Huffman means different bits/frequency $\sum_{bands}\sum_{freq} Bits(freq)$

Goodness/Sound Quality

- Error(freq) = |OrigFreq Amplitude – Encoded|
  - Whole OrigFreq if dropped
  - |OrigFreq-Quantize(OrigFreq,bits)| if quantized
- W(freq)
  - 0 if below hearing threshold
  - 0 if masked
  - Value between 0 and 1 if partially masked in critical band
  - Really depend on what already encoded $\sum_{freq} Error(freq) \times W(freq)$

Optimization Problem

- How fit in the resource constraints (128Kb/s) while maximizing goodness (sound quality)?
- Quantify bits used $\sum_{bands}\sum_{freq} Bits(freq)$
- Quantify goodness

$$\sum_{freq} Error(freq) \times W(freq)$$

Approach

$$\sum_{bands}\sum_{freq} Bits(freq)$$

$$\sum_{freq} Error(freq) \times W(freq)$$

- **Start with nothing**
- **Start with bitbudget**
- **While(bitbudget>0)**
  - Identify Largest Error component: Error(freq)×W(freq)
  - What do to reduce error?

Approach (Greedy)

- **Start with nothing**
- **Start with bitbudget**
- **While(bitbudget>0)**
  - Identify Largest Error component
  - Allocate some bits to reduce error
    - Add frequency
    - Add quantization bits to band
    - Pick one to most reduce the error

Approach (Refine Greedy)

- **Start with nothing**
- **Start with bitbudget**
- **While(bitbudget>0)**
  - Identify Largest Δerror/Δbits
- **Why might prefer?**
- **Large error, might take many bits to improve**
  - Maybe could be better spending those bits to fix many problems?

Adaptive Refinement

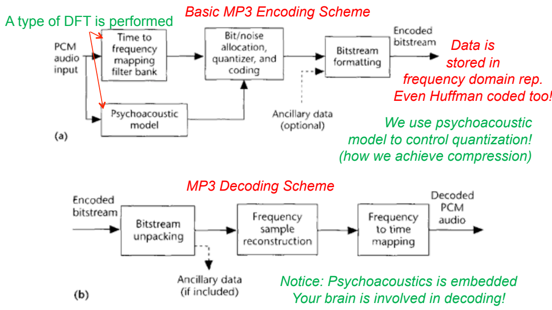
- **Rediscovering where to allocate everything every time may be laborious**
- **Maybe we can get close and adjust?**

Approach (Adaptive)

- **Start with budget guess**
  - Quantization in bands
  - Frequencies to keep in each band
- **Encode, compress**
- **What do if haven't used up all bits?**
- **What do if over budget?**

PERCEPTUAL CODING & MP3

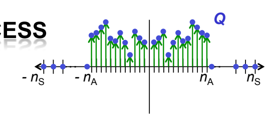
MP3 Flow Chart (Encoding/Decoding)



MP3 Encoding Process

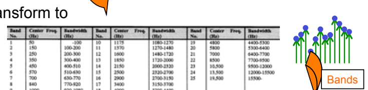
- **All MP3 files broken into "Frames"**
  - Each frame stores 1152 Audio Samples
  - Lasts for 26 ms
  - Frame also divided further into 2 "granuels"
    - Each granuel contains 576 samples
- **Bitrate determines size of each sample**
  - Increase the bitrate, increase the size of the frame!
- **Sampling rate also impacts frame size**
- **Formula:** $\frac{144 * bitrate}{samplefrequency} + Padding$ [bytes]
  - Example: frame size = 144 x 128kbs / 44.1kHz = 417

MP3 Encoding Process



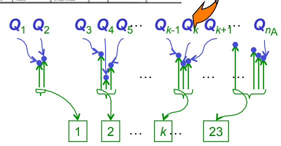
acquire & transform "frame"

- Bring in 1152 PCM samples
- Use DFT/MDCT to transform to frequency-domain



assign frequencies to bands

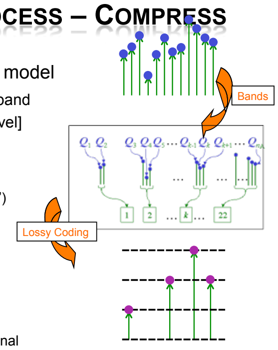
- use psychoacoustic model lookup to determine frequency bandwidth of each critical band
- separate samples into bands



MP3 Encoding Process – Compress

Compress based off masking model

- Goal: minimize bits per critical band
  - $C_{band\ k(j)}$ = Round[$Q_{band\ k(j)}$, Level]
- by appeal to "masking" models
  - $Q_{band\ k} = C_{band\ k} + D_{band\ k}$
- where distortion ("perceptual noise")
  - $D_{band\ k}$
- between retained signal
  - $C_{band\ k}$
- and actual signal
  - $Q_{band\ k}$
- should be "masked" by retained signal



Compress – Up close

E.g., Look at k[th] band

$Q_{band\ k} = (Q_{band\ k(1)}, \cdots, Q_{band\ k(m)})$

Determine masking paradigm:

- tone-masking-noise
- noise-masking-tone
- noise-masking-noise

E.g., for tone masker,

- Pick tone frequency, band k (j)
- at maximal amplitude in the band

Choose quantization level and compute compressed signal

- $C_{band\ k(j)}$ = Round[$Q_{band\ k(j)}$, Level]
- $C_{band\ k} = (0_{band\ k(1)}, \cdots, C_{band\ k(j)}, \cdots, 0_{band\ k(m)})$

Assess noise magnitude, $|D_{band\ k}|$

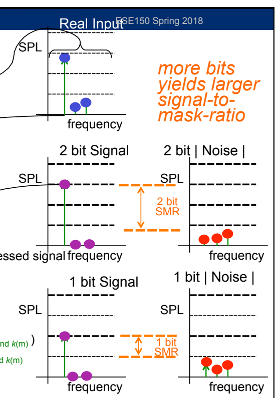
$D_{band\ k}$ =

$(Q_{band\ k(1)}, \cdots, Q_{band\ k(j)}, \cdots, Q_{band\ k(m)})$

$- (0_{band\ k(1)}, \cdots, C_{band\ k(j)}, \cdots, 0_{band\ k(m)})$

Use psychoacoustic model

- to determine whether compressed signal will mask the distortion noise for that band



Formal MP3 Perceptual Coding Algorithm

Commit to Observation Window

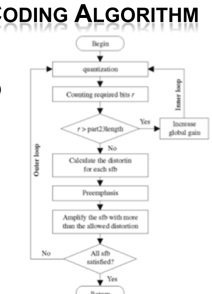
- "Long" (complex sound; frequency resolution)
- "Short" (transient sound; temporal resolution)

Estimate Perceptual Entropy

- Analyze Each Critical Band
- Characterize Masker
- Estimate Mask-to-Noise Threshold
- Update "bit reservoir"

Spectral Quantization/Coding Loop

- Allocate bits per critical band
- Quantize band to bits-allowed levels
- Run Huffmann & count actual bits



[You & Chen. *Multim.Tools & Appl.*, 40(3):341–359, 2008.]

Big Ideas

- **Can use pyschoacoustics to compress audio**
  - Eliminate portions of signal that human's don't notice
- **Optimization**
  - Identify Design Space (knobs)
  - Identify Costs and Constraints
  - Formulate quantitatively
  - Algorithms to approach
  - Iterative/adaptive approach
    - Deal with effects that aren't completely predictable

Learn More

- **Optimization –**
  - continuous mathematical optimization ESE204, ESE504, ESE605
  - discrete optimization CIS121, CIS320
- **Signal processing – ESE224**

Next Week

- **Monday start of 2 week lab:**
  - Perform perceptual compression
    - Convergence of everything seen first 6 weeks
  - Formal lab report on this 2 week lab
  - No weekly lab report Friday
- **Wednesday: midterm (class time)**
- **Office Hours:**
  - Cancel Wed. and Thursday evening
  - Add Tuesday 5-9pm
  - Wed. 1-3pm remain

Midterm – In Class

- **Closed book, no notes**
- **Calculators allowed**
- **4:35pm-5:50pm**
- **5% of grade**
  - prepare for final
- **Last year's final on syllabus**
  - Will be a bit different

Topics

- Data representation in bits
- Sounds waves
- Sampling
- Quantization
- Nyquist
- Lossy/lossless compression
- Common case
- Frequency domain
- Psychoacoustics
- Perceptual coding

References

Tutorials on Psychoacoustic Coding (in increasing order of abstraction and generality)

D. Pan, M. Inc, and I. L. Schaumburg. A tutorial on MPEG/audio compression. IEEE multimedia, 2(2):60–74, 1995.

Nikil Jayant, James Johnston, and Robert Safranek. Signal compression based on models of human perception. Proceedings of the IEEE, 81(10):1385–1422, 1993.

V. K. Goyal. Theoretical foundations of transform coding. IEEE Signal Processing Magazine, 18(5):9–21, 2001.

Lightweight Overview of MP3

Rassol Raissi. The theory behind mp3. Technical report, MP3' Tech, December 2002.

Scientific Basis of MP3 Coding Standard

J. D. Johnston. Transform coding of audio signals using perceptual noise criteria. IEEE Journal on selected areas in communications, 6(2):314–323, 1988.