CSc 461/561
Multimedia Systems
Audio coding

Jianping Pan

Spring 2015

Audio is difficult to *compress*

high

- Lossless: without "information" loss
  - e.g., LPAC, FLAC, Monkey's Audio
    - MPEG-4 audio lossless coding (ALS): ~2 C/R
  - and many more (e.g., Apple Lossless ALAC)
- Lossy: with information loss
  - MPEG audio layer 3 (MP3): ~12 C/R
- Or other ways to represent audio
  - music: MIDI; speech: synthesized voice (TTS)

bit-rate

low

* lossless after sampling and quantization

Lossless compression

- Why lossless compression?
  - to preserve audio quality (easy to decode too)
  - for further processing etc
    - “What is lost is not (fully) recoverable.”
- Why plain entropy encoding fails for audio?
  - equally likely “letters”; too many “words”
  - very low compression ratio (C/R): ~1
    - e.g., winzip, gzip, etc directly on audio streams

* lossless audio compression ratio: ~2

Lossless predictive coding

- Recall 64Kbps PCM vs 32Kbps ADPCM
  - Prediction! Prediction! Prediction!
- Correlation among consecutive samples!
  - residual = sample - prediction(last_samples)
- Correlation between (stereo) channels!
  - L, R => (L+R)/2, (L-R)/2
- Then attempt entropy encoding
  - code smaller values

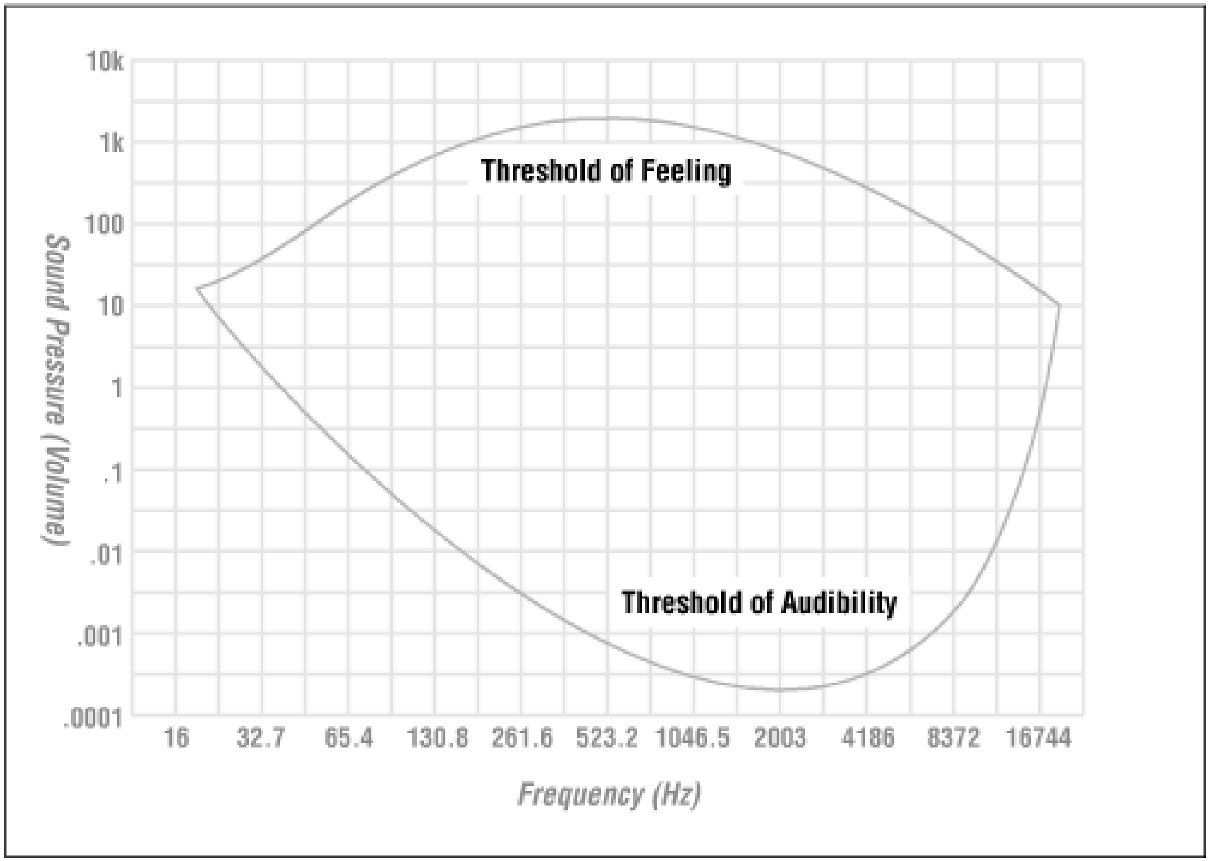
* the art of coding small values: e.g., differential, logarithm, etc

Lossy compression

- Why lossy compression?
  - to get higher compression ratio
  - without degrading audio quality too much
- Why lossy compression is possible?
  - audio is a wave of "waves"
  - not all waves are equal for *human* ears
  - wave: frequency, amplitude
- Perceptual audio encoding

* others: represent audio waves by connected line segments (LPC)

Not all waves are equal

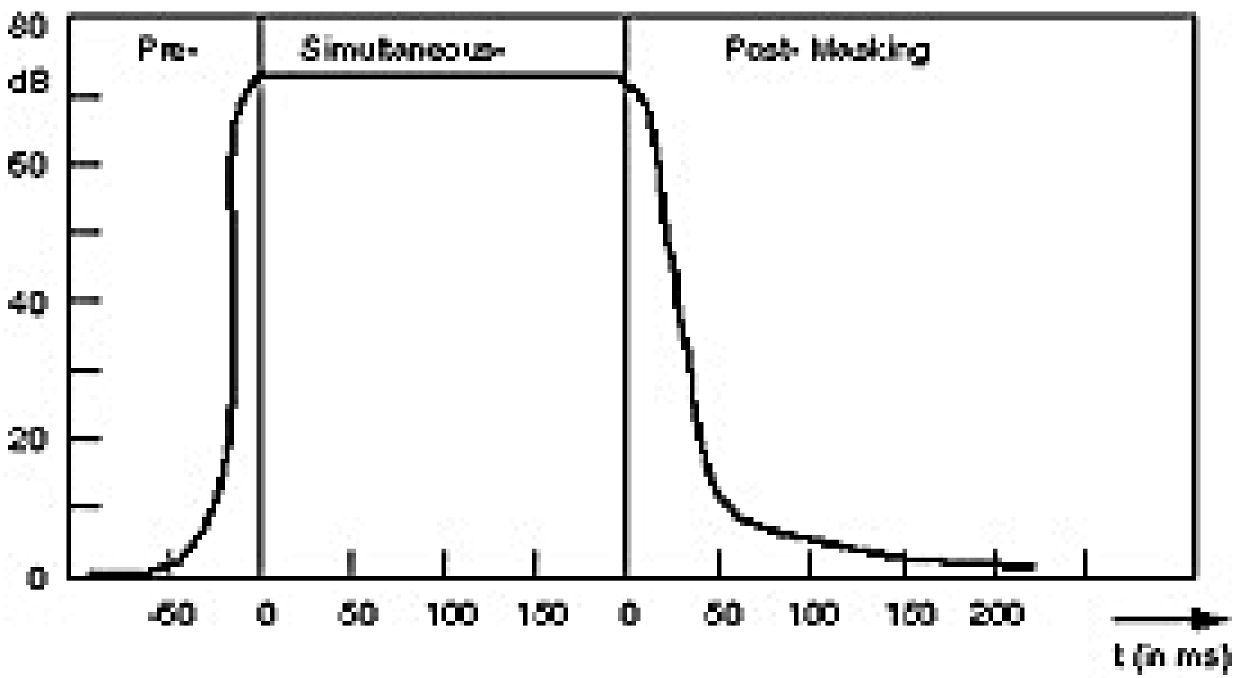


* too loud or too low to hear (for eardrums)

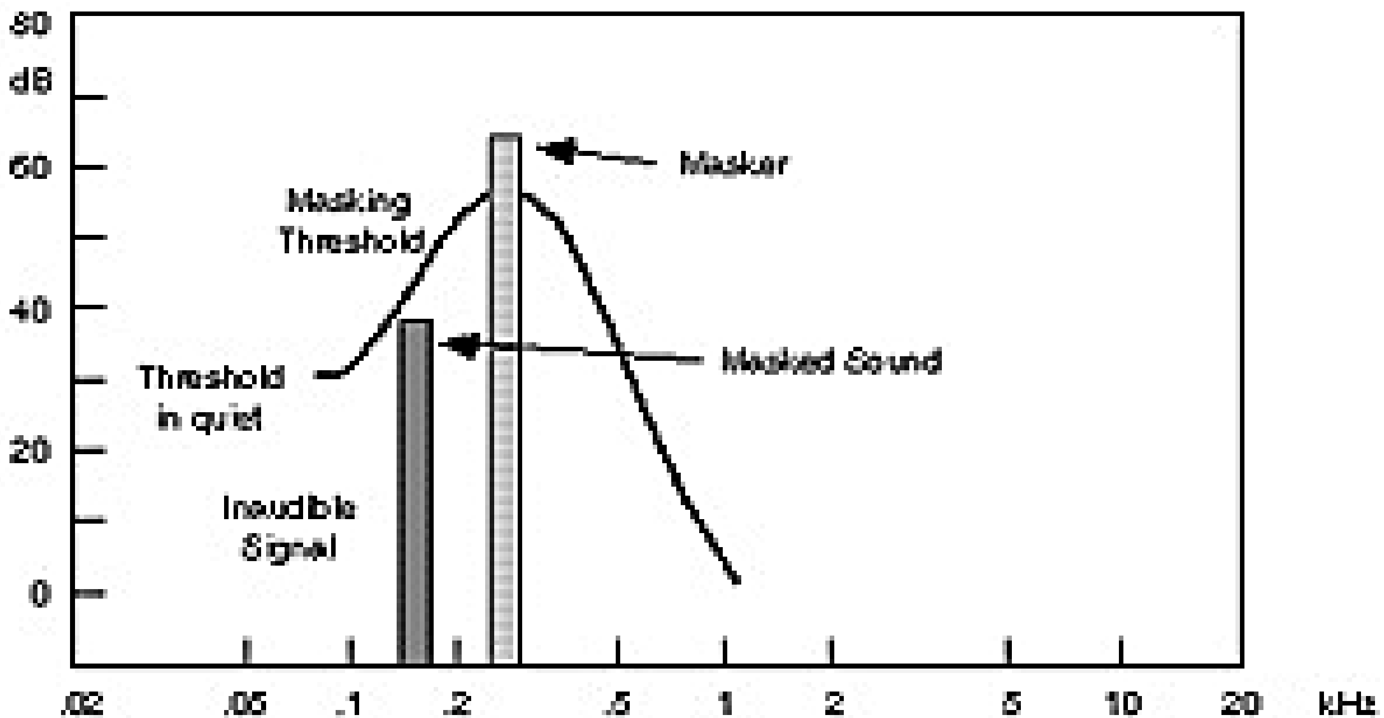
We only *hear* some waves

- Human psycho-acoustic model
  - frequency range: 20Hz - 20KHz
    - most sensitive: 2KHz - 4KHz
  - amplitude range: about 96 dB
- Temporal masking
  - "I cannot hear anything now; it was too loud!"
- Frequency masking
  - "I cannot hear this tone while that is around!"

Temporal masking



Frequency masking



* modified discrete cosine transform (MDCT)

MPEG-1 audio

- MPEG-1: VCD (VCR-like quality)
  - 1.2Mbps video (352x240, 30fps)
  - 256Kbps audio (mono or stereo)
- MPEG-1 audio to *approximate* CD quality
  - divide into 32 sub-bands (sub-band coding)
  - consider masking effects
    - discard a sub-band if it’s masked by neighbors
    - assign a smaller # of bits given the noise “floor”

MPEG-1 audio layers

- Layer 1: ~4 C/R; 384Kbps for CD quality
  - frequency masking
  - uniform sub-bands (12*32=384 samples/frame)
- Layer 2: ~6-8 C/R; 192-256Kbps; broadcast
  - also temporal masking (3 frames;1152 samples)
- Layer 3 (MP3): ~10-12 C/R; 112-128Kbps
  - both types of masking effect and stereo effect
  - *non-uniform* sub-band & quantization, Huffman coding

* MP3 vs Ogg Vorbis

MPEG-1 audio performance

- Mean Opinion Score (MOS): score 1~5
  - excellent (4.5); very good (4); good (3.6)
  - fair (3.1); poor (2.6); bad (1.0)

Layer	Target Bit-rate	Ratio	Quality at 64 kb/s	Quality at 128 kb/s	Theoretical Min. Delay
Layer 1	192 kb/s	4:1	---	---	19 ms
Layer 2	128 kb/s	6:1	2.1 to 2.6	4+	35 ms
Layer 3	64 kb/s	12:1	3.6 to 3.8	4+	59 ms

* non-uniform sampling and quantization

MPEG-2 audio

- MPEG-2: DVD (HDTV quality)
  – e.g., DVD movie: 10Mbps
- MPEG-2 (backward compatible) audio
  – mechanisms similar to MPEG-1 audio
  – more sampling rates: <u>16/22/24</u>/32/44/48KHz
  – expanded range of data rates: 8~640Kbps
    - MPEG-1 audio: 32~448Kbps
  – support 5.1/7.1-channel (MPEG-1 audio: 2)

* surround sound systems

Advanced Audio Coding (AAC)

- Not backward compatible with MPEG-1 audio
- MPEG-2 AAC
  - 8~96KHz sampling rate (MP3: 32-48KHz)
  - up to 48 main channels
  - data rate: up to 576Kbps
    - CD quality: AAC 96Kbps ~ 128Kbps MP3
- MPEG-4 AAC: LC/HE/SSR-AAC
  - e.g., iPod, PSP

* Apple iTune; Google Youtube; etc

Voice codecs

- Telephone (corded, cordless, mobile)
  – ITU-T: G.711 64Kbps (PCM), G.721/6 32Kbps (ADPCM);G.728 16Kbps(CELP),G.729 8Kbps
  – GSM:6.5~13Kbps(LPC);4.75~12.2Kbps(AMR)
    - voice detection, discontinuous TX, comfort noise
- Internet (VoIP, music streaming, etc)
  – iLBC (low bitrate): 15Kbps; iSAC: 10~32Kbps
  – SILK: 8~24KHz, 6~40Kbps (used in Skype)
  – Opus: 8~48KHz, 6~512Kbps; SILK, CELT

* more at multimedia delivery

This lecture

- Multimedia manipulation
  - audio compression
    - lossless compression
      - predictive coding
    - lossy compression
      - perceptual coding: frequency/temporal masking
- Explore further
  - FLAC: http://flac.sourceforge.net/ => xiph.org
  - http://www.mpeg.org/MPEG/audio

Next lecture

- Multimedia manipulation
  – image compression [Ref: Li&Drew Chap 9]
    - JPEG [9.1-3]