digitalmultimedia.org

Sound

Digital Multimedia, 2nd edition

Nigel Chapman & Jenny Chapman

Chapter 9

The Nature of Sound

- Conversion of energy into vibrations in the air (or some other elastic medium)
- Most sound sources vibrate in complex ways leading to sounds with components at several different frequencies
- *Frequency spectrum* – relative amplitudes of the frequency components
- Range of human hearing: roughly 20Hz–20kHz, falling off with age

Waveforms

- Sounds change over time
  - e.g. musical note has attack and decay, speech changes constantly
- Frequency spectrum alters as sound changes
- *Waveform* is a plot of amplitude against time
  - Provides a graphical view of characteristics of a changing sound
  - Can identify syllables of speech, rhythm of music, quiet and loud passages, etc

Digitization – Sampling

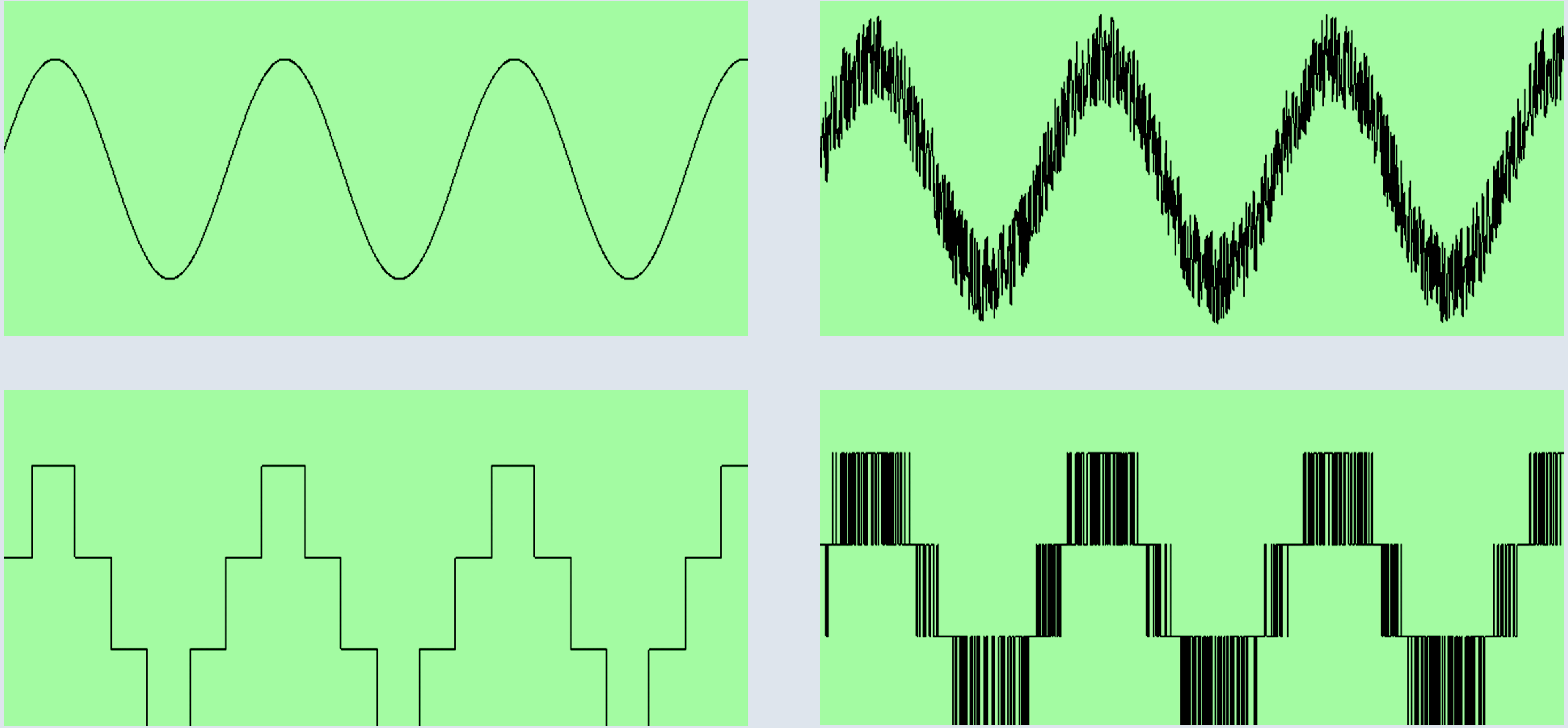
- Sampling Theorem implies minimum rate of 40kHz to reproduce sound up to limit of hearing
  - CD: 44.1kHz
    - Sub-multiples often used for low bandwidth – e.g. 22.05kHz for Internet audio
  - DAT: 48kHz

  (Hence mixing sounds from CD and DAT will require some resampling, best avoided)

Digitization – Quantization

- 16 bits, 65536 quantization levels, CD quality
- 8 bits: audible quantization noise, can only use if some distortion is acceptable, e.g. voice communication
- *Dithering* – introduce small amount of random noise before sampling
    - Noise causes samples to alternate rapidly between quantization levels, effectively smoothing sharp transitions

Undersampling & Dithering

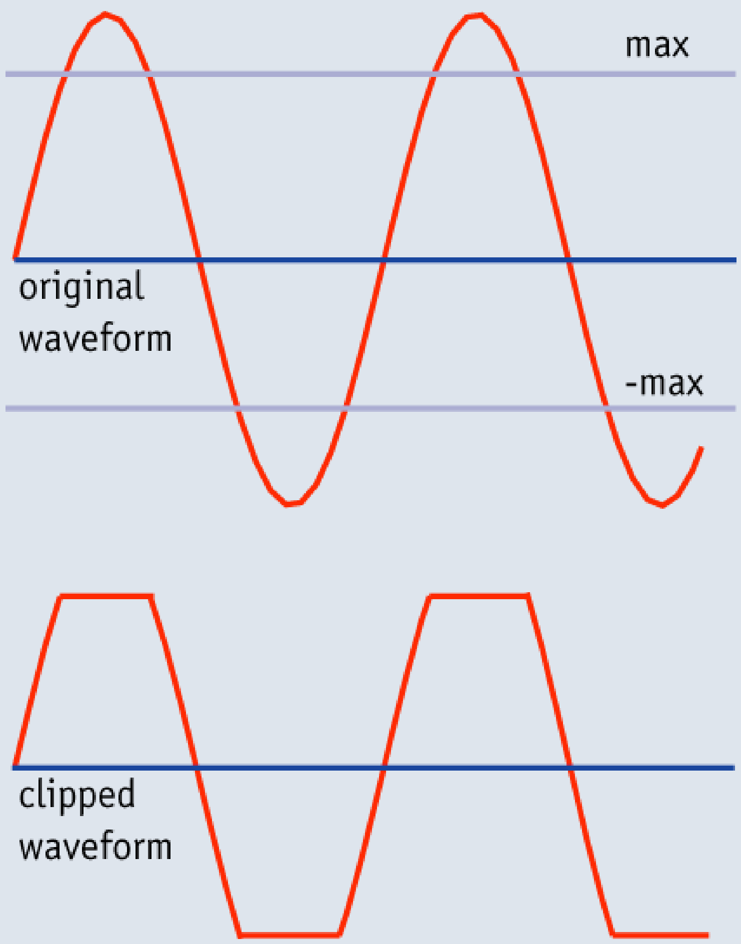


Data Size

- Sampling rate *r* is the number of samples per second
- Sample size *s* bits
- Each second of digitized audio requires *rs*/8 bytes
- CD quality: *r* = 44100, *s* = 16, hence each second requires just over 86 kbytes (k=1024), each minute roughly 5Mbytes (mono)

Clipping

- If recording level is set too high, signal amplitude will exceed maximum that can be recorded, leading to unpleasant distortion
- But if level is set too low, dynamic range will be restricted



Sound Editing

- Timeline divided into *tracks*
- Sound on each track displayed as a waveform
- 'Scrub' over part of a track e.g. to find pauses
- Cut and paste, drag and drop
- May combine many tracks from different recordings (mix-down)

Effects and Filters

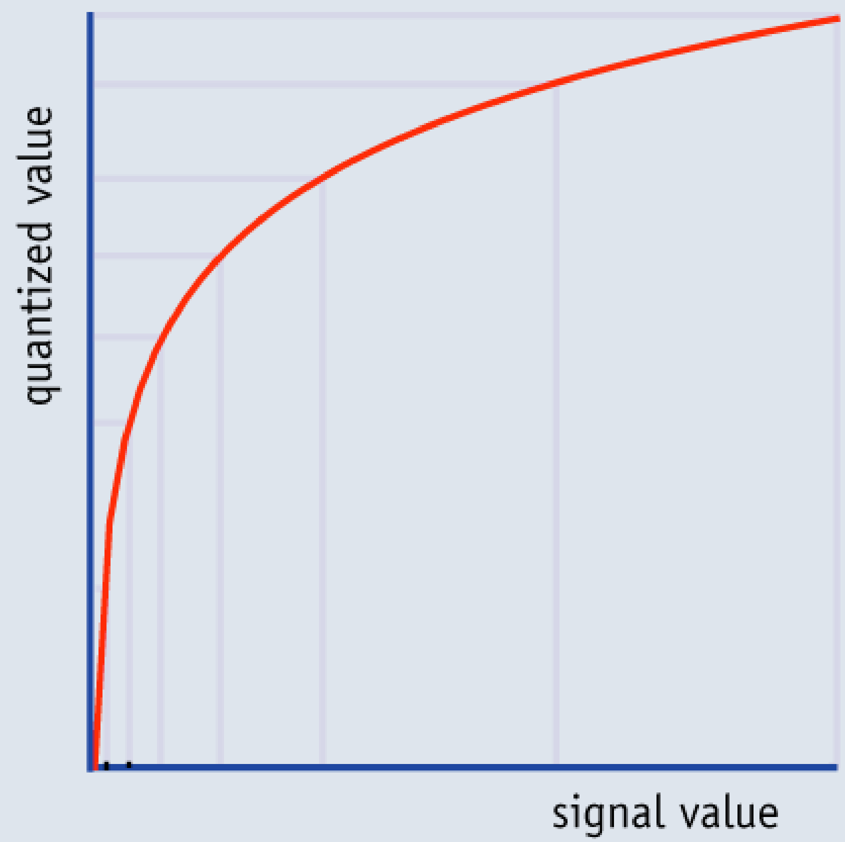
- Noise gate
- Low pass and high pass filters
- Notch filter
- De-esser
- Click repairer
- Reverb
- Graphic equalizer
- Envelope Shaping
- Pitch alteration and time stretching
- etc

Compression

- In general, *lossy* methods required because of complex and unpredictable nature of audio data
- CD quality, stereo, 3-minute song requires over 25 Mbytes
  - Data rate exceeds bandwidth of dial-up Internet connection
- Difference in the way we perceive sound and image means different approach from image compression is needed

Companding

- Non-linear quantization
- Higher quantization levels spaced further apart than lower ones
- Quiet sounds represented in greater detail than loud ones
- -law, A-law



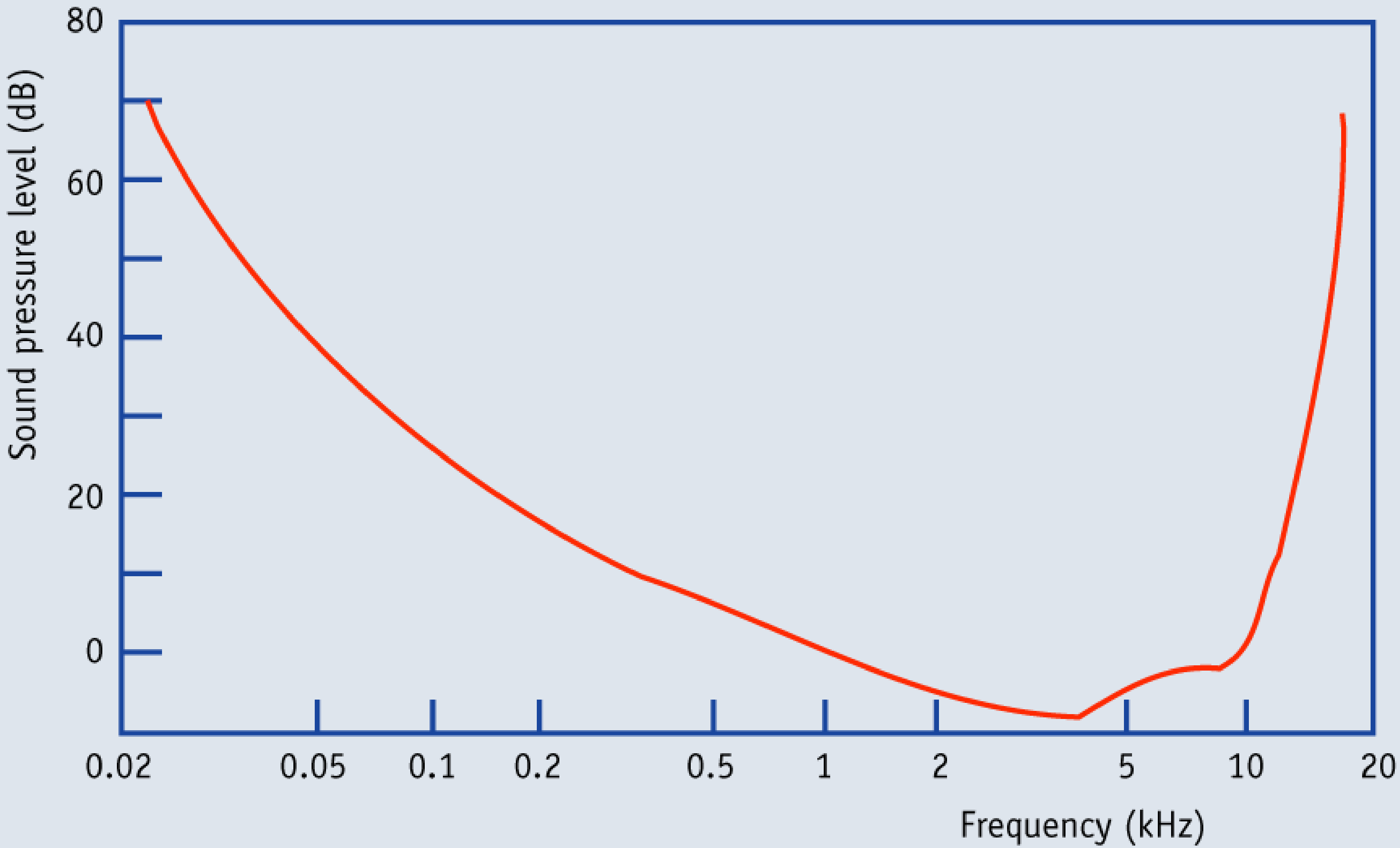
ADPCM

- Differential Pulse Code Modulation
  - Similar to video inter-frame compression
  - Compute a predicted value for next sample, store the difference between prediction and actual value
- *Adaptive* Differential Pulse Code Modulation
  - Dynamically vary step size used to store quantized differences

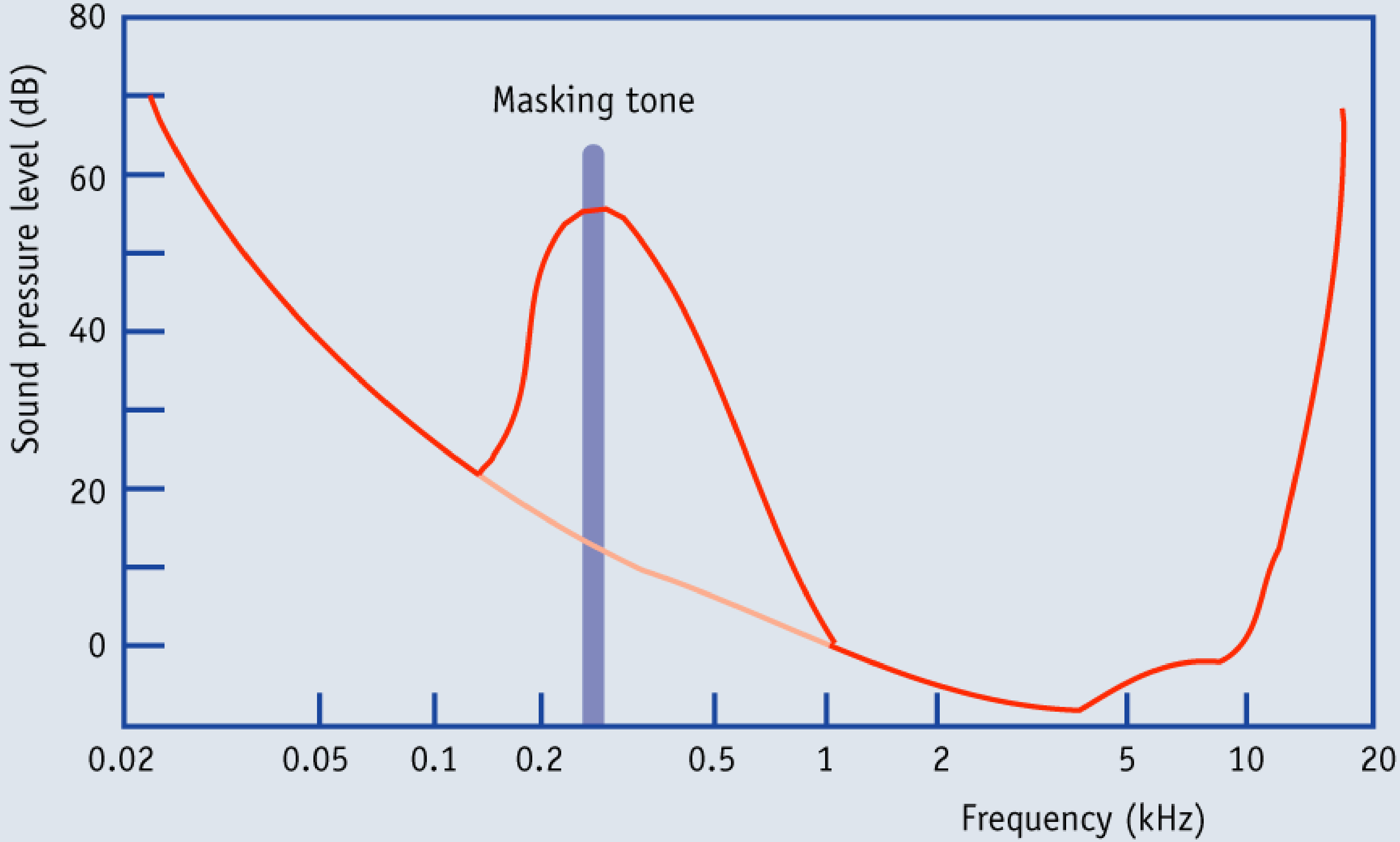
Perceptually-Based Compression

- Identify and discard data that doesn't affect the perception of the signal
  - Needs a *psycho-acoustical model*, since ear and brain do not respond to sound waves in a simple way
- *Threshold of hearing* – sounds too quiet to hear
- *Masking* – sound obscured by some other sound

The Threshold of Hearing



Masking



80
60
40
20
0
Sound pressure level (dB)
Masking tone
0.02
0.05
0.1
0.2
0.5
1
2
5
10
20
Frequency (kHz)

Compression Algorithm

- Split signal into bands of frequencies using filters
  - Commonly use 32 bands
- Compute *masking level* for each band, based on its average value and a psycho-acoustical model
  - i.e. approximate masking curve by a single value for each band
- Discard signal if it is below masking level
- Otherwise quantize using the minimum number of bits that will mask quantization noise

MP3

- *MPEG Audio, Layer 3*
- Three *layers* of audio compression in MPEG-1 (MPEG-2 essentially identical)
- Layer 1 →Layer 3, encoding proces increases in complexity, data rate for same quality decreases
  - e.g. Same quality 192kbps at Layer 1, 128kbps at Layer 2, 64kbps at Layer 3
- 10:1 compression ratio at high quality
- Variable bit rate coding (VBR)

AAC

- *Advanced Audio Coding*
- Defined in MPEG-2 standard, extended and incorporated into MPEG-4
- Not backward compatible with earlier standards
- Higher compression ratios and lower bit rates than MP3
- Subjectively better quality than MP3 at the same bit rate

Audio Formats

- Platform-specific file formats
  - AIFF, WAV, AU
- Multimedia formats used as 'container formats' for sound compressed with different codecs
  - QuickTime, Windows Media, RealAudio
- MP3 has its own file format, but MP3 data can be included as audio tracks in QuickTime movies and SWFs

MIDI

- *Musical Instruments Digital Interface*
- Instructions about how to produce music, which can be interpreted by suitable hardware and/or software
    - cf. vector graphics as drawing instructions
- Standard protocol for communicating between electronic instruments (synthesizers, samplers, drum machines)
- Allows instruments to be controlled by hardware or software *sequencers*

MIDI and Computers

- MIDI interface allows computer to send MIDI data to instruments
- Store MIDI sequences in files, exchange them between computers, incorporate into multimedia
- Computer can synthesize sounds on a sound card, or play back samples from disk in response to MIDI instructions
  - Computer becomes primitive musical instrument (quality of sound inferior to dedicated instruments)

MIDI Messages

- Instructions that control some aspect of the performance of an instrument
- *Status byte* – indicates type of message
- 2 *data bytes* – values of parameters
  - e.g. Note On + note number (0..127) + key velocity
- Running status – omit status byte if it is the same as preceding one

General MIDI

- Synths and samplers provide a variety of *voices*
- MIDI Program Change message selects a new voice, but mapping from values to voices is not defined in the MIDI standard
- General MIDI (addendum to standard) specifies 128 standard voices for Program Change values
  - Actually GM specifies voice *names*, no guarantee that identical sounds will be produced on different instruments