Introduction to Multimedia Computing

Audio Media Type

Topics

- Properties of Audio Media
- Audio Data Compression
  - Motivation
  - Redundancy
  - Redundant Audio Data
  - MPEG Layer 3 (mp3) standard

Properties of Sound (review)

- The **amplitude** measures the relative loudness of the sound, which is the distance between a valley and a crest.
- The **amplitude** determines the volume of the sound.
- The unit of measurement of volume is a **decibel**.

Sample Sound Amplitudes

Source	Intensity Level
Threshold of Hearing (TOH)	0 dB
Rustling Leaves	10 dB
Whisper	20 dB
Normal Conversation	60 dB
Busy Street Traffic	70 dB
Vacuum Cleaner	80 dB
Large Orchestra	98 dB
Walkman at Maximum Level	100 dB
Front Rows of Rock Concert	110 dB
Threshold of Pain	130 dB
Military Jet Takeoff	140 dB
Instant Perforation of Eardrum	160 dB

Properties of Sound (Cont.)

- **Frequency**: The number of peaks that happen in one second is the frequency.
- Another term associated with frequency is **pitch**. If an object oscillates rapidly (high frequency), it creates a "**high-pitched**" sound (treble).
- A low-frequency sound on the other hand is produced by an object that vibrates slowly, such as the thicker strings of a piano or guitar (bass).

Properties of Sound (Cont.)

- **Wavelength:** Wavelength is the distance from the midpoint of one crest to the midpoint of the next crest.
- Wavelength is represented by the symbol λ

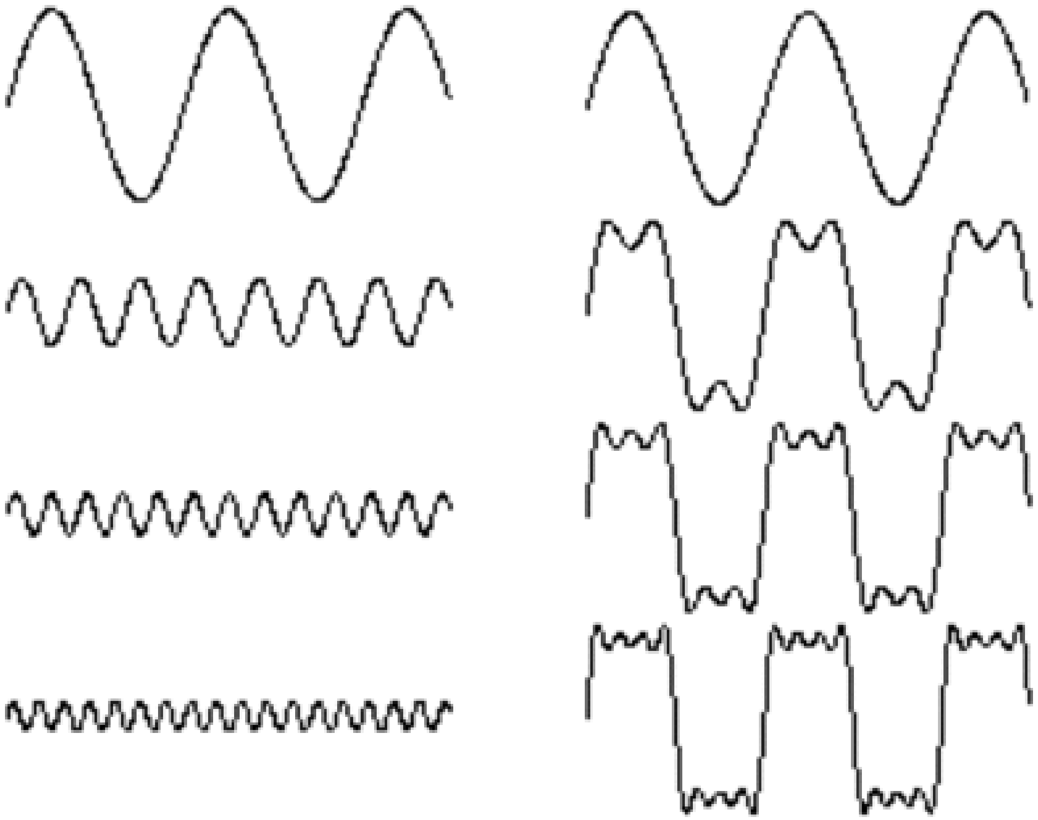
- **Bandwidth:** Bandwidth is defined as the difference between the highest and the lowest frequency contained in a signal.
- A signal which spans the range of 200-3200 Hz has a bandwidth (BW) of:

  BW= 3200-200 = 3000 Hz

Harmonics

- Few objects produce sound of a single frequency. Most musical instruments, for example, generate multiple frequencies for each note.
- The combinations of frequencies generated by an instrument are known as the **timbre**.
- A timbre consists of a fundamental or main frequency and other minor frequencies known as **overtones** or **harmonics.**

Harmonics of a Compound Wave



Velocity of Sound

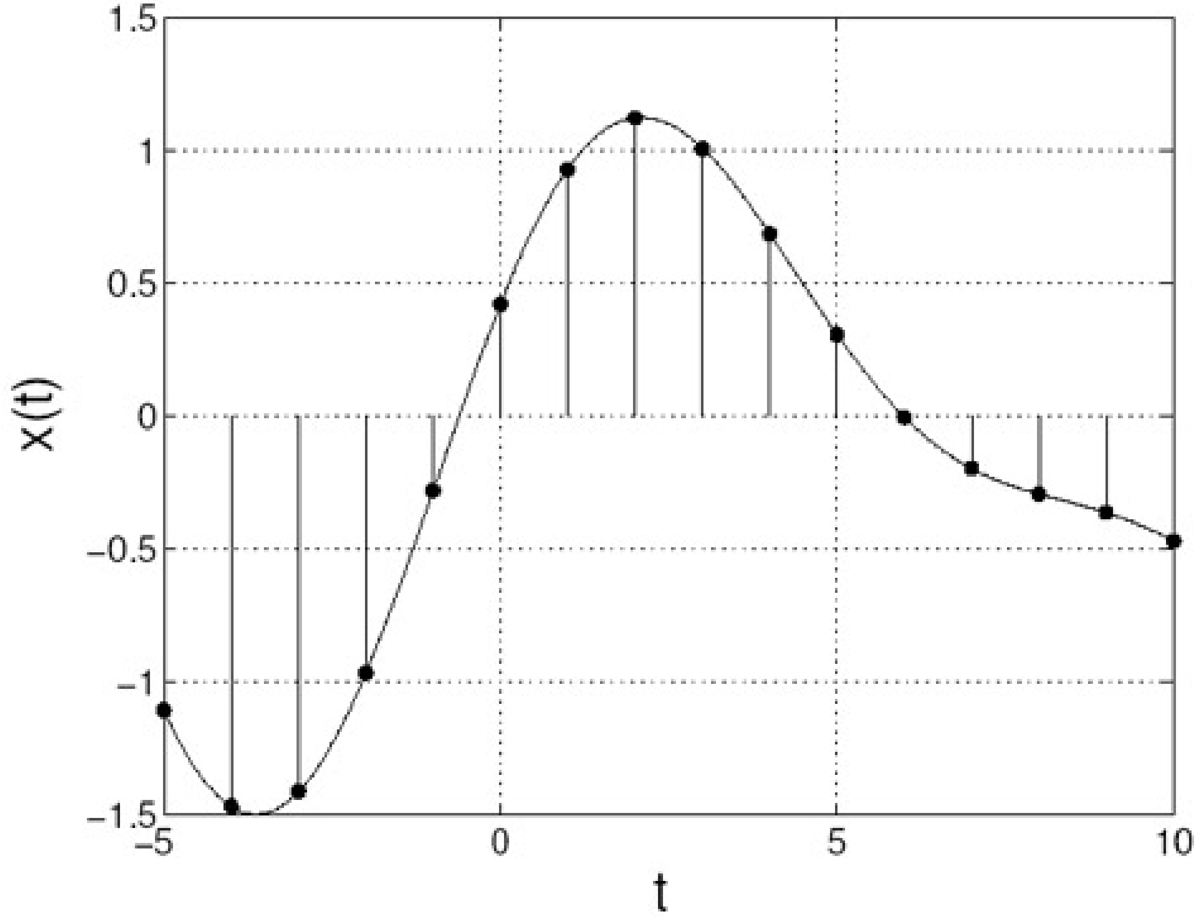
- The velocity of sound may be found directly by measuring the time required for the waves to travel a measured distance.
- The velocity varies greatly with the medium through which it travels.

Medium	Meters/second
Air	331.5
Hydrogen	1270
Carbon dioxide	258
Water	1450
Iron	5100
Glass	5500

Sampling

- Audio signal is an analog signal
- Computers store digital signal
- To convert analog signal into digital signal we have to do sampling

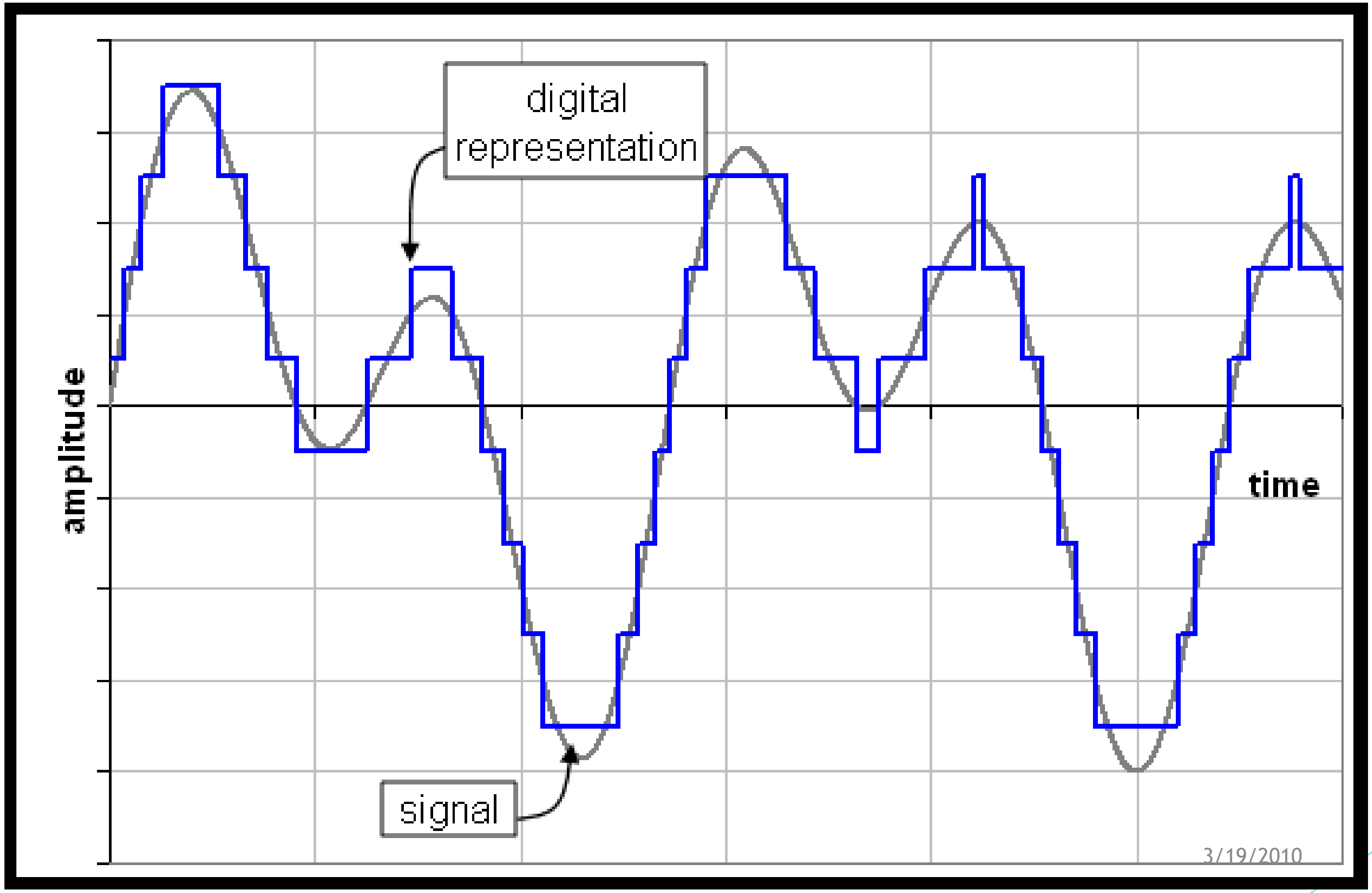
Sampling



Quantization

- Sample values can be in a large range and include real numbers.
- To store sample values with less number of bits quantization is used.
- In quantization the values in a range are mapped to a single number

Quantization



Sample Audio Data

- 20,000 Hz sound, sampled at 40,000 sample/second
- Quantization level: 256 (8 bits)
- 40,000 * 1 bytes /sec data

- With 65,536 (16 bits) quantization levels, data size will be 40,000 * 2 = 80,000 bytes/second

Fidelity

- Fidelity is defined as the closeness of the recorded version to the original sound. In the case of digital speech, it depends on the number of bits per sample (quantization) and the sampling rate.

- A really high-fidelity (hi-fi) recording takes up a lot of memory space (176.4 Kb for every second of audio of stereo quality sampled at 16 bits, 44.1 kHz per channel).

- Fortunately for most computer multimedia applications, it is not necessary to have very high fidelity sound.

Sound Formats and Settings

- Mono Recording:

  File size = Sampling rate x duration of recording in seconds x (bits per sample/8) x 1

- Stereo Recording:

  File size = Sampling rate x duration of recording in seconds x (bits per sample/8) x 2

Sound Quality

Sampling rate (kHz)	Bit-resolution (bits)	Stereo/Mono	Bytes needed for one minute (MB)	Comments
44	16	Stereo	10.5	CD-quality sound.
44	16	Mono	5.25	Good quality for voice-overs.
44	8	Stereo	5.25	Good quality for playback.
44	8	Mono	2.6	Good quality for recording a mono audio source.
22	16	Stereo	5.25	Good quality reproduction but not CD quality.
22	16	Mono	2.6	OK for narration.
22	8	Stereo	2.6	Good for stereo recording when playback equipment quality is low.
22	8	Mono	1.3	Sounds like good AM radio quality.
11	16	Stereo	2.64	No advantage in using stereo.
11	16	Mono	1.32	Sounds muffled. The lowest standard you should use, unless you are using for telephony applications.

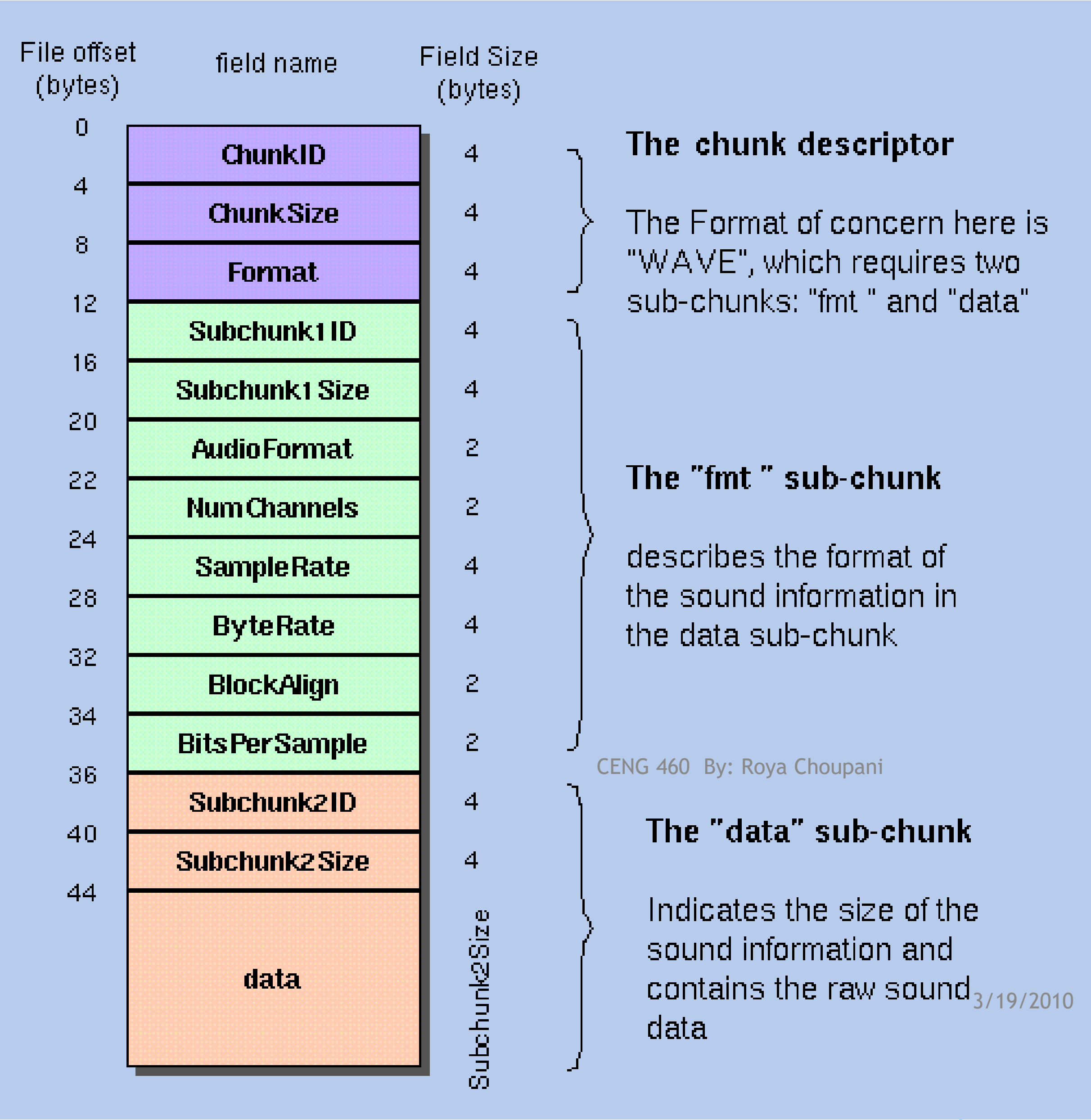
Audio Representation Standards

- *wav* - standard audio file format used mainly in Windows PCs. Commonly used for storing uncompressed, CD-quality sound files.
- *mp3* - the MPEG Layer-3 format is the most popular format for downloading and storing music.
- *au* - the standard audio file format used by Sun, Unix and Java. The audio in *au* files can be uncompressed or compressed.

WAV Audio Standard

- The WAV file format is a Microsoft specification for the storage of multimedia files.
- A WAV file starts out with a file header followed by a sequence of data chunks.
- A WAV file consists of two sub-chunks
  - a "fmt " chunk specifying the data format
  - a "data" chunk containing the actual sample data.

WAV File Format



The WAVE format header

Length	Name	Description
4	**ChunkID**	Contains the letters "RIFF" in ASCII form
4	**ChunkSize**	This is the size of the entire file in bytes minus 8 bytes for the two fields not included in this count: ChunkID and ChunkSize.
4	**Format**	Contains the letters "WAVE"

The WAVE format header

- The "WAVE" format consists of two subchunks: "fmt " and "data":
- The "fmt " subchunk describes the sound data's format:
- The "data" subchunk contains the size of the data and the actual sound:

Length	Name	Description
4	**Subchunk1ID**	Contains the letters "fmt "
4	**Subchunk1Size**	16 for uncompressed case. This is the size of the rest of the Sub-chunk which follows this number.
2	**AudioFormat**	Uncompressed = 1 , Values other than 1 indicate some form of compression.
2	**NumChannels**	Mono = 1, Stereo = 2, etc.
4	**SampleRate**	8000, 44100, etc.
4	**ByteRate**	== SampleRate * NumChannels * BitsPerSample/8
2	**BlockAlign**	== NumChannels * BitsPerSample/8
2	**BitsPerSample**	8 bits = 8, 16 bits = 16, etc.
2	**ExtraParamSize**	if uncompressed, then doesn't exist
X	**ExtraParams**	space for extra parameters
4	**Subchunk2ID**	Contains the letters "data"
4	**Subchunk2Size**	== NumSamples * NumChannels * BitsPerSample/8 This is the number of bytes in the data.
*	**Data**	The actual sound data.

Need for Compression

- Uncompressed digital CD-quality audio signals include a large amount of data and are therefore not suited for storage and transmission.
- The need to reduce this amount without any noticeable quality loss was stated in the late 80ies by the International Organization for Standardization (ISO).

Data Redundancy

- Redundancy in information theory is the number of bits used to transmit a message minus the number of bits of actual information in the message.

- Informally, it is the amount of wasted "space" used to transmit certain data

Data Compression

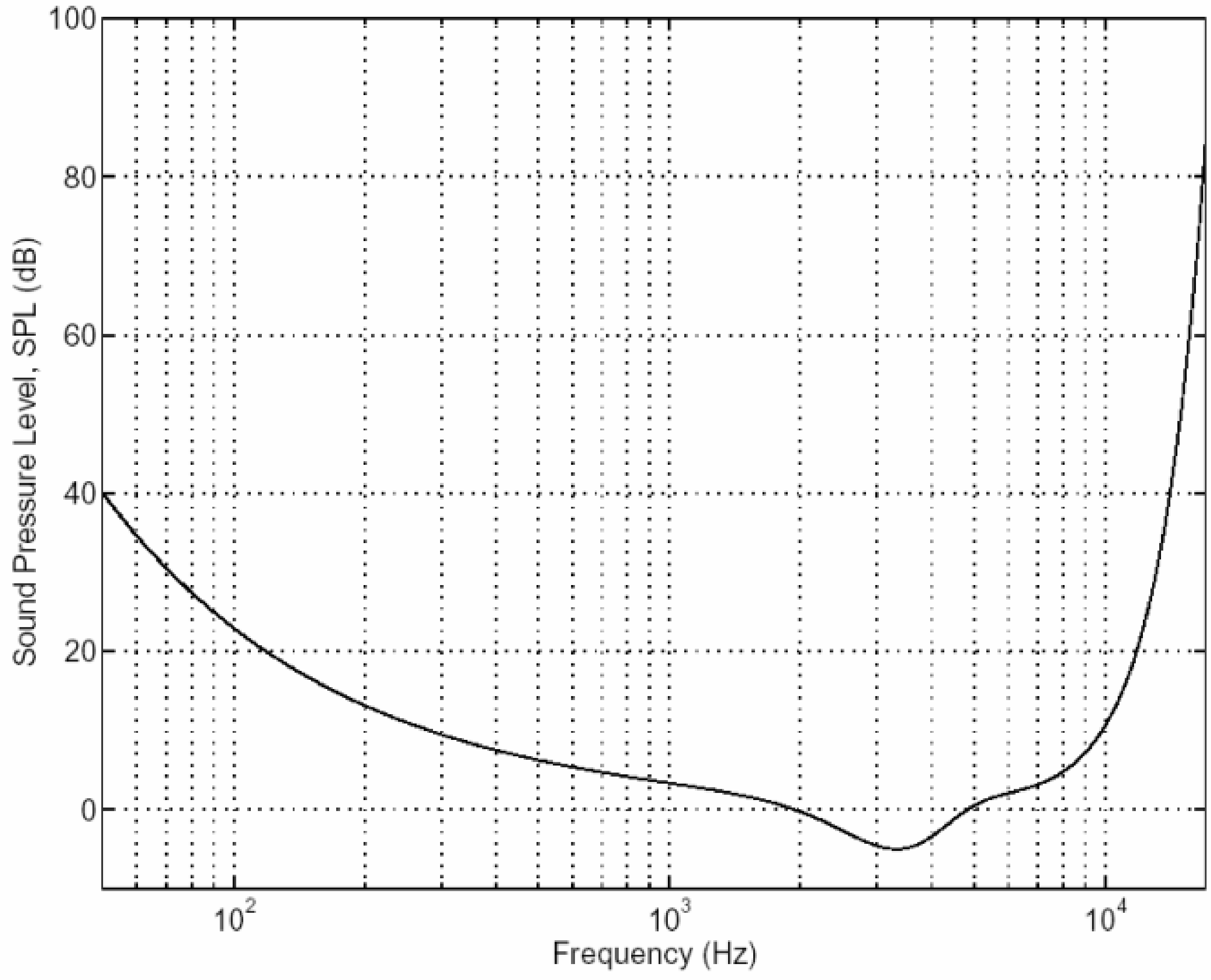
- Data compression or source coding is the process of encoding information using fewer bits than an un-encoded representation.

Redundancy Types in Audio

Psycho-acoustics & Perceptual Coding

- Psycho-acoustics is the research where we aim to understand how the ear and brain interact as various sounds enter the ear.
  - A middle aged man will not hear much above 16 kHz.
  - Frequencies ranging from 2 kHz to 4 kHz are easiest to perceive, they are detectable at a relatively low volume.
  - As the frequencies changes towards the ends of the audible bandwidth, the volume must also be increased for us to detect them

The absolute threshold of hearing



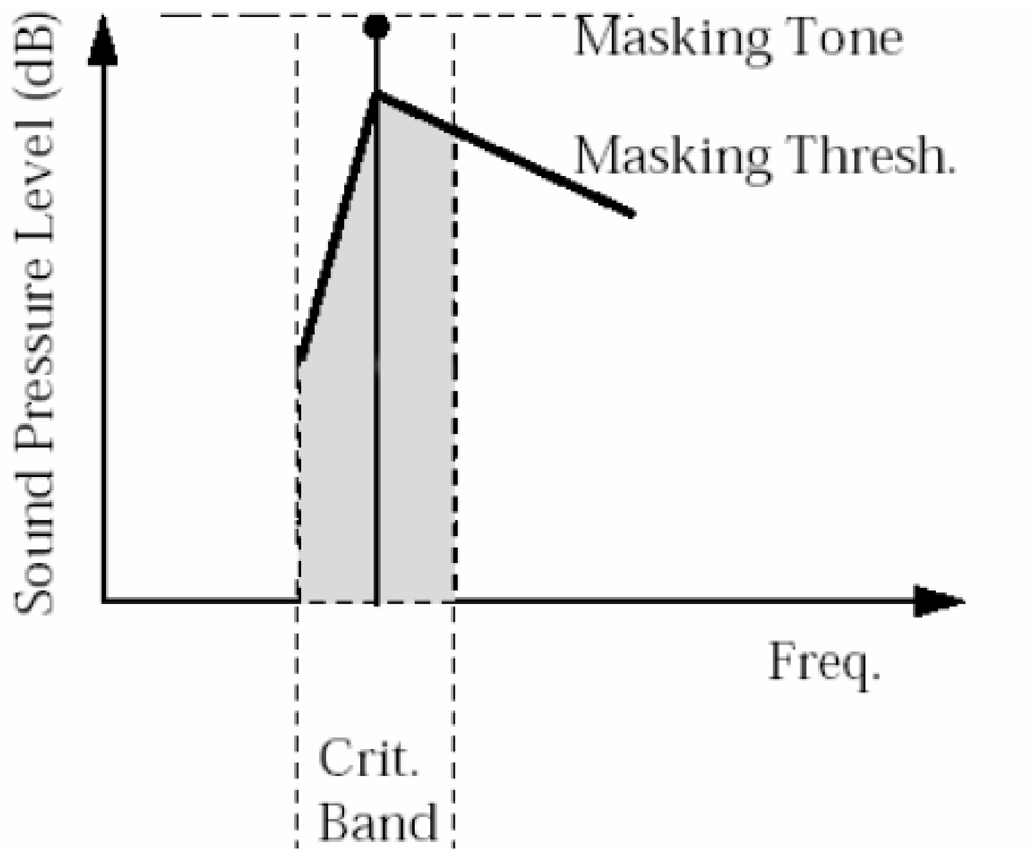
Masking

- While playing a CD it is impossible to percept all data reaching your ears, so there is no point in storing the part of the music that will be inaudible.
- The process that makes certain samples inaudible is called masking.
- There are two masking effects that the perceptual codec need to be aware of;
  - simultaneous masking
  - temporal masking.

Simultaneous Masking

- Experiments have shown that the human ear has 24 frequency bands.
- Frequencies in these so called critical bands are harder to distinguish by the human ear.
- If there is a dominant tonal component present in an audio signal it will introduce a masking threshold that will mask out frequencies in the same critical band
- This frequency-domain phenomenon is known as simultaneous masking, which has been observed within critical bands

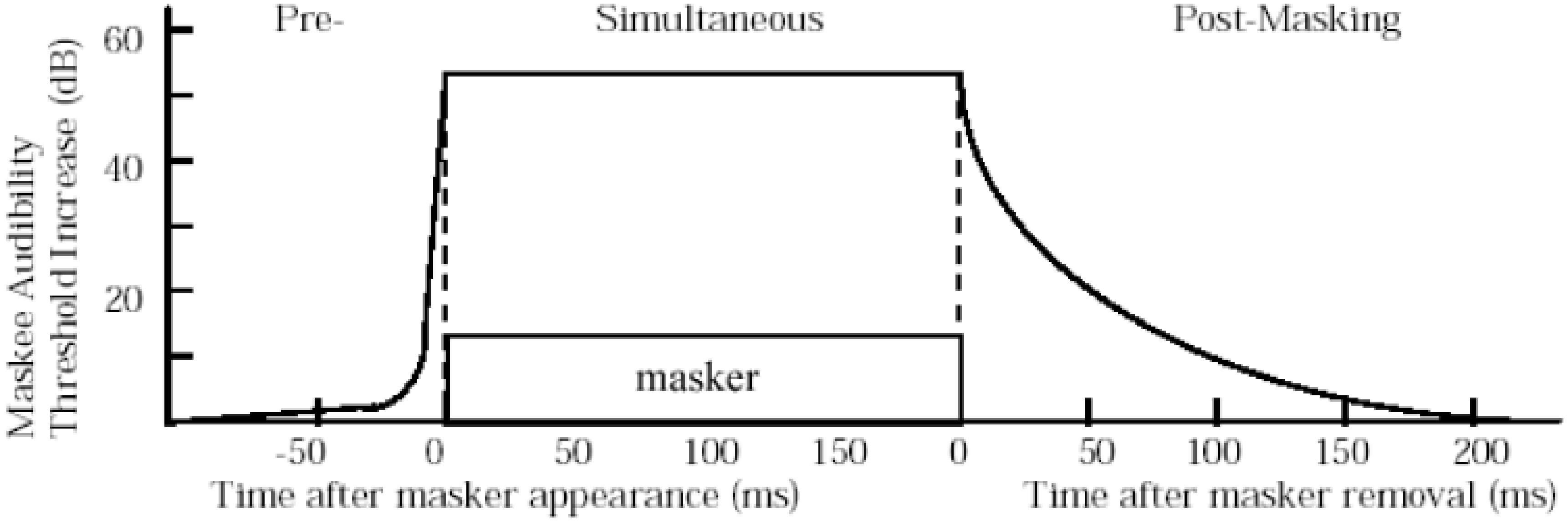
Simultaneous Masking



Temporal Masking

- Temporal masking occurs in the time-domain.
- A stronger tonal component (masker) will mask a weaker one (maskee) if they appear within a small interval of time.
- The masking threshold will mask weaker signals pre and post to the masker.
- Pre-masking usually lasts about 50 ms while post-masking will last from 50 to 300 ms

Temporal Masking



The MPEG-1 Standard

- The development began in 1988 and was finalized in 1992 given the name MPEG-1. The standard consisted of three different parts:
  - An audio part
  - A video part
  - A System part

MPEG-1 Audio

- For the audio part there were three levels of compression and complexity defined; Layer I, Layer II and Layer III. Increased complexity requires less transmission bandwidth since the compression scheme becomes more effective.

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

Audio Data Compression

- MEPG standard does not include the sounds that are not heard. Therefore, masking concepts are used for compressing data
- MPEG standard considers other types of redundancies. These redundancies will be discussed later