Introduction to Multimedia Computing

Audio Media Type (II)

Topics

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

Digital Audio (recap)

- What is Sound?

  Sound is the brain's interpretation of electrical impulses being sent by the inner ear through the nervous system.

- Audio is sound within the acoustic range available to humans.

- An audio is an alternating signal with frequency (AF) within the 20 to 20,000 hertz (cycles per second) range.

Sound Categories (recap)

- ## Content Sound

  Content sound provides information to audiences, for example, dialogs in movies or

  theater.

- ## Ambient Sound

  Ambient sound consists of an array of background and sound effects.

Sample Audio Data (recap)

- 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

Sound Formats and Settings (recap)

- 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

Audio Representation Standards (recap)

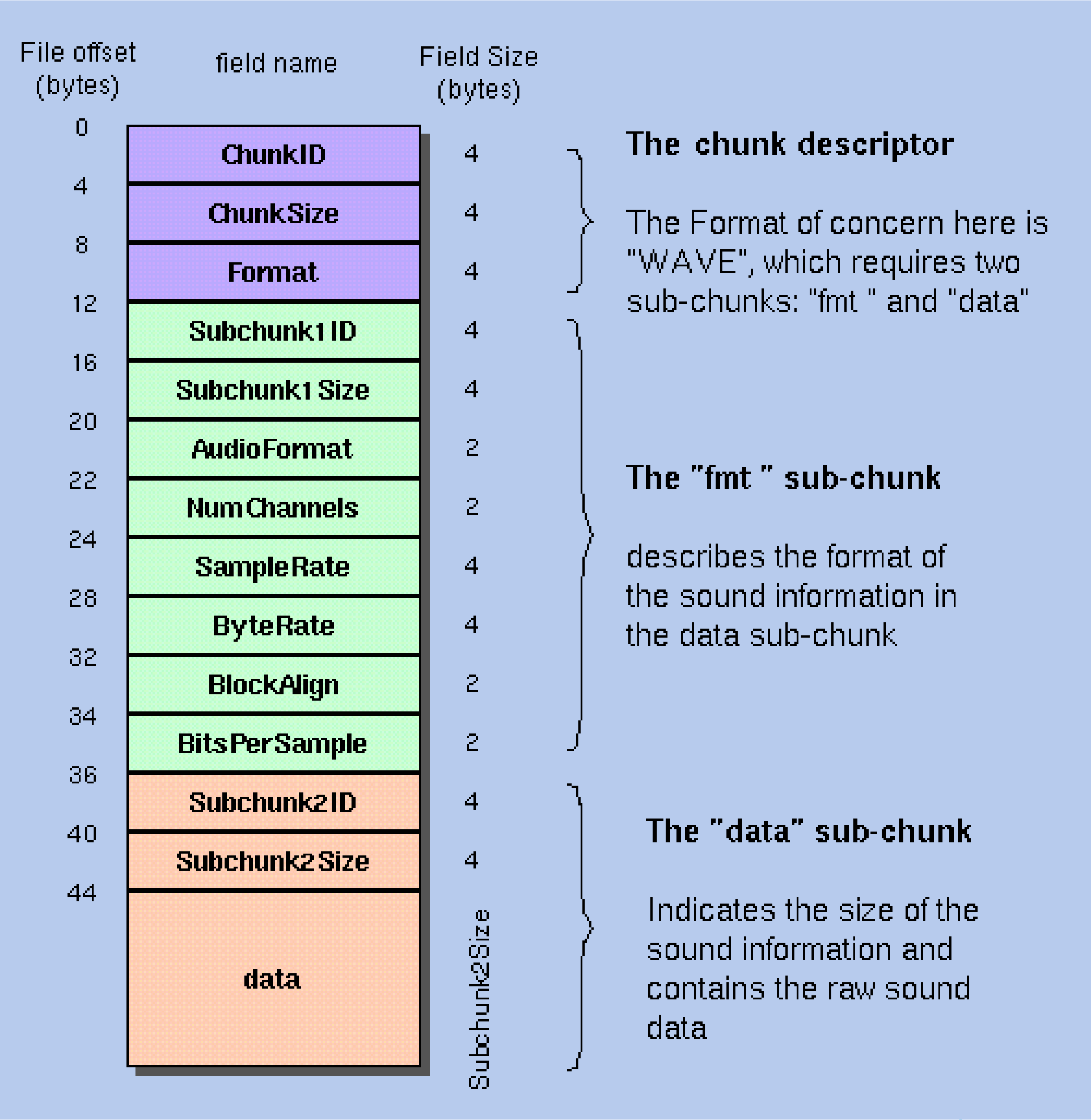
- *wav* - standard audio file format used mainly in Windows PCs. Commonly used for storing uncompressed, CD-quality sound files.
- *mp3* - the MPEG Layer-3 format is the most popular format for downloading and storing music.
- *au* - the standard audio file format used by Sun, Unix and Java. The audio 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 sub-chunks: "fmt " and "data":
- The "fmt " sub-chunk describes the sound data's format:
- The "data" sub-chunk 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 80s 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

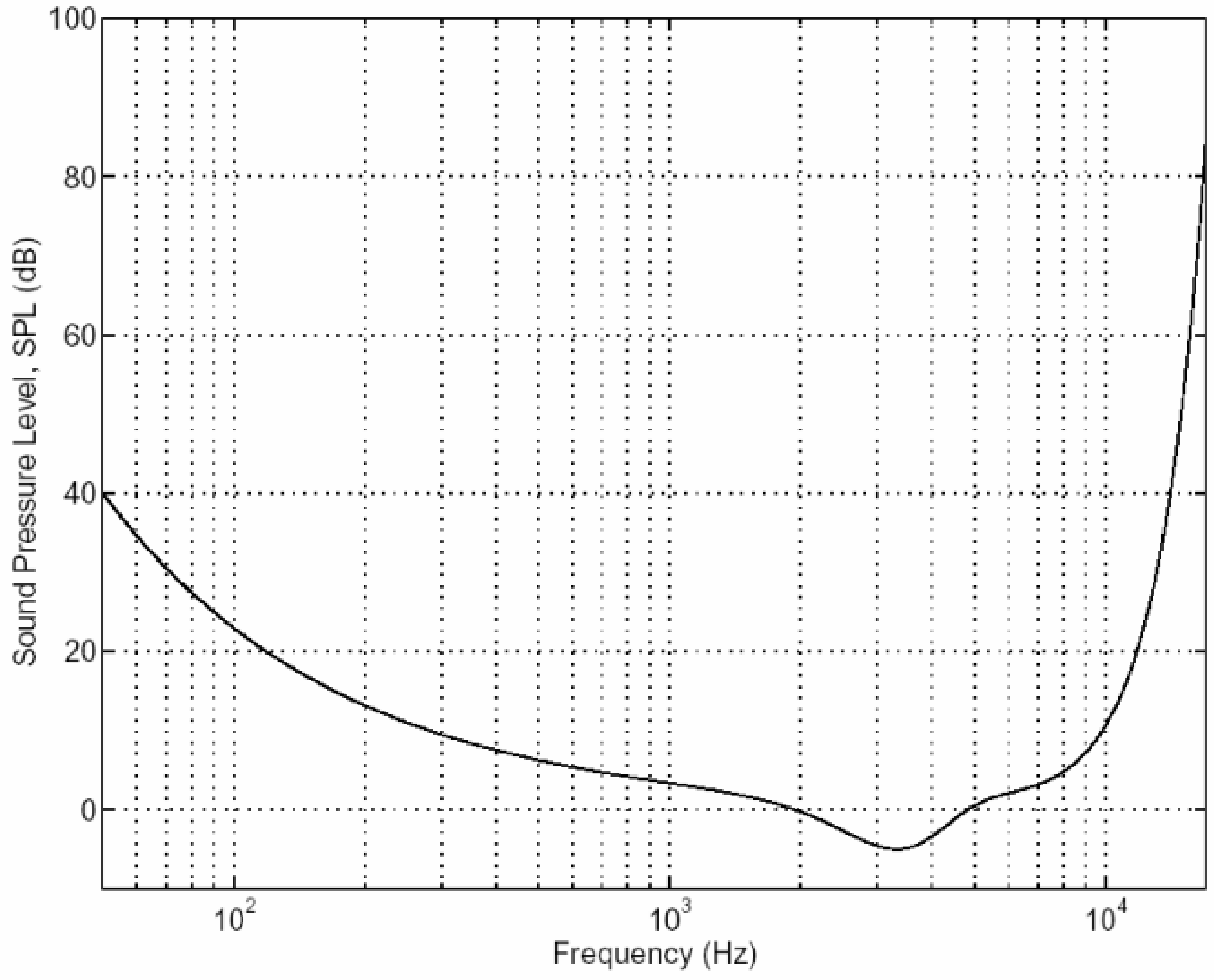
- Data compression or source coding is the process of encoding information using fewer bits than an un-encoded representation.

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



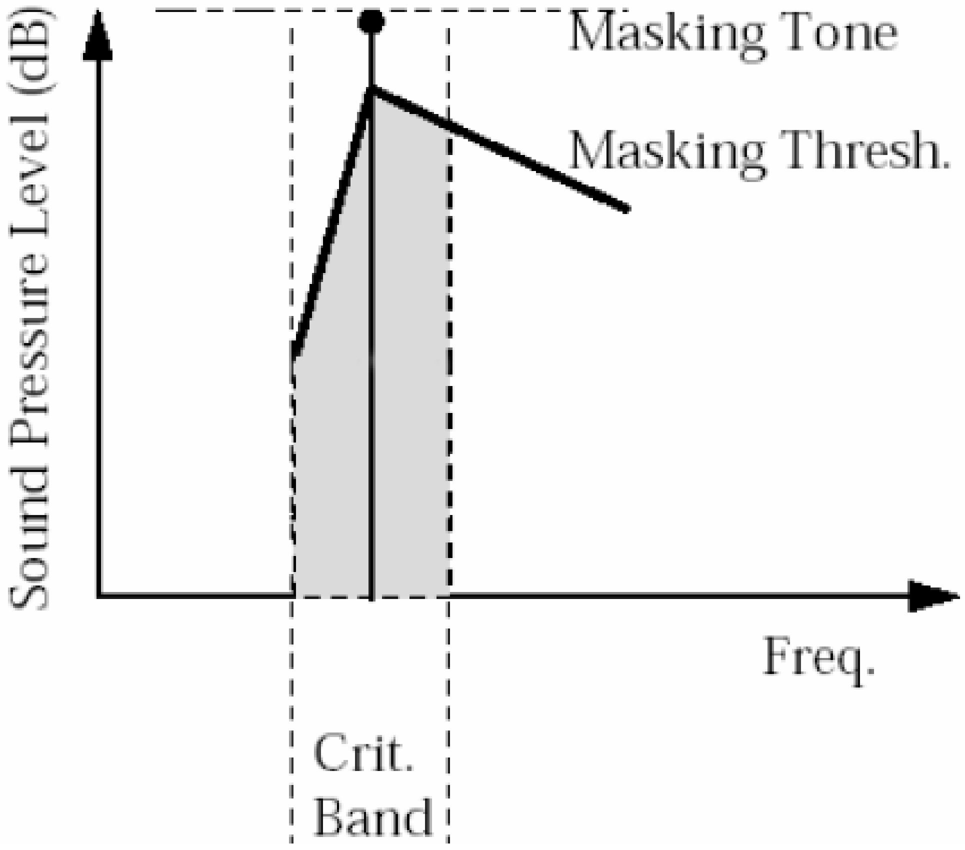
Masking

- 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

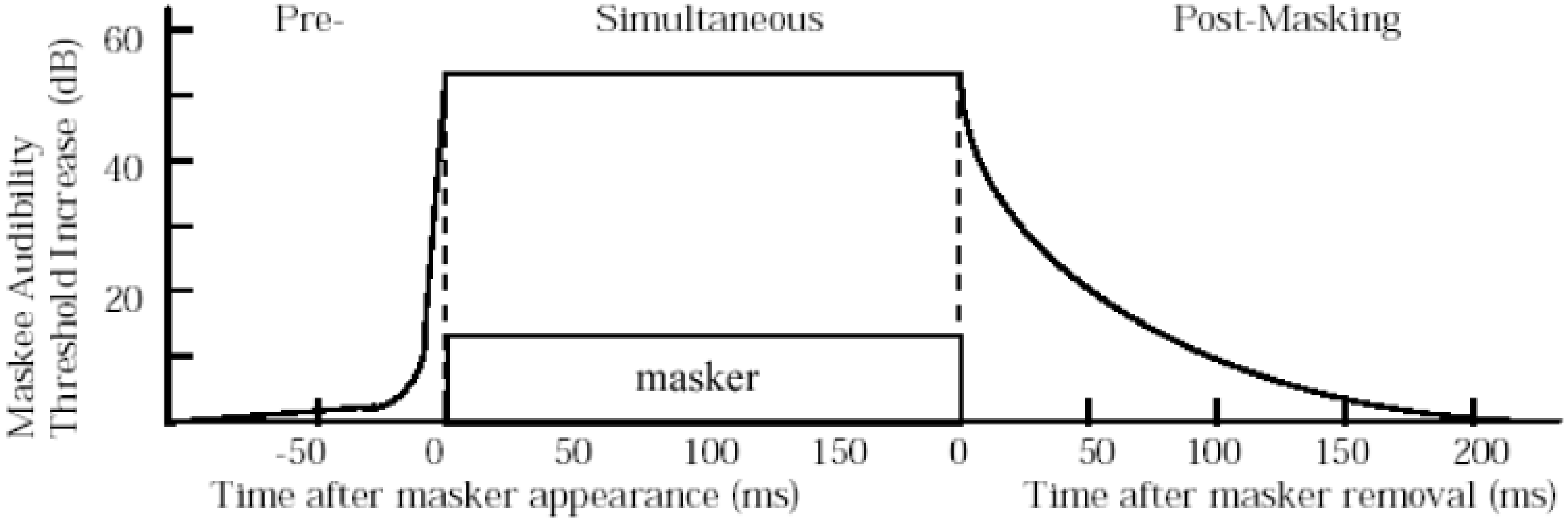
Simultaneous Masking



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

Summary

- Audio is an important media type used for transferring information, creating sound effects, music, and so on.
- A large number of real-time applications transmit audio data over the Internet. Therefore, compressing audio is necessary.
- Human ear cannot every frequency. Some frequencies mask others. These properties are used for removing data redundancy