

Course Code 005636 (Fall 2017)

Multimedia

Audio Compression Principles

Prof. S. M. Riazul Islam, Dept. of Computer Engineering, Sejong University, Korea

E-mail: riaz@sejong.ac.kr

Contents

Overview of A/V compression

- Principles of Audio Compression
 - Predictive Encoding (DPCM)
 - Perceptual Encoding
 - Hearing threshold, Frequency and Temporal Masking
- MP3 (MPEG-1 Layer 3)

Why we need A/V Compression

- Same as of Other data (i.e., Image) compression
- Reduces storage space
- Reduces Bandwidth
- Lower Communication Cost
- New Applications

Audio Compression

Audio Compression

Two Broad Categories

- **Predictive Encoding**
 - Encode the difference between the samples instead of the sample data (For example, DPCM).
- **Perceptual Encoding**
 - Use of Flaws in our auditory system based on the study of how people perceive sound (For example, MP3)

Differential Pulse Code Modulation (DPCM)

- How DPCM reduces the bit rate?

Send the difference between the samples

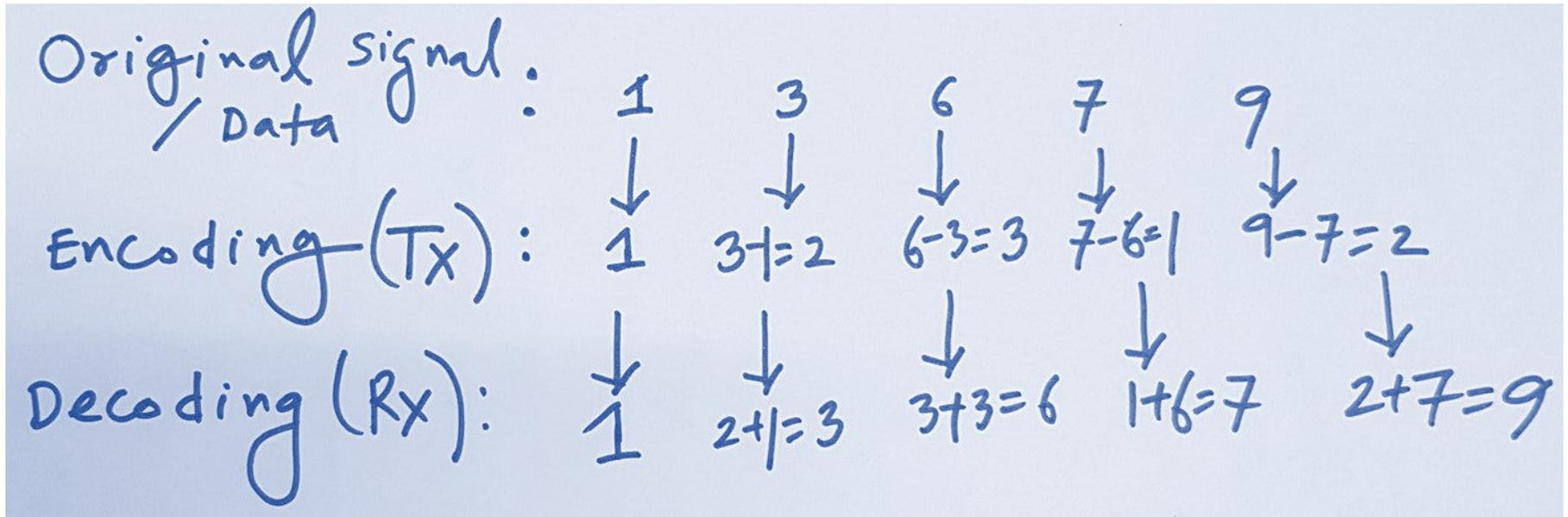
$$\begin{aligned} \text{Max Quantization Error} &= \frac{\Delta}{2} \\ \Delta &= \frac{\text{Range of I/p voltage}}{L} \\ &= \frac{V_{\max} - V_{\min}}{2^n} \end{aligned}$$

L ↑
Number of Q. level

2^n ← no of bits/Sample.

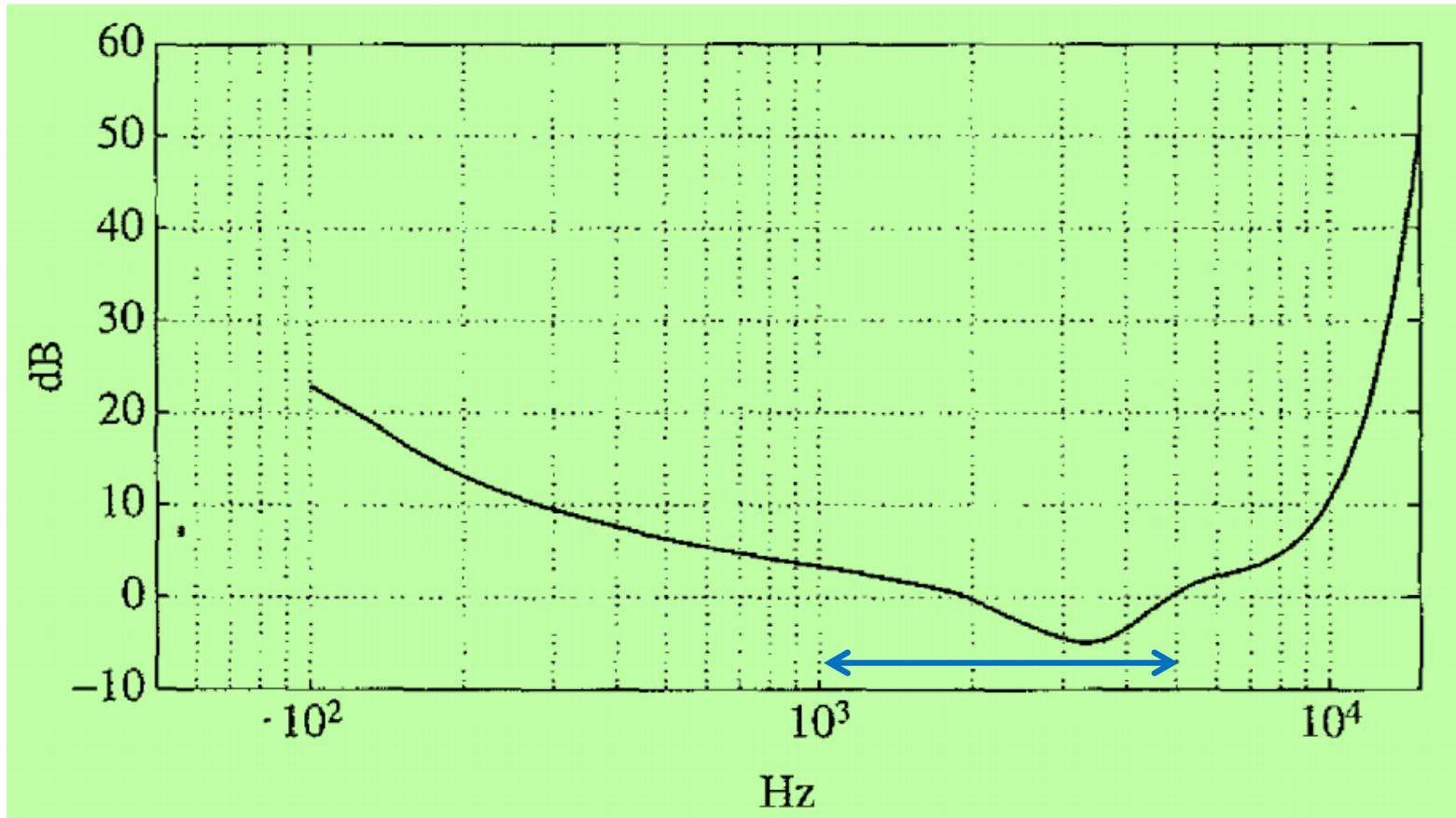
$Q_E \downarrow \rightarrow L \uparrow \rightarrow n \uparrow$
BW ↑ storage ↑

DPCM: Example



- 1) The range of input voltage is reduced
- 2) Step size is reduced without increasing the number of bits per sample

Perceptual Encoding: Hearing Threshold

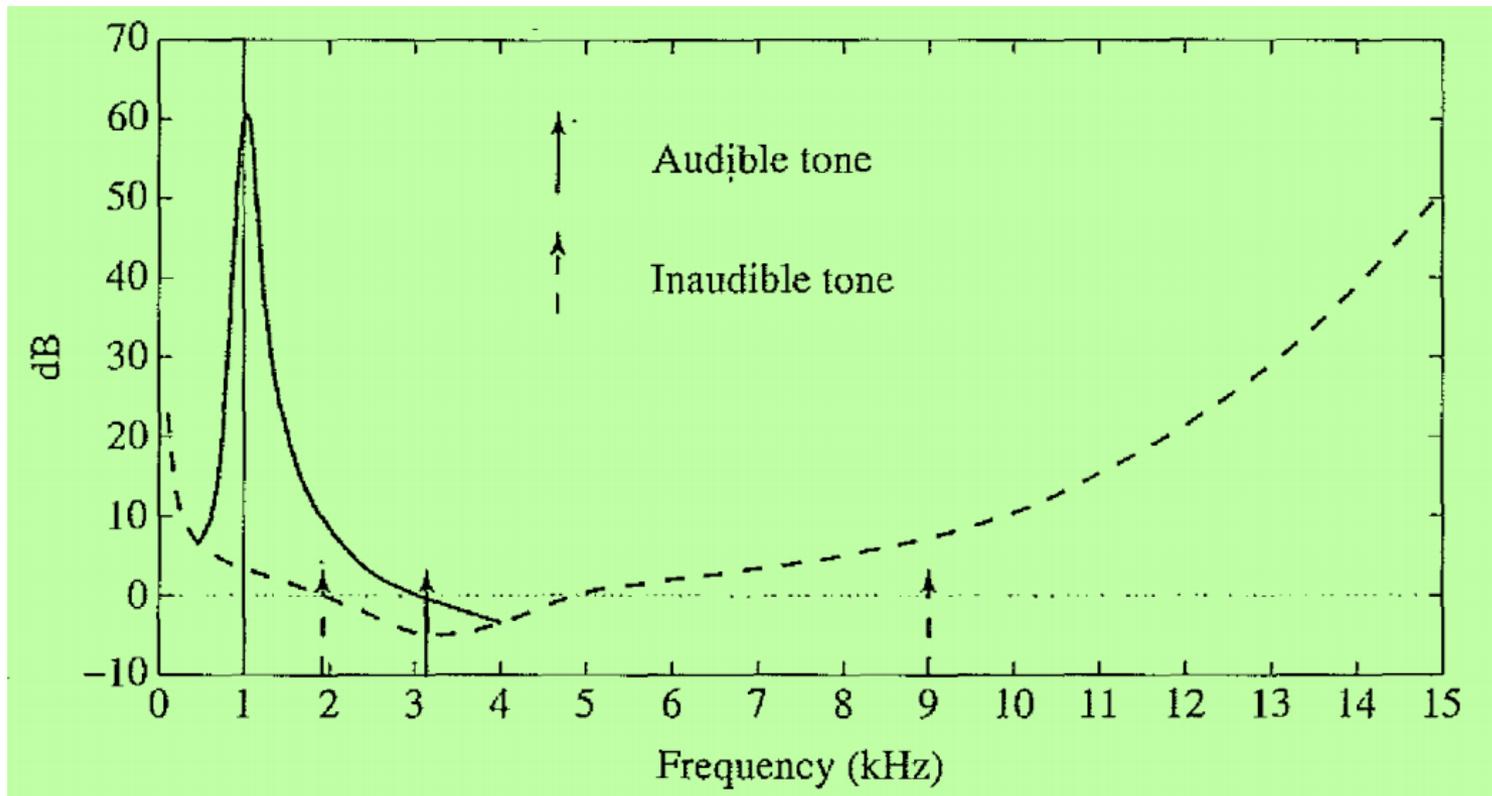


- In particular, at normal sound volume levels, the ear is most sensitive to frequencies between 1 kHz and 5 kHz.
- Audio samples that are below the threshold can be deleted.

Perceptual Encoding: Masking

- Frequency Masking

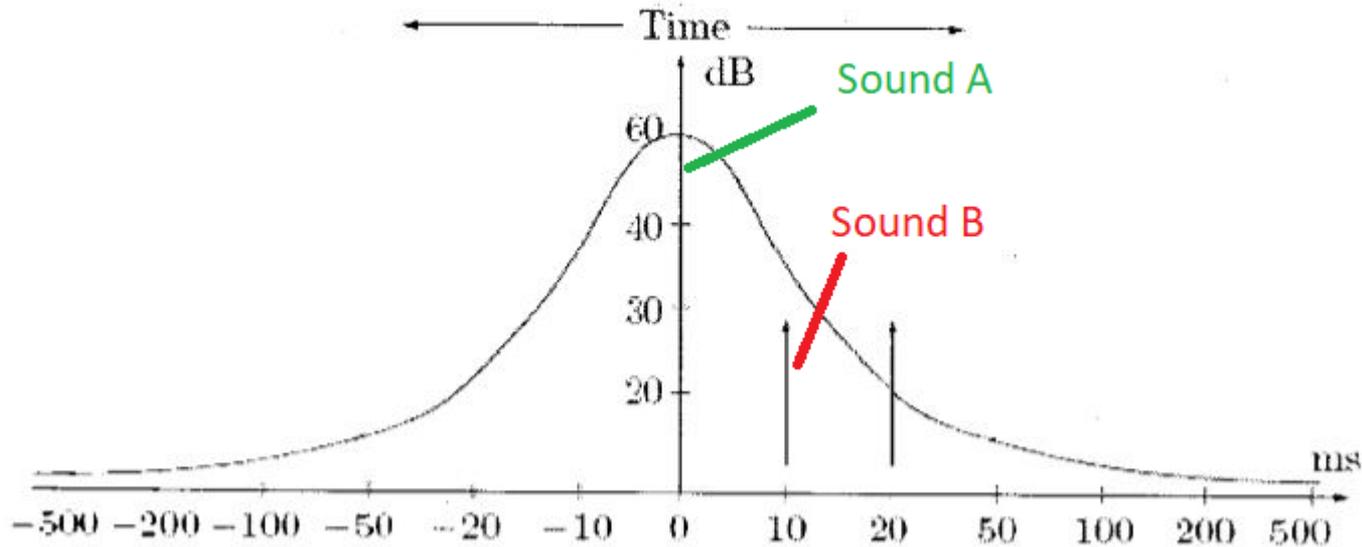
A loud sound in a frequency range can partially or fully masks another sound in the nearby frequency range.



Perceptual Encoding: Masking

- Temporal Masking

A loud sound can numb our ears for a short duration even after the sound has stopped.



** Sound B may not be audible at the same or nearby frequency as of A

Moving Picture Experts Group Phase 1 (MPEG-1)

MPEG-1 Audio Compression

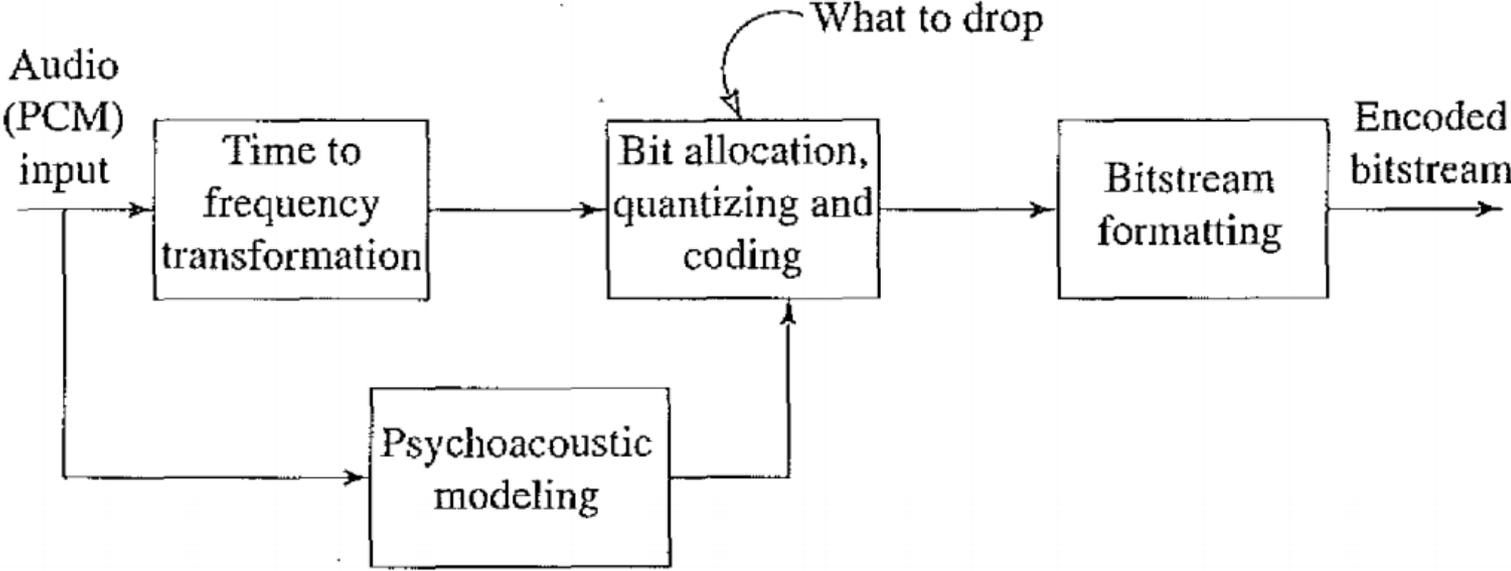
- Sampling: 32 KHz, 44.1 KHz, and 48 KHz
- Time-domain to Frequency domain conversion is performed by FFT. The resulting spectrum is divided in at most 32 frequency bands each of which is processed separately.
- Frequency bands that are to be **completely masked** are allocated zero bits.
- Frequency bands that are to be **partially masked** are allocated small number of bits.
- Frequency bands that are **not to be masked** are allocated larger number of bits.
- The audio stream is adjustable from 32 Kbps to 448 Kbps.

Moving Picture Experts Group Phase 1 (MPEG-1)

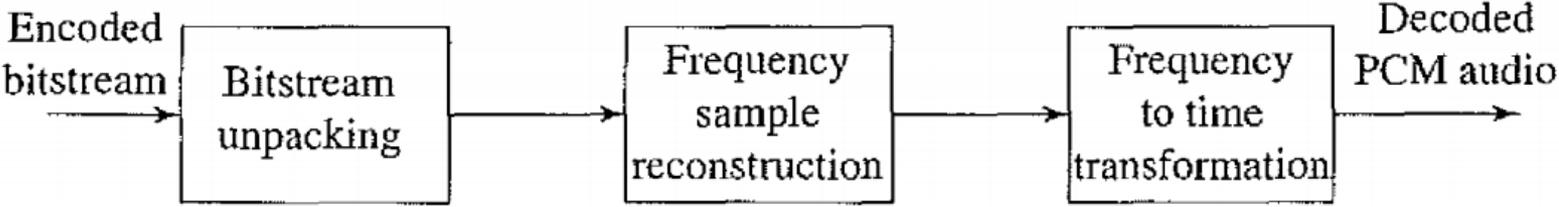
MP3

- MP3 is a popular audio compression standard. The "3" stands for Layer 3, and "MP" stands for the MPEG-I standard. .
- The MPEG standard actually delineates three different aspects of multimedia: audio, video, and systems.
- MP3 forms part of the audio component of this first phase of MPEG.
- It was released in 1992 and resulted in the international standard ISO/IEC 11172-3, published in 1993.

Basic MPEG Audio Encoder and Decoder



(a)



(b)

Bandwidth Reduction

Audio Quality	NC	Sampling Frequency (KHz)	Bits per sample	Data rate (Bandwidth)
Voice (UC)	1	8	8	64 Kbps
Voice (C)	1	8	8	4 to 32 Kbps
CD (UC)	2	44.1	16	1.411 Mbps
CD (C)	2	44.1	16	64 to 192 Kbps

UC → Uncompressed
C → Compressed
NC → Number of Channels

Q&A

