Special Module on Media Processing and Communication

Dayalbagh Educational Institute
(DEI)
Dayalbagh Agra

Indian Institute of Technology Delhi
(IITD)
New Delhi

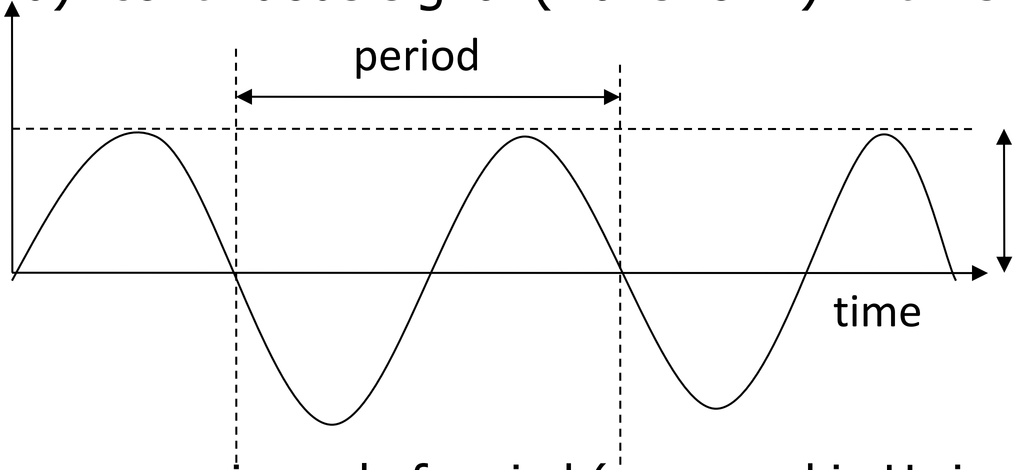
SIV864

Audio Compression

Audio (Revisit)

Digital Representation

Audio (Sound): continuous signal (wave form) in time 1D function f(x)



Frequency: reciprocal of period (measured in Hz i.e., cycles/sec)
relates to the pitch of sound

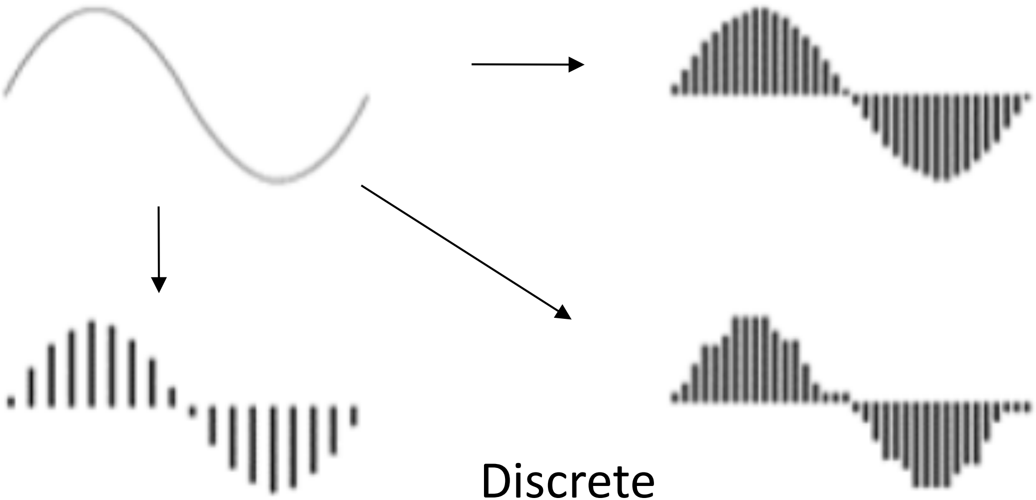
Amplitude: relates to the loudness of sound (measured in decibels –db)

Audio (Revisit)

Digital Representation

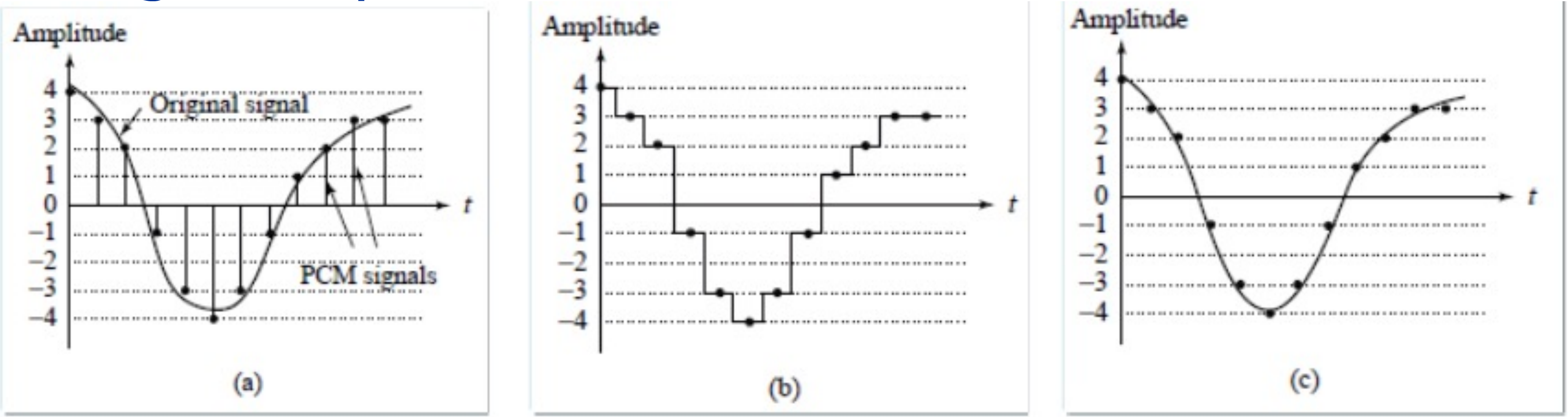
Audio (Sound): continuous signal (wave form) in time 1D function f(x)

Continuous



Discrete

Digital Representation: Pulse Code Modulation

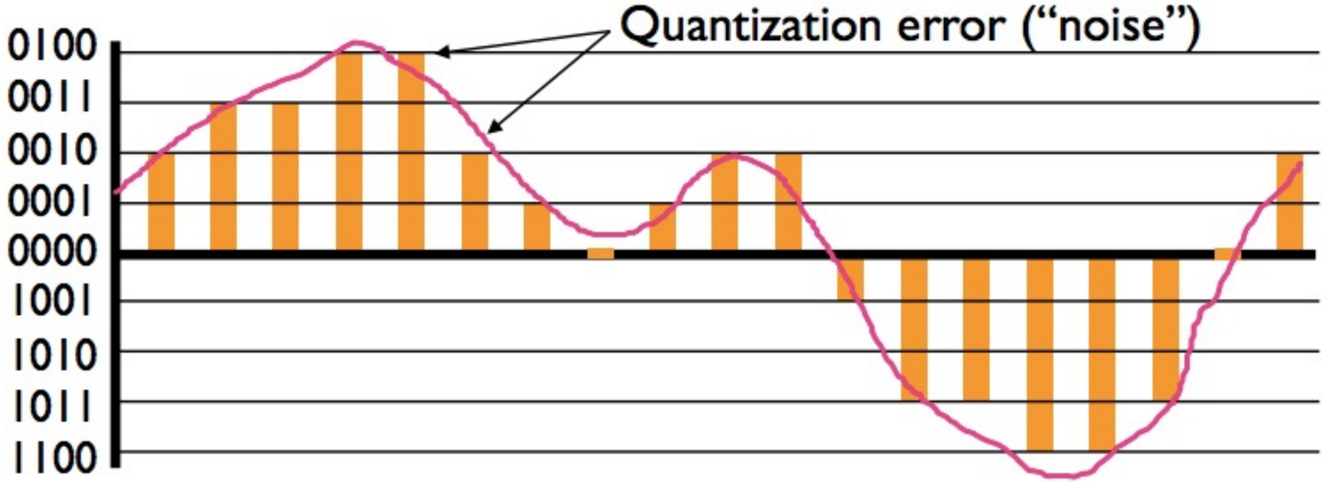


Original analog signal and its corresponding PCM signals.	Decoded staircase signal.	Reconstructed signal after low-pass filtering.

Source [3]

Pulse Code Modulation

Linear Quantization

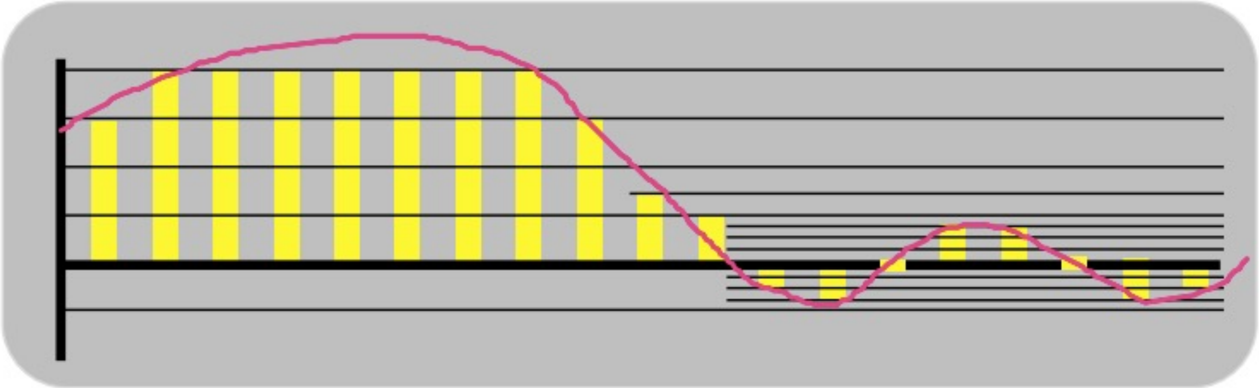


Quantization levels are equally spaced.

More number of bits (greater quantization levels), lower the quantization noise.

Source [2]

Non-Linear Quantization



Non-linear quantization of the signal's amplitude:

- Quantization step-size decreases logarithmically with signal level.
- Low-amplitude samples represented with greater accuracy than high amplitude samples.
- Logarithmic-compressed quantizer.

Source [2]

Pulse Code Modulation

Companding

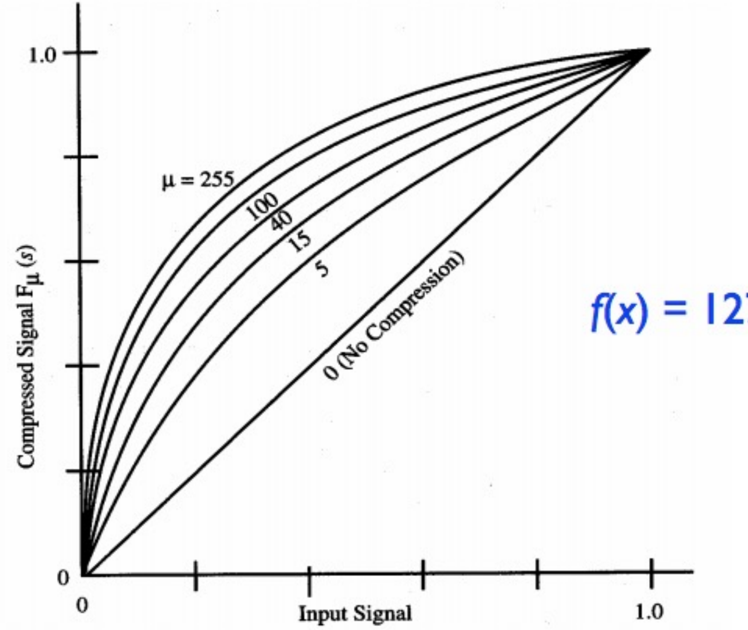
Compressing – Expanding

Uses the fact that the ear requires more precise samples at low amplitudes and more forgiving at higher amplitudes – logarithmic law

μ-Law and A-Law companding

Pulse Code Modulation

μ-Law Companding



$$f(x) = 127 \times \text{sign}(x) \times \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)}$$

(x normalized to [-1, 1])

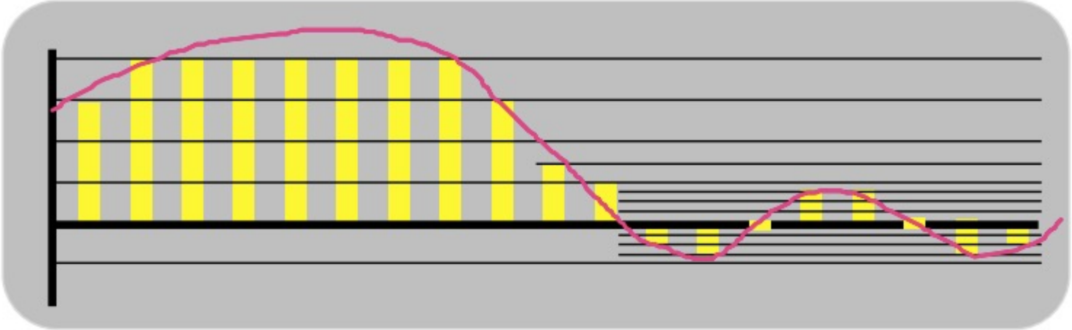
Typically $\mu = 255$

Adopted in North America and Japan

Source [2]

Pulse Code Modulation

μ-Law Companding



- Provides 14-bit quality (dynamic range) with an 8-bit encoding
- Used in North American & Japanese ISDN voice service
- Simple to compute encoding.
- Compression factor 2:1.
- Example file formats:
  - .wav (Microsoft), .au (Sun).

Source [2]

Pulse Code Modulation

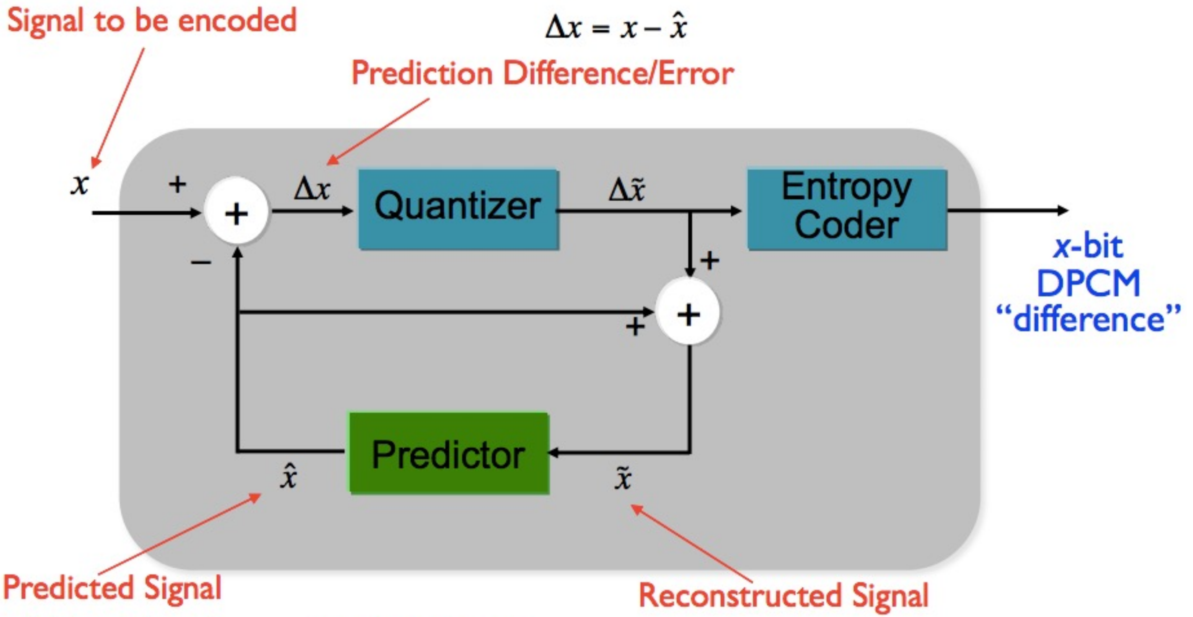
A-Law Companding

$$F(x) = \operatorname{sgn}(x) \begin{cases} \dfrac{A|x|}{1+\ln(A)}, & |x| < \dfrac{1}{A} \\ \dfrac{1+\ln(A|x|)}{1+\ln(A)}, & \dfrac{1}{A} \leq |x| \leq 1, \end{cases}$$

Typically, A=87.6

Adopted in Europe

Differential Pulse Code Modulation(DPCM)



$\tilde{x} = \hat{x} + \Delta\tilde{x}$

$\hat{x}(n) = \tilde{x}(n-1)$ Simple difference

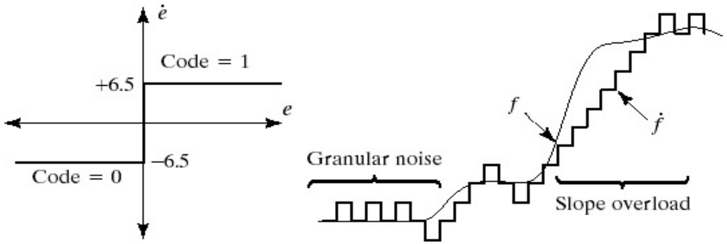
$\hat{x}(n) = \sum_{k=1}^{N} h_k \tilde{x}(n-k)$ Linear predictor => minimizes quantization error

$\hat{x}(n) = 0$ PCM

Source [2]

Image Compression (Revisit)

Predictive Coding: Lossy



a b
c

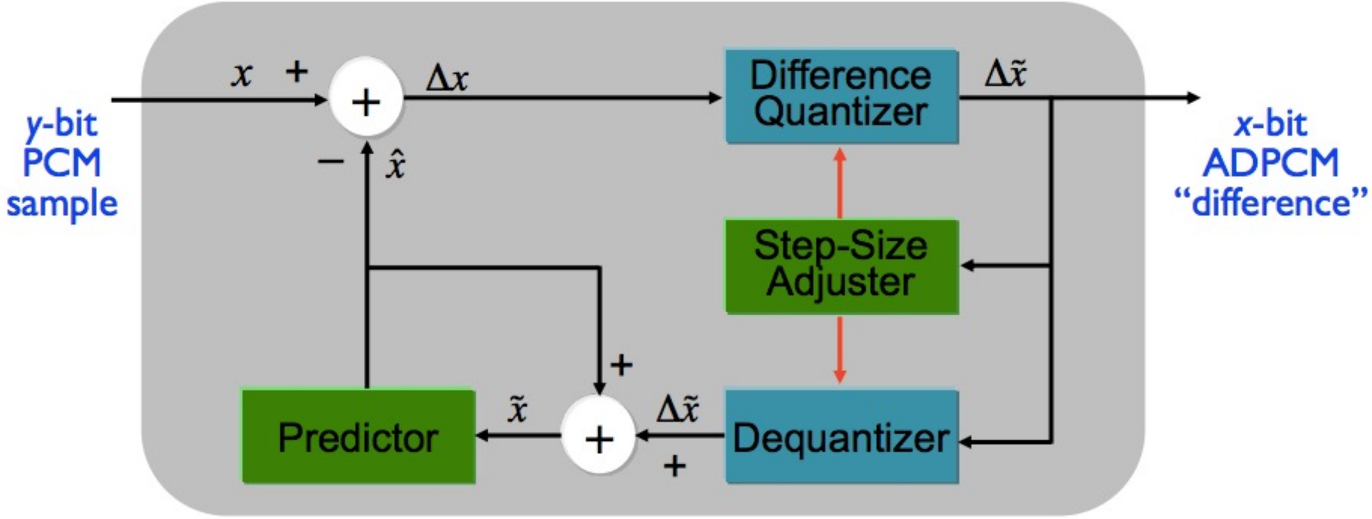
FIGURE 8.22 An example of delta modulation.

Input		Encoder				Decoder		Error
n	f	$\hat{f}$	e	$\dot{e}$	$\dot{f}$	$\hat{f}$	$\dot{f}$	$[f - \dot{f}]$
0	14	—	—	—	14.0	—	14.0	0.0
1	15	14.0	1.0	6.5	20.5	14.0	20.5	–5.5
2	14	20.5	–6.5	–6.5	14.0	20.5	14.0	0.0
3	15	14.0	1.0	6.5	20.5	14.0	20.5	–5.5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
14	29	20.5	8.5	6.5	27.0	20.5	27.0	2.0
15	37	27.0	10.0	6.5	33.5	27.0	33.5	3.5
16	47	33.5	13.5	6.5	40.0	33.5	40.0	7.0
17	62	40.0	22.0	6.5	46.5	40.0	46.5	15.5
18	75	46.5	28.5	6.5	53.0	46.5	53.0	22.0
19	77	53.0	24.0	6.5	59.6	53.0	59.6	17.5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

Adaptive DPCM

To ensure differences are always small…

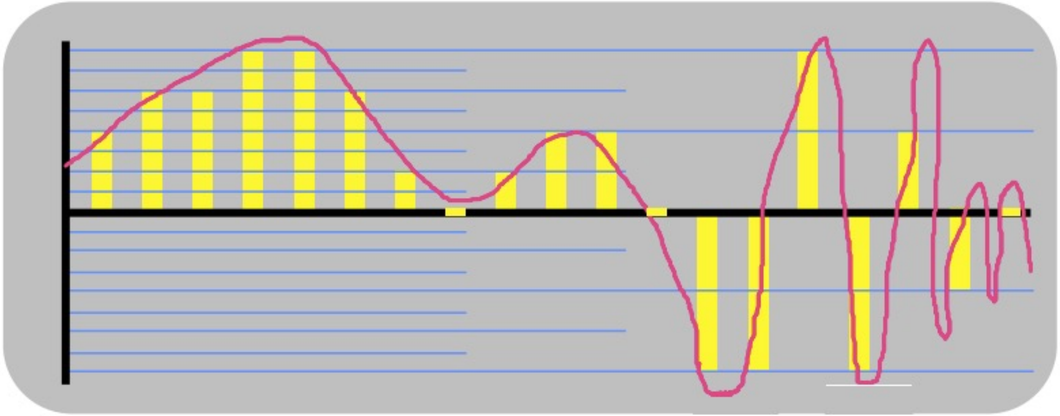
- Adaptively change the step-size (quanta).
- (Adaptively) attempt to predict next sample value.



$$\tilde{x} = \hat{x} + \Delta\tilde{x}$$

Source [2]

Adaptive DPCM



- Use a larger step-size to encode differences between high-frequency samples & a smaller step-size for differences between low-frequency samples.
- Use previous sample values to estimate changes in the signal in the near future.

Source [2]

Psychoacoustic Coding

Based on extensive studies of human perception:

- We do not hear all frequencies the same way.

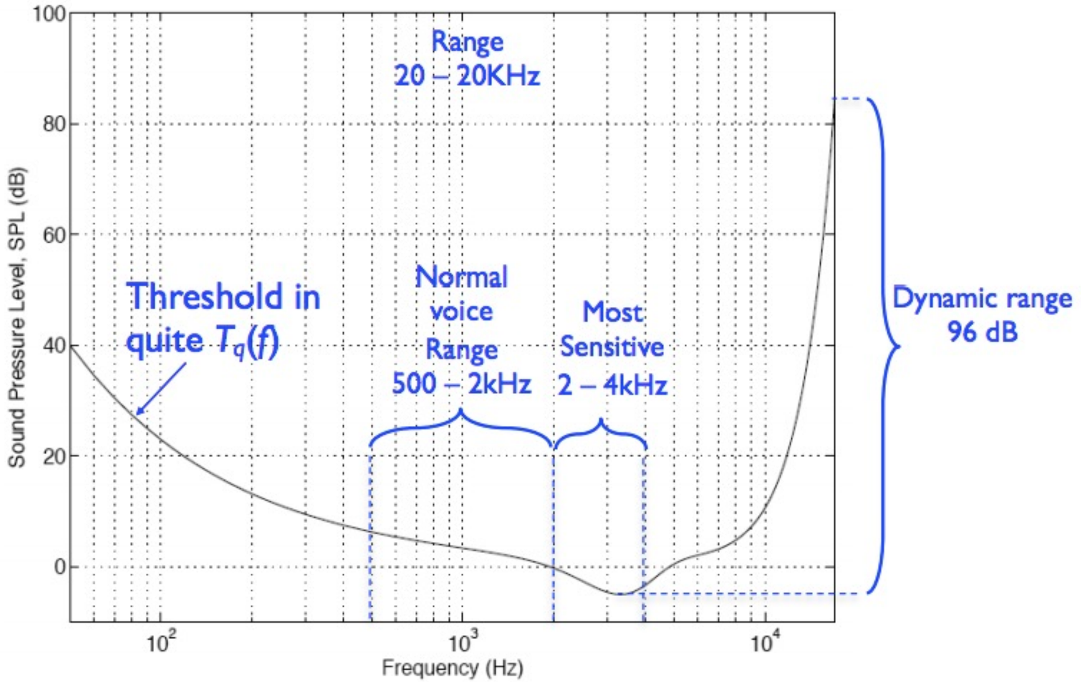
Limitations of the human sensory system => cut out unnecessary data in an audio signal!

Two main properties of the human auditory system:

- Absolute threshold of hearing
- Auditory masking:
  - Simultaneous masking
  - Temporal masking

Psychoacoustic Coding

Absolute threshold of h
in quite environment

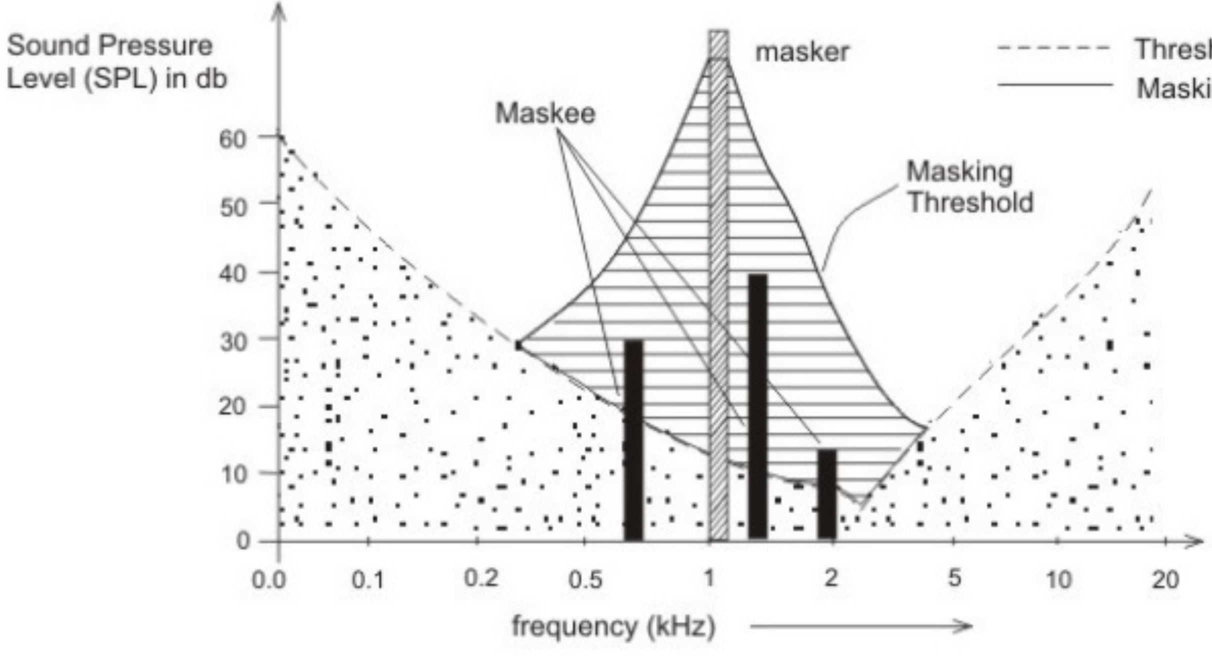


Source [2]

$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4$$

Psychoacoustic Coding

Simultaneous Masking



Source [1]

Psychoacoustic Coding

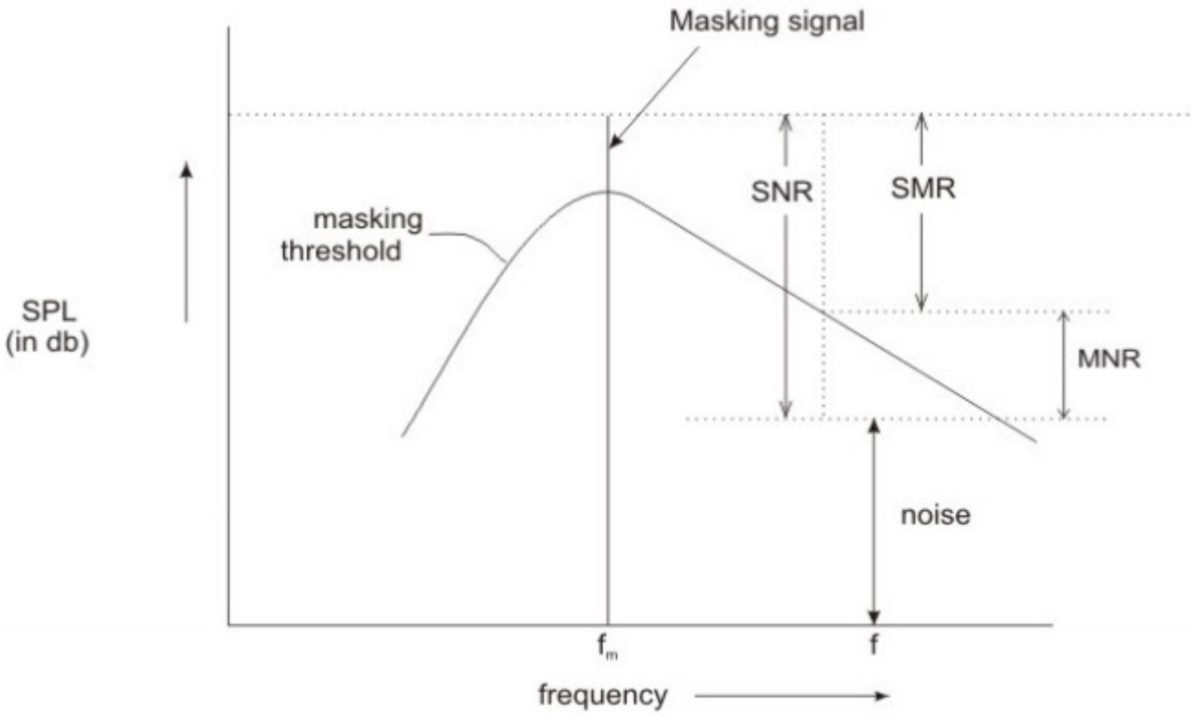
SOME DEFINITIONS:

Signal to noise ratio

Signal to mask ratio (SMR)

Mask to noise ratio (MNR)

SMR (f) = SNR (f) – MNR (f)

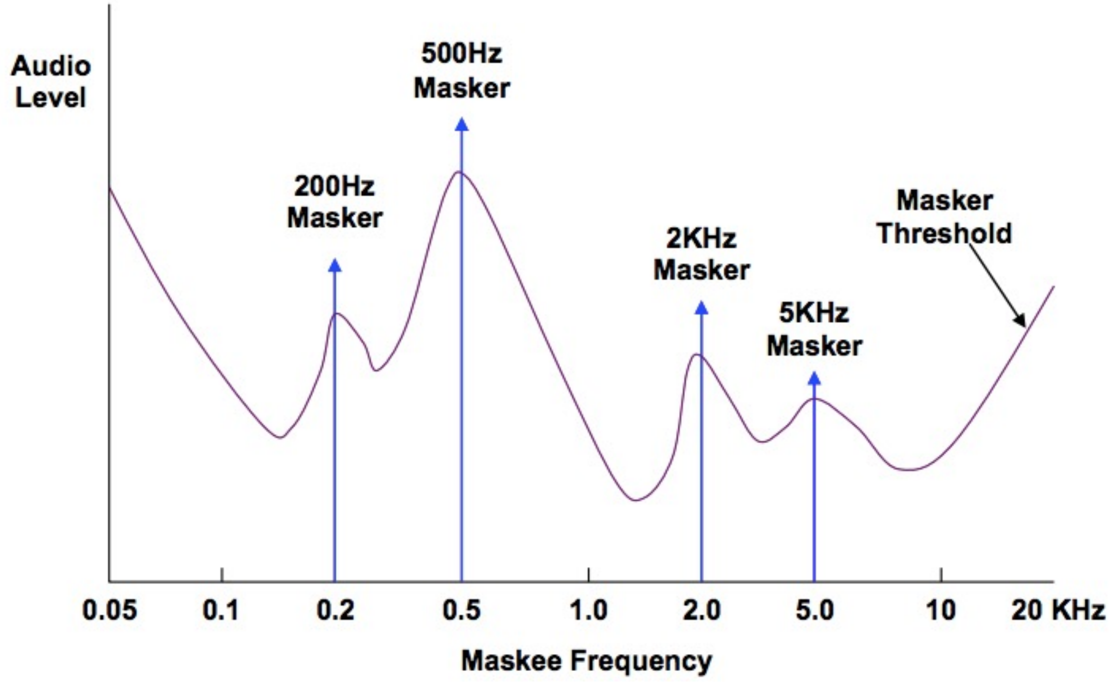


Source [1]

Psychoacoustic Coding

GLOBAL THRESHOLDING

Multiple maskers

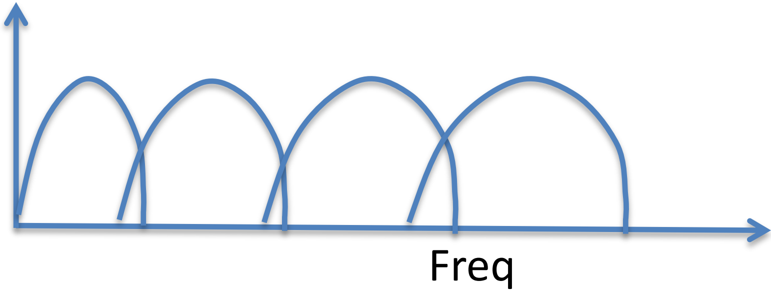


Source [2]

Psychoacoustic Coding

- Human auditory:
  - Limited and frequency dependent resolution.
  - 25 critical bands.
- Bark - Barkhausen
  - 1 Bark = width of one critical band
  - Critical band number (Bark) for a given frequency, $z(f)$:
    - $f < 500\text{Hz} \Rightarrow z(f) \approx f/100$
    - $f > 500\text{Hz} \Rightarrow z(f) \approx 9 + 4\log_2(f/1000)$
  - Also given as

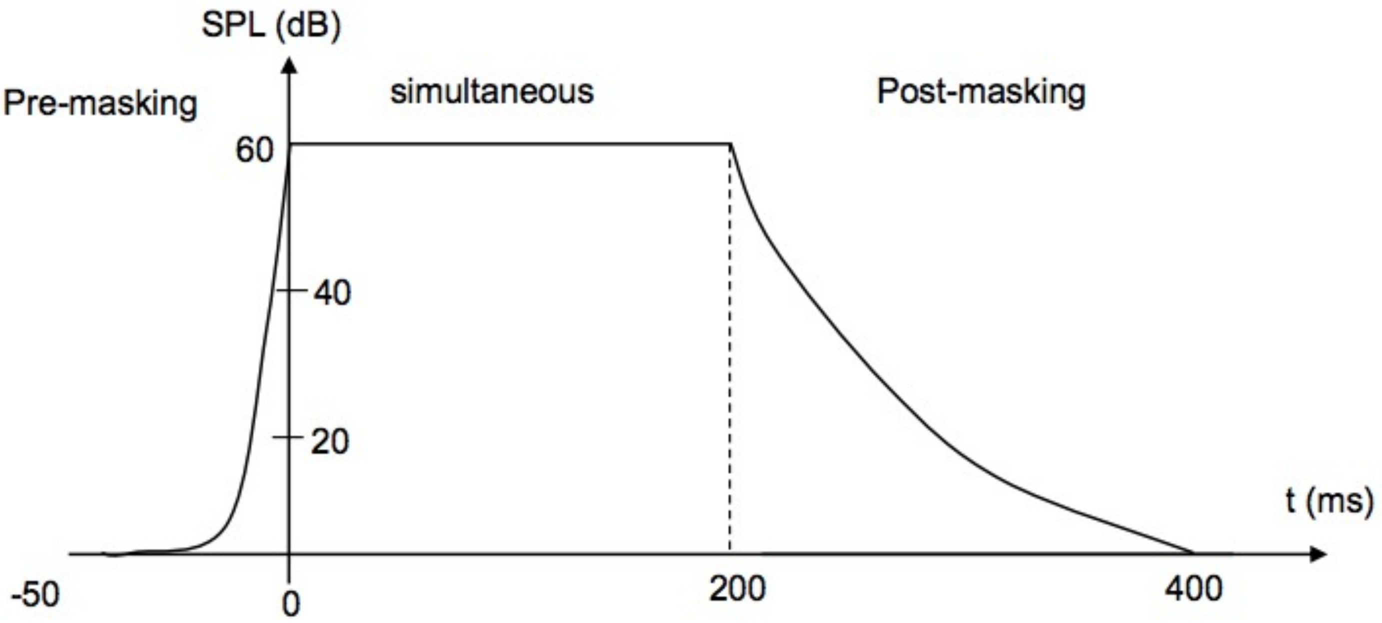
$$z(f) = 13.0\arctan(0.76f) + 3.5\arctan(f^2/56.25)$$



Source [2]

Psychoacoustic Coding

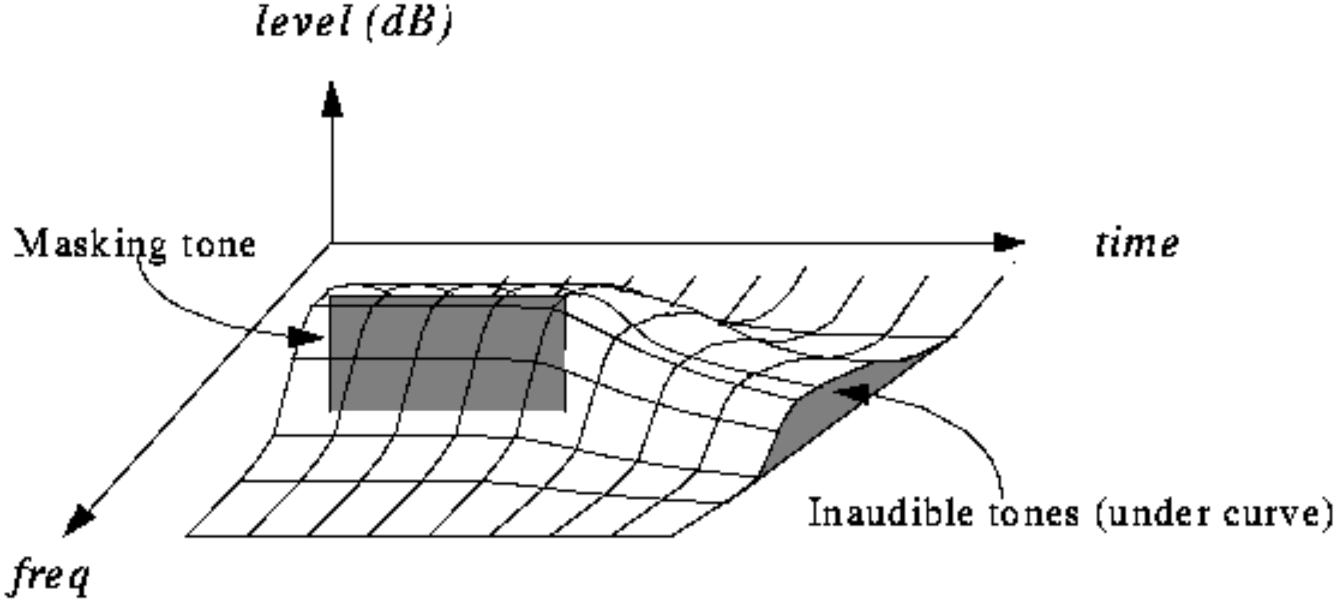
TEMPORAL MASKING



Source [2]

Psychoacoustic Coding

Net Effect



MPEG-1

Layer 1

- Psychoacoustic model only uses simultaneous masking.

Layer 2

- This models a little bit of the temporal masking.

Layer 3

- Psychoacoustic model includes temporal masking effects, and takes into account stereo redundancy.
- Huffman coder.
- Known as MP3

References

1. NPTEL course on Multimedia Processing by Prof S Sengupta.
2. http://web.engr.oregonstate.edu/~benl/Courses/ece477.sp20/Lectures/
3. Sharif University of Technology, Department of Computer Engineering, Multimedia Systems Course