



CS 598KN

Advanced Multimedia Systems Design
Digital Audio + MP3

Klara Nahrstedt
Fall 2018



Overview

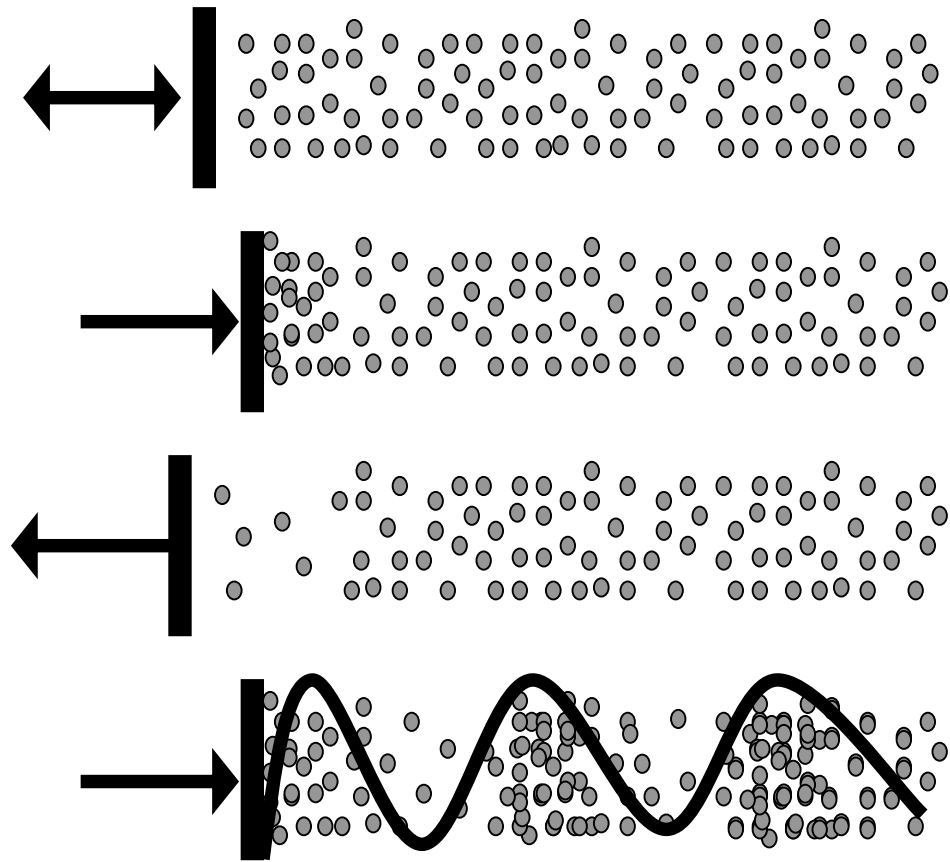
- Human Auditory System
- Digital Audio
- MP3 Encoding



DIGITAL AUDIO CHARACTERISTICS

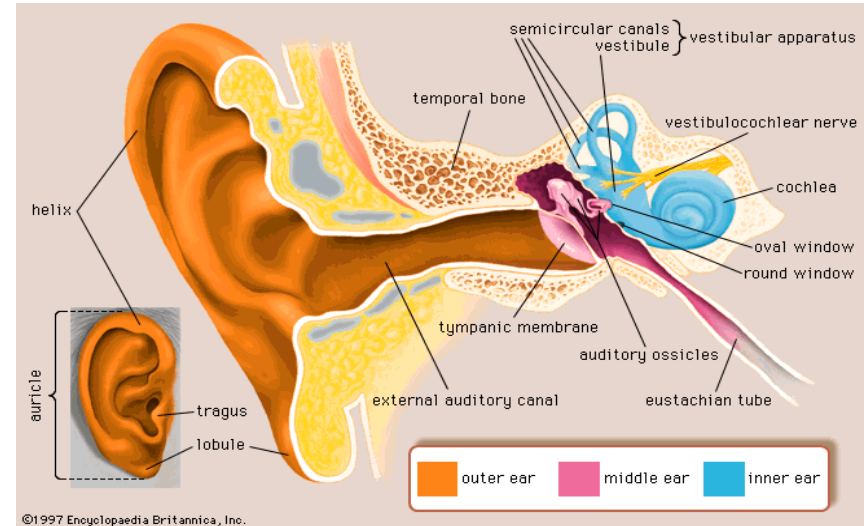
Auditory Perception

- Sound – physical phenomenon caused by vibration of material
- These vibrations trigger pressure wave fluctuations in the air
- Wave forms

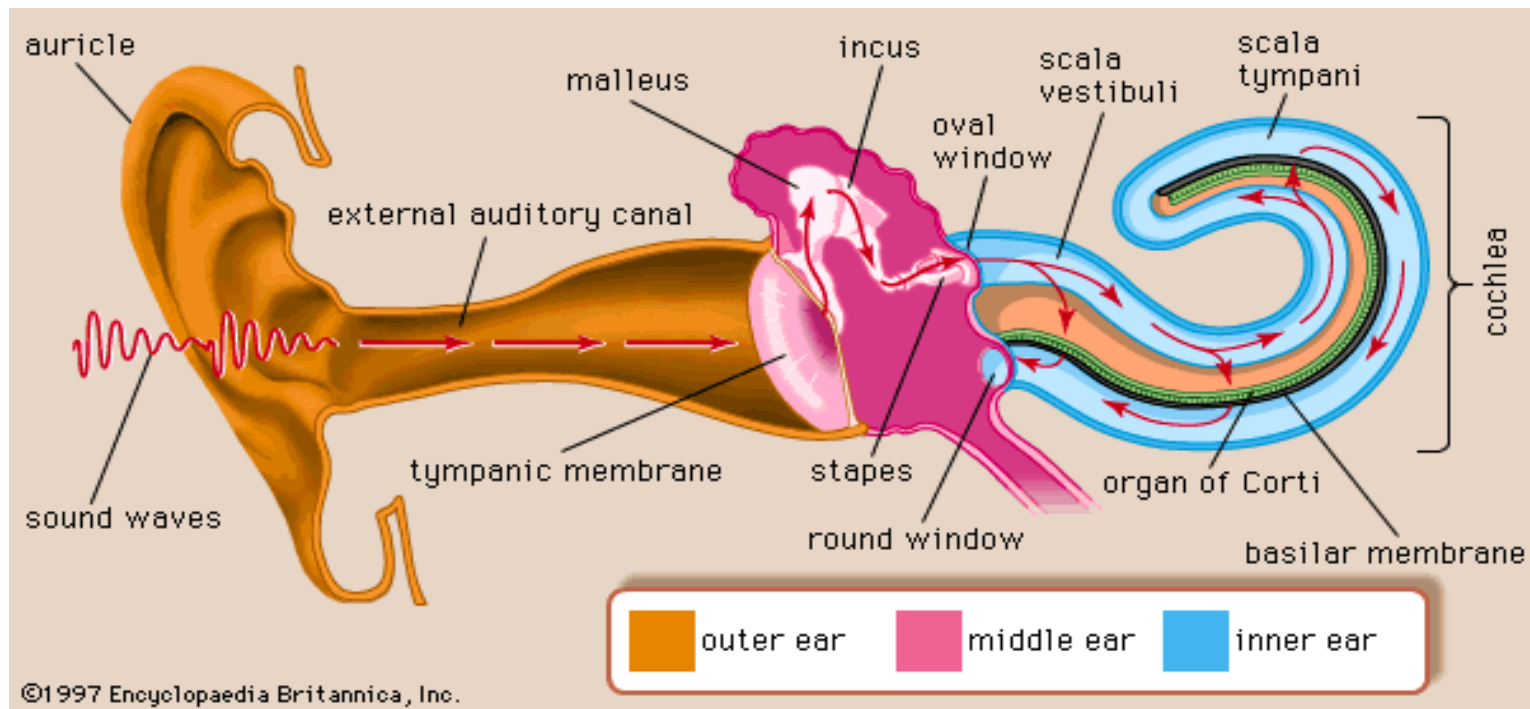


Auditory System

- Ears, parts of brain, and neural pathways
- Changes in pressure move hair-like fibers within the inner ear
- Movements result in electrical impulses sent to the brain

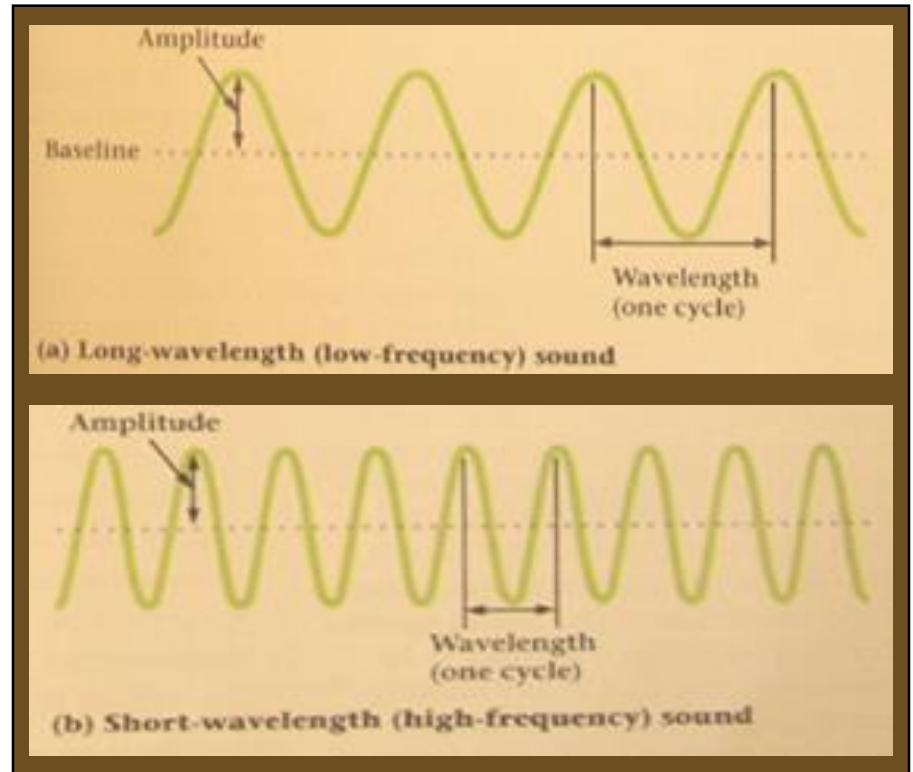


Process of Hearing (Transduction)



Physical Dimensions

- Amplitude
 - height of a cycle
 - relates to loudness
- Wavelength (w)
 - distance between peaks
- Frequency (λ)
 - cycles per second
 - relates to pitch
 - $\lambda w = \text{velocity}$
- Most sounds mix many frequencies & amplitudes



Sound is repetitive changes
in air pressure over time

Sound Perception and Psychoacoustics

■ **Psychoacoustics**

- Study the correlation between the physics of acoustical stimuli and hearing sensations
- Experimental data and models are useful for audio codec

■ **Modeling human hearing mechanisms**

- Allows to reduce the data rate while keeping distortion from being audible

Psychological Dimensions

■ Loudness

- higher amplitude results in louder sounds
- measured in decibels (db), 0 db represents hearing threshold

■ Pitch

- higher frequencies perceived as higher pitch
- hear sounds in 20 Hz to 20,000 Hz range

Decibel Scale

- Describes intensity relative to threshold of hearing based on multiples of 10

$$dB = 10 \log \frac{I}{I_0}$$

I_0 is reference level = 10^{-12} W/m^2

Decibels of Everyday Sounds

Sound	Decibels
Rustling leaves	10
Whisper	30
Ambient office noise	45
Conversation	60
Auto traffic	80
Concert	120
Jet motor	140
Spacecraft launch	180

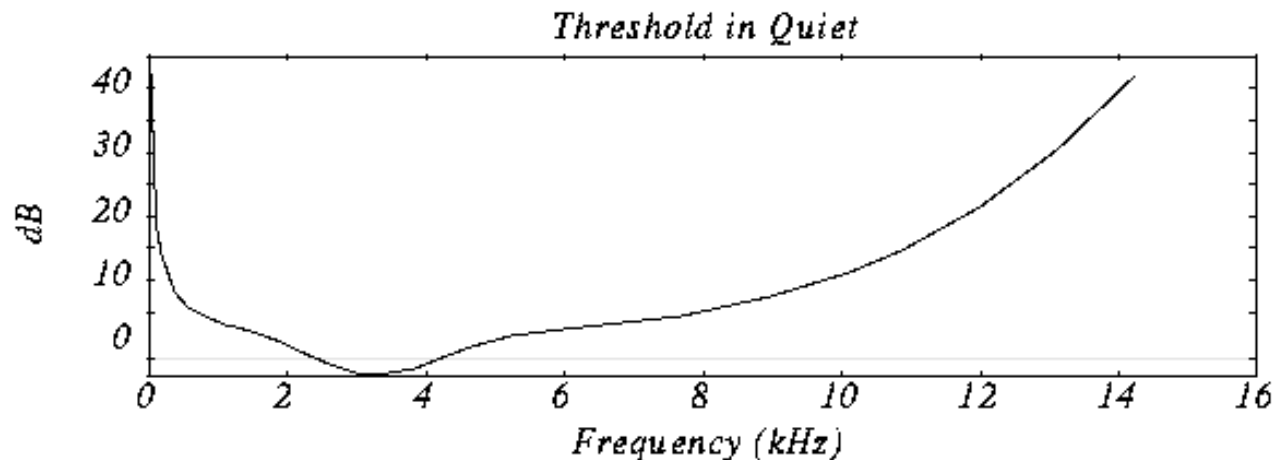


Masking

- Perception of one sound interferes with another
- Frequency masking
- Temporal masking

Frequency Masking

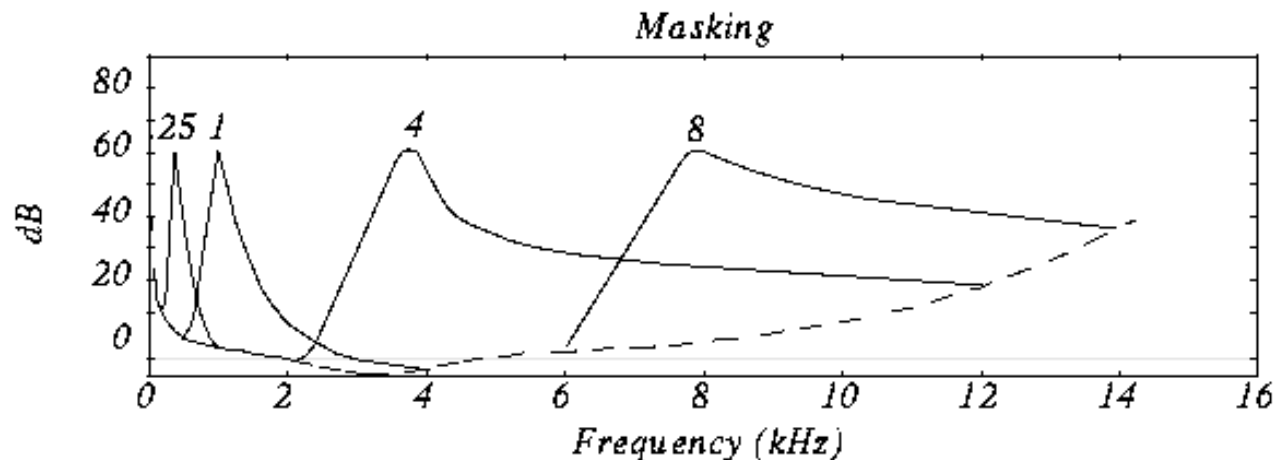
- Louder, lower frequency sounds tend to mask weaker, higher frequency sounds



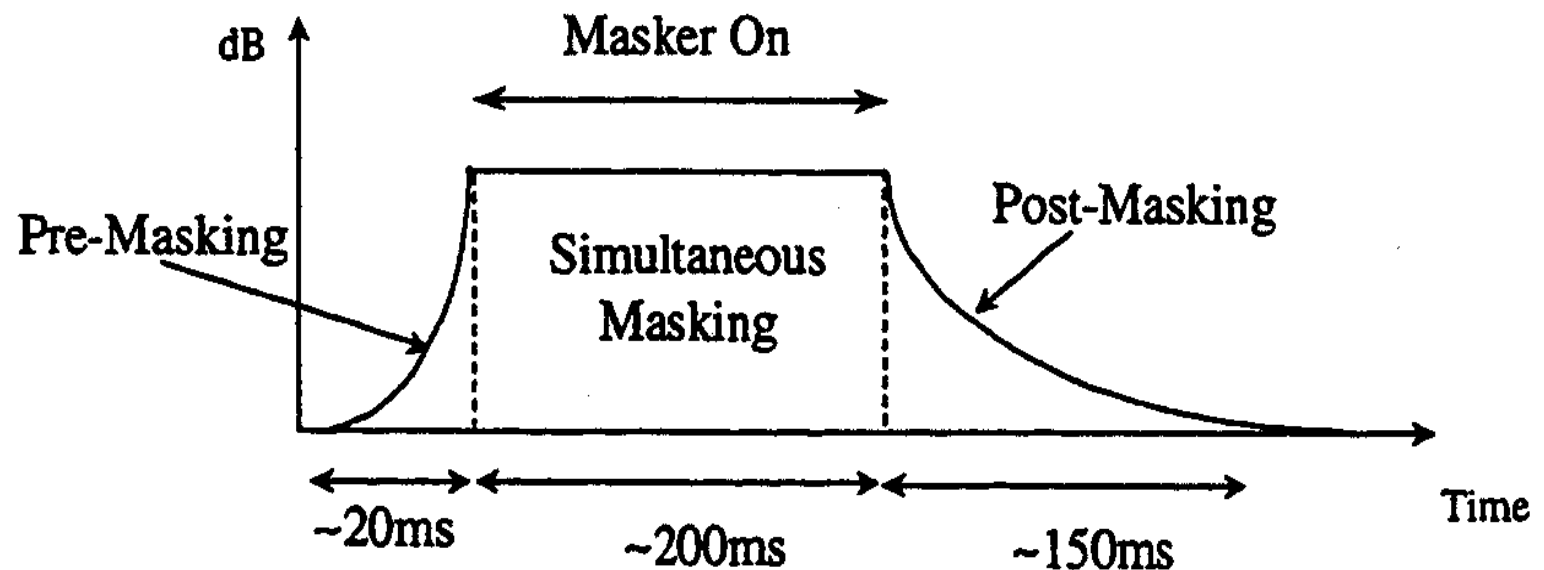
From <http://www.cs.sfu.ca/CourseCentral/365/>

Frequency Masking

- Louder, lower frequency sounds tend to mask weaker, higher frequency sounds



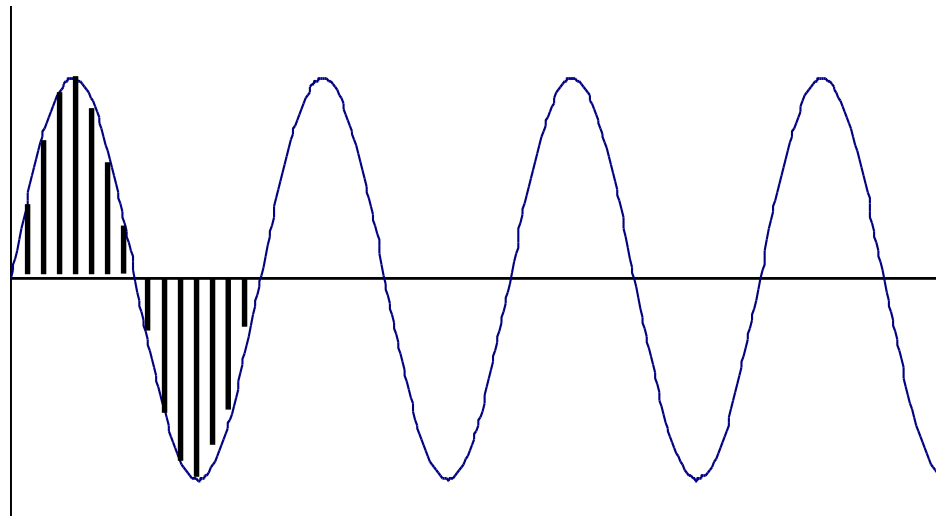
Temporal Masking



Digital Representation of Audio

- Must convert wave form to digital

- ☐ sample
- ☐ quantize
- ☐ compress



Nyquist Theorem

For lossless digitization, the sampling rate should be at least twice the maximum frequency response.

- In mathematical terms:

$$f_s > 2 * f_m$$


- where f_s is sampling frequency and f_m is the maximum frequency in the signal

Nyquist Limit

- max data rate = $2 H \log_2 V$ *bits/second*, where
 - H = bandwidth (in Hz)
 - V = discrete levels (bits per signal change)
- Shows the maximum number of bits that can be sent per second on a *noiseless* channel with a bandwidth of H, if V bits are sent per signal
 - Example: what is the maximum data rate for a 3kHz channel that transmits data using 2 levels (binary) ?
 - $(2 \times 3,000 \times \ln 2 = 6,000 \text{ bits/second})$

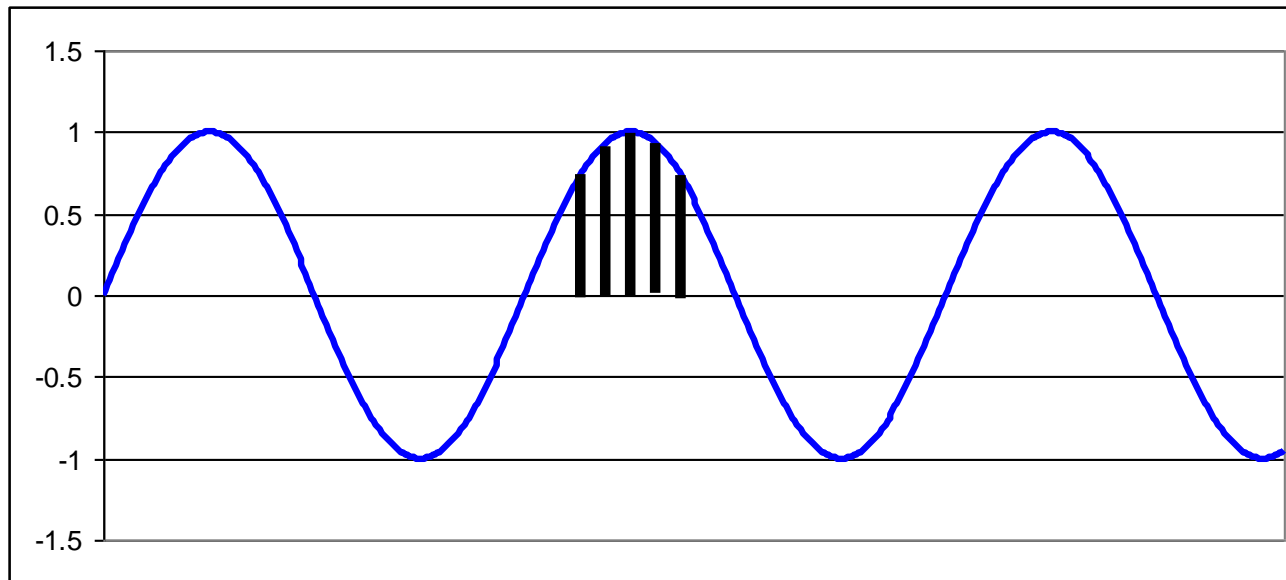
Sampling Ranges

- Auditory range 20Hz to 22.05 kHz
 - must sample up to 44.1kHz
 - common examples are 8.000 kHz, 11.025 kHz, 16.000 kHz, 22.05 kHz, and 44.1 KHz
- Speech frequency [200 Hz, 8 kHz]
 - sample up to 16 kHz
 - but typically 4 kHz to 11 kHz is used



Sampling Rates	Used As...
8000	Telephony Standard, Popular in UNIX Workstations
11000	Quarter of CD rate, Popular on Macintosh
16000	G.722 Standard (Federal Standard)
18900	CD-ROM XA Rate
22000	Half CD rate, Macintosh rate
32000	Japanese HDTV, British TV audio, Long play DAT
37800	CD XA Standard
44056	Professional audio industry
44100	CD Rate
48000	DAT Rate

Quantization



Quantization

- Typically use
 - 8 bits = 256 levels
 - 16 bits = 65,536 levels
- How should the levels be distributed?
 - Linearly? (PCM)
 - Perceptually? (u-Law)
 - Differential? (DPCM)
 - Adaptively? (ADPCM)

Pulse Code Modulation

■ Pulse modulation

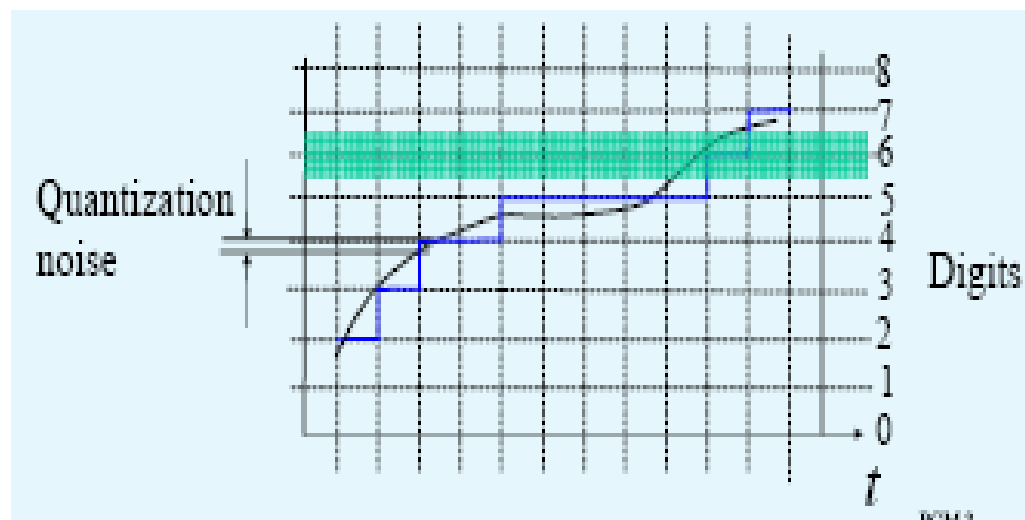
- Use discrete time samples of analog signals
- Transmission is composed of analog information sent at different times
- Variation of pulse amplitude or pulse timing allowed to vary continuously over all values

■ PCM

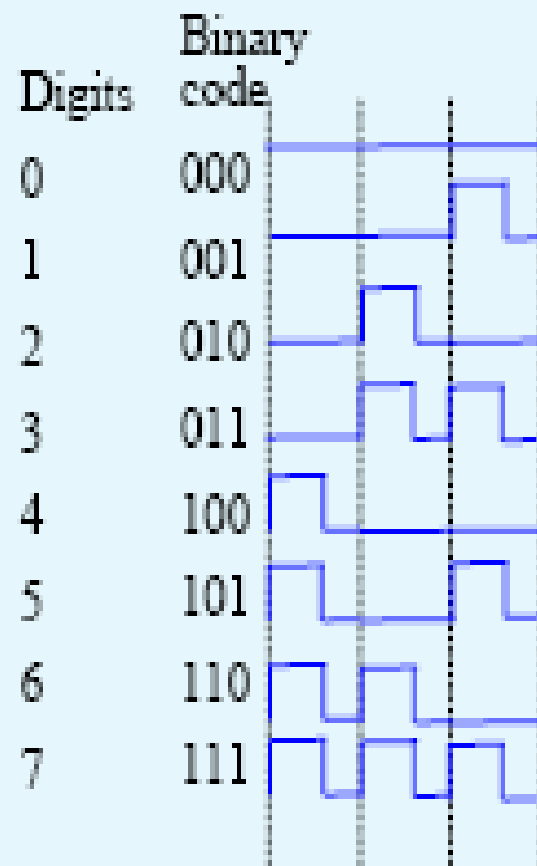
- Analog signal is quantized into a number of discrete levels

PCM Quantization and Digitization

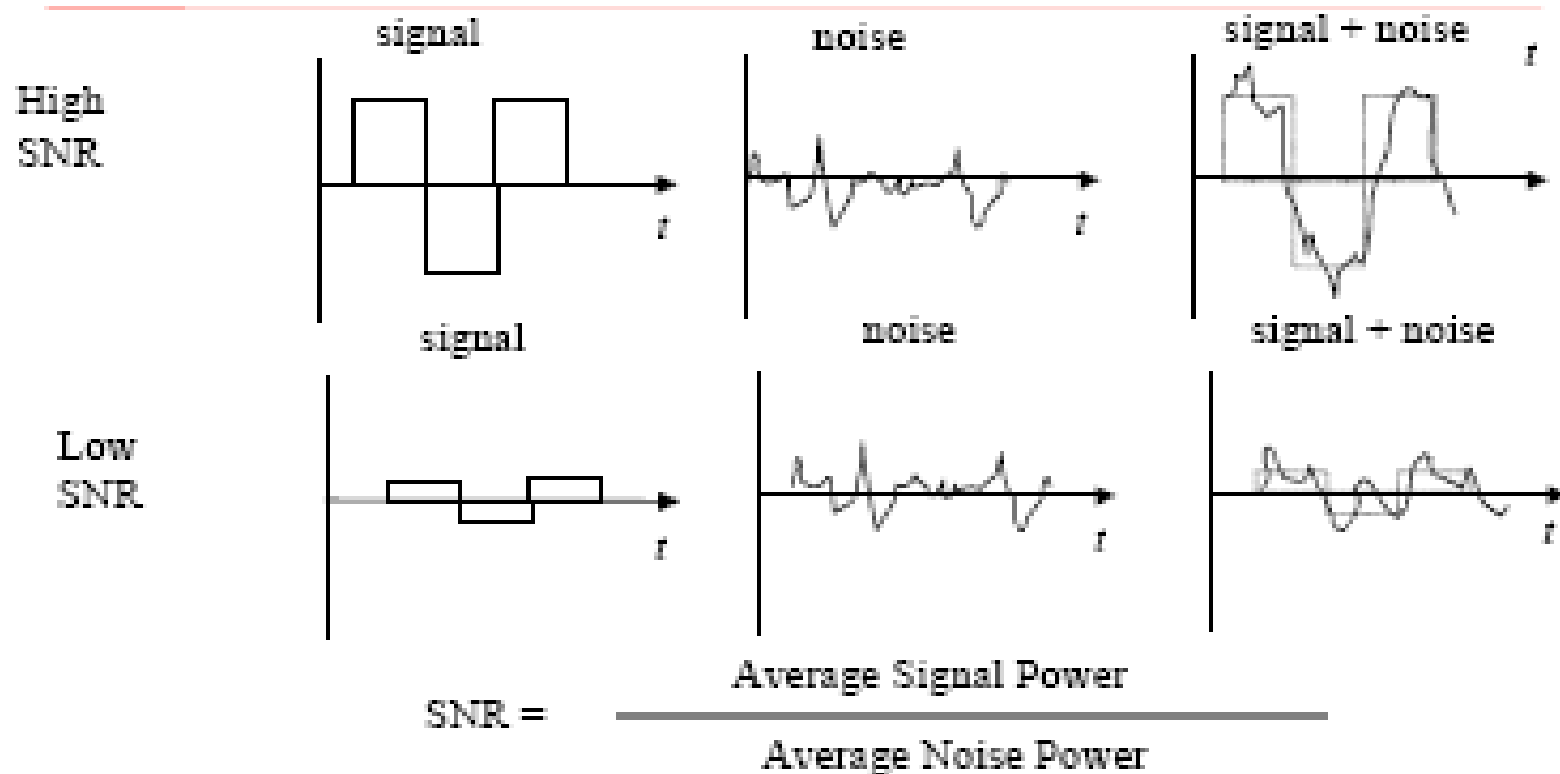
Quantization



Digitization



Signal-to-Noise Ratio



$$\text{SNR (dB)} = 10 \log_{10} \text{SNR}$$

Data Rates

- Data rate = sample rate * quantization * channel
- Compare rates for CD vs. mono audio
 - 8000 samples/second * 8 bits/sample * 1 channel
= 8 kBytes / second
 - 44,100 samples/second * 16 bits/sample *
2 channel = 176 kBytes / second \approx 10MB / minute



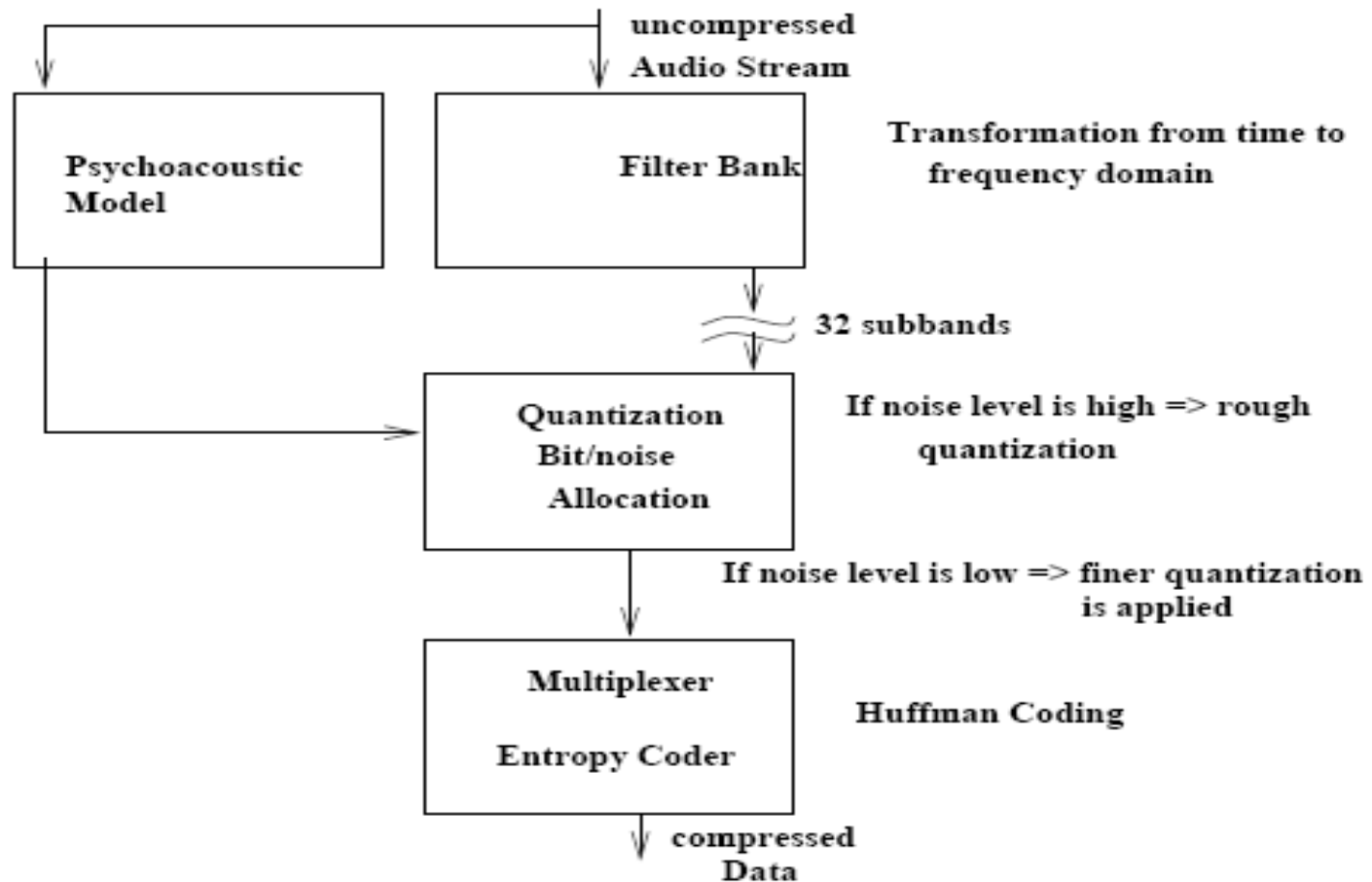
MPEG AUDIO CODING

MPEG Audio Encoding

■ Characteristics

- Precision 16 bits
- Sampling frequency: 32KHz, 44.1 KHz, 48 KHz
- 3 compression layers: Layer 1, Layer 2, Layer 3 (MP3)
 - Layer 3: 32-320 kbps, target 64 kbps
 - Layer 2: 32-384 kbps, target 128 kbps
 - Layer 1: 32-448 kbps, target 192 kbps

MPEG Audio Encoding Steps



MPEG Audio Filter Bank

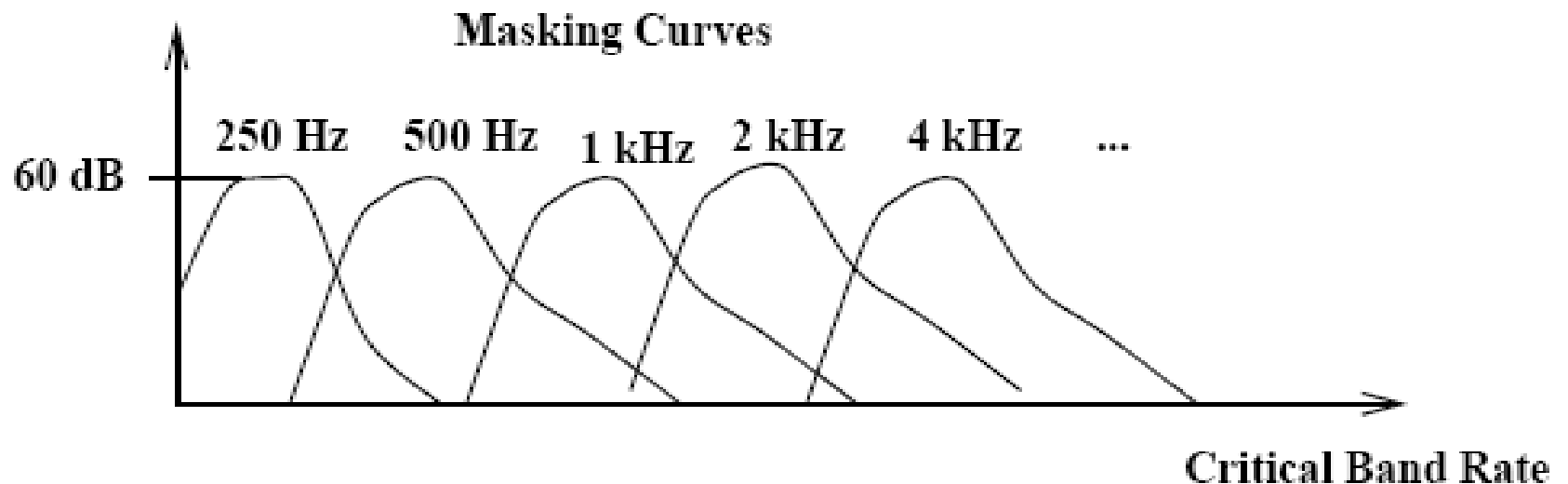
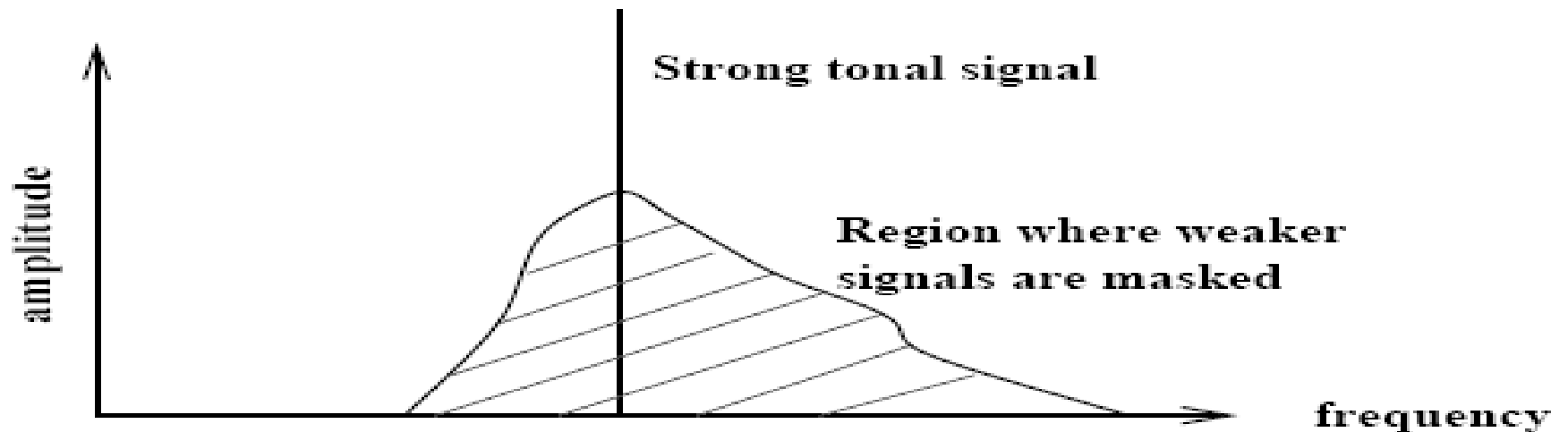
- Filter bank divides input into multiple sub-bands (32 equal frequency sub-bands)
- Sub-band i defined

$$S_t[i] = \sum_{k=0}^7 3 \sum_{j=0}^7 \cos\left(\frac{(2i+1)(k-16)\pi}{64}\right) * (C[k+64j] * x[k+64j])$$

- $i \in [0,31]$, $S_t[i]$ - filter output sample for sub-band i at time t , $C[n]$ – one of 512 coefficients, $x[n]$ – audio input sample from 512 sample buffer

MPEG Audio Psycho-acoustic Model

- MPEG audio compresses by removing acoustically irrelevant parts of audio signals
- Takes advantage of human auditory systems inability to hear quantization noise under auditory masking
- Auditory masking: occurs when ever the presence of a strong audio signal makes a temporal or spectral neighborhood of weaker audio signals imperceptible.



MPEG/audio divides audio signal into frequency sub-bands that approximate critical bands. Then we quantize each sub-band according to the audibility of quantization noise within the band

MPEG Audio Bit Allocation

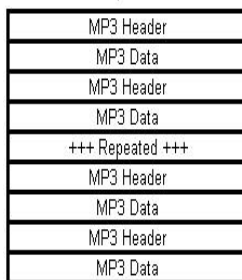
- This process determines number of code bits allocated to each sub-band based on information from the psycho-acoustic model
- Algorithm:
 1. Compute mask-to-noise ratio: $MNR = SNR - SMR$
 - Standard provides tables that give estimates for SNR resulting from quantizing to a given number of quantizer levels
 2. Get MNR for each sub-band
 3. Search for sub-band with the lowest MNR
 4. Allocate code bits to this sub-band.
 - If sub-band gets allocated more code bits than appropriate, look up new estimate of SNR and repeat step 1

MP3 Audio Format

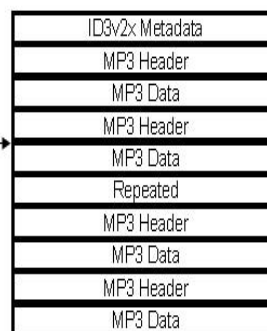
An MP3 File



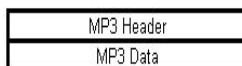
Internal Structure of An MP3 File



Note That The MP3 File Structure
Maybe 'encapsulated'
within an ID3 Tag



An MP3 Frame



Example
MP3 Header

FFFBA040

Color Coding shows binary bit mapping to hex values below

Detail Of An MP3
Header

Bits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32						
Binary	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0						
Hex	F			F			F						B						A				0									0						
																										Mode Extension (Used With Joint Stereo)												
Meaning	MP3 Sync Word												Version	Layer			Error Protection					Bit Rate					Frequency			Pad. Bit	Priv. Bit	Mode				Copy	Original	Emphasis
Value	Sync Word												1 = MPEG	01 = Layer 3			1=No					1010 = 160					00 = 44100 Hz			0 = Frame is not padded	Unknown	01=Joint Stereo		0 = Intensity Stereo Off	0 = MS Stereo Off	0=Not Copyrighted	0=Copy Of Original Media	00=None

Source: <http://wiki.hydrogenaudio.org/images/e/ee/Mp3filestructure.jpg>

MPEG Audio Comments

- Precision of 16 bits per sample is needed to get good SNR ratio
- Noise we are getting is quantization noise from the digitization process
- For each added bit, we get 6dB better SNR ratio
- Masking effect means that we can raise the noise floor around a strong sound because the noise will be masked away
- Raising noise floor is the same as using less bits and using less bits is the same as compression

Conclusions

- Audio is very important part of multimedia with its own characteristics and challenges
 - Audio quality is more important than video quality!!
- Systems and networks need to carefully consider the **compression techniques** audio uses to deliver high quality audio
 - Most well known MP3 and now coming MP4
- Systems and networks need to carefully consider psycho-acoustic features, head-related transfer functions, to enable **immersive audio**
 - **Advanced audio coding (AAC developed after MP3), Dolby AC-4, MPEG-H 3D Audio**



References

- Slides based on Literature Reference
 - D. Pan, “A Tutorial on MPEG/Audio Compression”, 2002, Readings in Multimedia Computing and Networking, Editors Kevin Jeffay and HongJiang Zhang