Multimedia signal representation

Omar A. Nasr

omaranasr@ieee.org

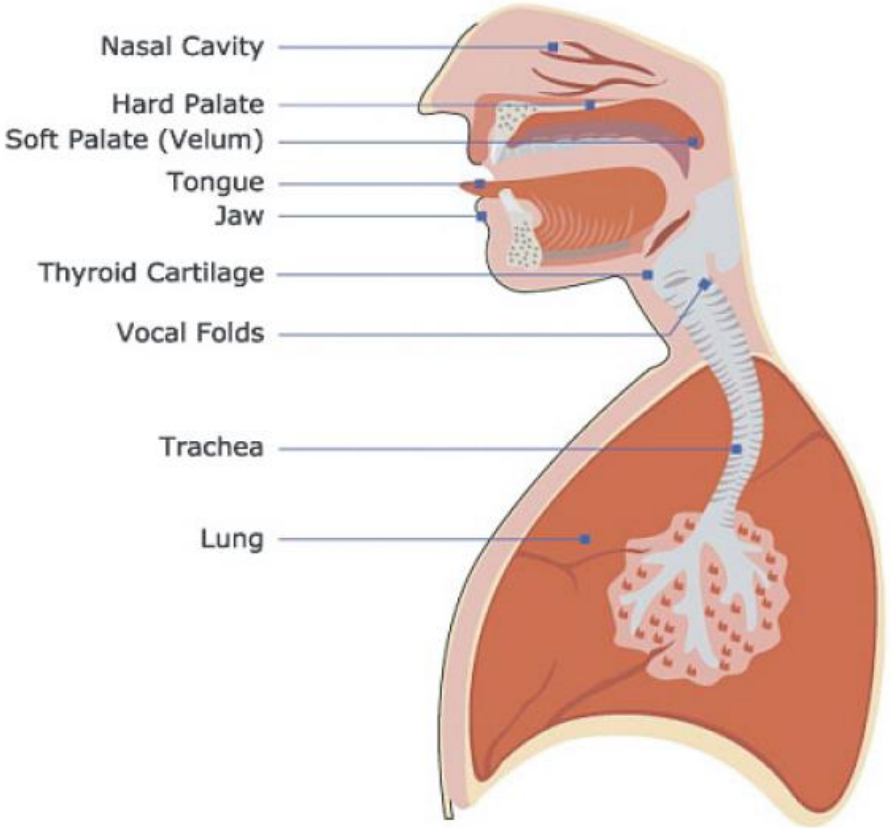
Feb, 2015

Contents

- Speech/audio signal representation
- Image representation
- Video representation (next lecture)
- This lecture is about “uncoded” formats

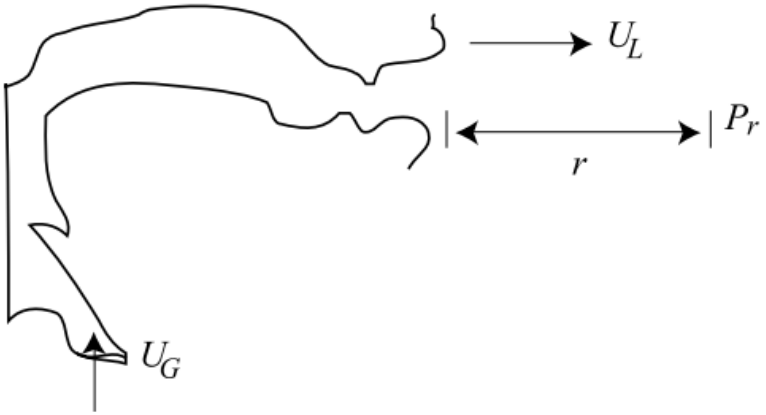
Speech/audio signal representation

- Speech production system



Acoustic Theory of Speech Production

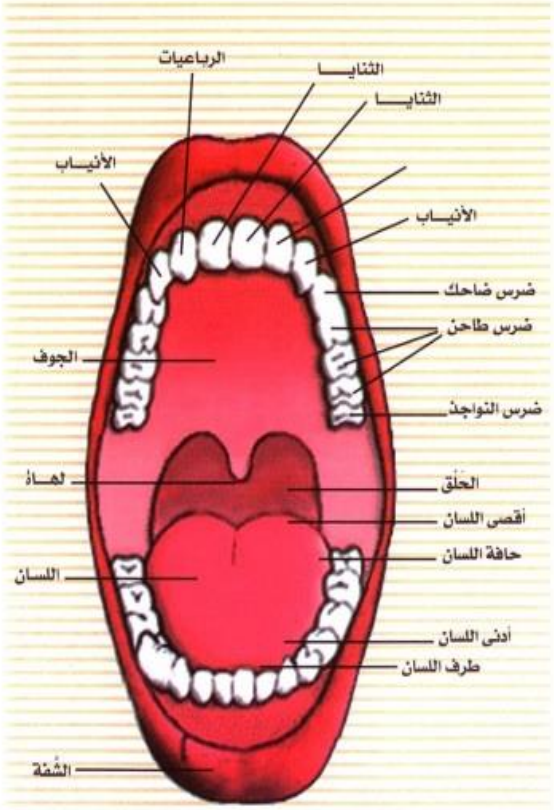
- The acoustic characteristics of speech are usually modelled as a sequence of source, vocal tract filter, and radiation characteristi



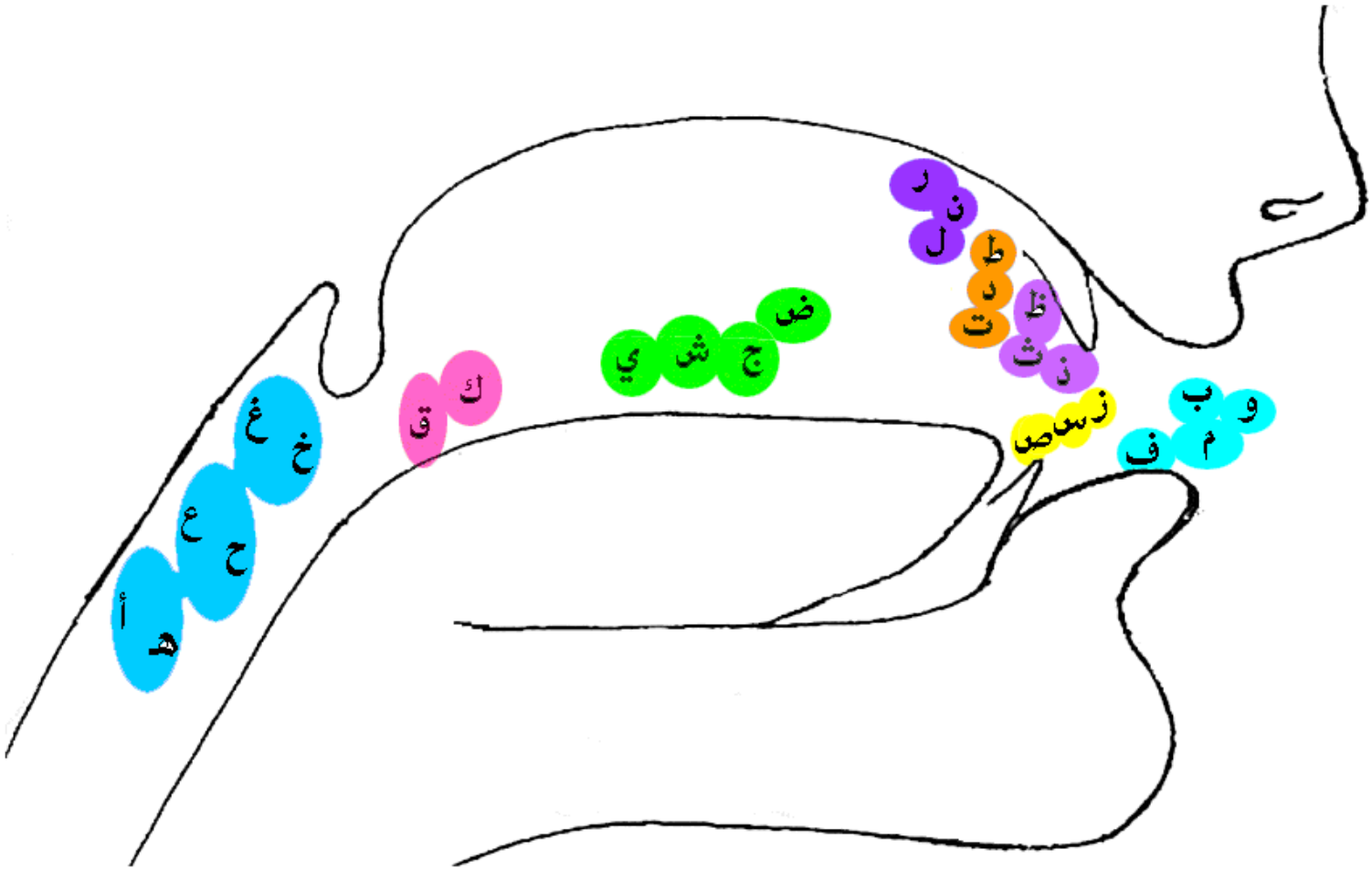
$$P_r(j\Omega) = S(j\Omega)\,T(j\Omega)\,R(j\Omega)$$

مخارج الحروف الـ ١٧

تسلسل	المخرج	تفصيل المخرج	الحروف التي تخرج منه
١	الجوف (١) مخرج	هو الخلاء داخل الفم.	الألف المفتوح ما قبلها والواو المضموم ما قبلها والياء المكسور ما قبلها ا - و - ي
٢	الحلق (٣) مخارج	من أقصاه (البعيد أي يكون أقرب للصدر).	ء - هـ
٣		من وسطه (من الحَنجرة).	ع - ح
٤		من أدناه (أقرب للفم).	غ - خ
٥	اللسان (١٠) مخارج	من أقصاه أي ءاخره مما يلي الحلق من فوق.	ق
٦		من أقصاه وما تحته من الحنك الأعلى.	ك
٧		من وسطه ومع ما يحاذيه من الحنك الأعلى.	ج - ش - ي
٨		من حافتيه أي جانبيه.	ض
٩		من أول حافة اللسان مع ما يليها من الحنك الأعلى	ل
١٠		من طرف اللسان مع ما يليه من الحنك الأعلى تحت اللام قليلاً.	ن
١١		من طرف اللسان وهو أدخل إلى ظهر اللسان قليلاً لانحرافه عن اللام.	ر
١٢		من طرف اللسان وأصول عليا الثنايا.	ط - د - ت
١٣		من طرف اللسان من بين الثنايا العليا والسفلى.	ص - ز - س
١٤		من طرف اللسان والثنايا العليا.	ظ - ذ - ث
١٥	الشفتان (٢) مخارج	من باطن الشفة السفلى.	ف
١٦		من بين الشفتين.	و - ب - م
١٧	الخيشوم (١) مخرج	من منتهى الأنف أي أقصاه.	الغُنّة



الحروف الحلقية وهي التي تخرج من الحلق وهي:
الألف، الهاء، العين، الغين، الحاء، الخاء


ر
ن
ل
ط
د
ت
ظ
ث
ذ
ض
ي
ش
ج
ك
ق
غ
خ
ع
ح
أ
هـ
ز
س
ص
ب
و
م
ف

صفات الحروف

الرقم	الصفة	التعريف	الحروف
١	الهمس	جريان النفس عند النطق بالحرف لضعف الاعتماد على المخرج	فحثه شخص سكت
٢	الجهر	انحباس جري النفس عند النطق بحروفه لقوة الاعتماد على المخرج	الباقي بعد الهمس
٣	الشدة	انحباس جري الصوت عند النطق بالحرف لكمال الاعتماد على المخرج	أجد قط بكت
٤	التوسط	اعتدال الصوت عند النطق بالحرف لعدم كمال انحباسه كما في الشدة، وعدم كمال جريانه كما في الرخاوة	لن عمر
٥	الرخاوة	جريان الصوت مع الحرف لضعف الاعتماد على المخرج	ما عدا حروف التوسط والشدة
٦	الاستعلاء	ارتفاع اللسان إلى الحنك الأعلى عند النطق بالحرف	خص ضغط قظ
٧	الاستفال	انخفاض اللسان (انحطاطه عن الحنك الأعلى إلى قاع الفم)	الباقي بعد الاستعلاء
٨	الإطباق	تلاصق ما يحاذي اللسان من الحنك الأعلى على اللسان	ص - ض - ط - ظ
٩	الانفتاح	تجافي كل من طائفتي اللسان والحنك الأعلى عن الأخرى حتى يخرج النفس	ما عدا حروف الإطباق
١٠	المد واللين	امتداد الصوت وخروج الحرف في لين وعدم كلفة	ا - و - ي
١١	الصفير	حدة الصوت	ص - س - ز
١٢	التفشي	انتشار خروج الريح وانبساطه	ش - (ث على خلاف)

٦	الاستعلاء	ارتفاع اللسان إلى الحنك الأعلى عند النطق بالحرف	خص ضغط قظ
٧	الاستفال	انخفاض اللسان (انحطاطه عن الحنك الأعلى إلى قاع الفم)	الباقي بعد الاستعلاء
٨	الإطباق	تلاصق ما يحاذي اللسان من الحنك الأعلى على اللسان	ص - ض - ط - ظ
٩	الانفتاح	تجافي كل من طائفتي اللسان والحنك الأعلى عن الأخرى حتى يخرج النفس	ما عدا حروف الإطباق
١٠	المد واللين	امتداد الصوت وخروج الحرف في لين وعدم كلفة	ا - و - ي
١١	الصفير	حدة الصوت	ص - س - ز
١٢	التفشي	انتشار خروج الريح وانبساطه	ش - (ث على خلاف)
١٣	الاستطالة	امتداد الصوت من أول إحدى حافتي اللسان إلى أخرها	ض
١٤	التكرير	تضعيف يوجد في جسم الراء لارتعاد طرف اللسان بها	ر
١٥	الانحراف	خروج من صفة إلى صفة	ل - ر
١٦	الغنة	صفة لازمة للنون والميم (وهو الصوت الزائد المنبعث عن الخيشوم)	م - ن
١٧	القلقلة	اضطراب المخرج عند النطق بالحرف ساكنا حتى يسمع له نبرة قوية	قطب جد [أو] بجد قط
١٨	النفخ	صوت حادث عند خروج حرفه لضغطه عن موضعه وهو دون القلقلة	ظ - ز - ض - ذ

Phonemes in American English

PHONEME	EXAMPLE	PHONEME	EXAMPLE	PHONEME	EXAMPLE
/iʸ/	beat	/s/	see	/w/	wet
/ɪ/	bit	/š/	she	/r/	red
/eʸ/	bait	/f/	fee	/l/	let
/ɛ/	bet	/θ/	thief	/y/	yet
/æ/	bat	/z/	z	/m/	meet
/ɑ/	Bob	/ž/	Gigi	/n/	neat
/ɔ/	bought	/v/	v	/ŋ/	sing
/ʌ/	but	/ð/	thee	/č/	church
/oʷ/	boat	/p/	pea	/ǰ/	judge
/ʊ/	book	/t/	tea	/h/	heat
/uʷ/	boot	/k/	key		
/ɝ/	Burt	/b/	bee		
/ɑʸ/	bite	/d/	Dee		
/ɔʸ/	Boyd	/g/	geese		
/ɑʷ/	bout				
/ə/	about				

MIT

Speech Sounds of American English

- There are over 40 speech sounds in American English which can be organized by their basic manner of production

Manner Class	Number
Vowels	18
Fricatives	8
Stops	6
Nasals	3
Semivowels	4
Affricates	2
Aspirant	1

- Vowels, glides, and consonants differ in degree of constriction
- Sonorant consonants have no pressure build up at constriction
- Nasal consonants lower the velum allowing airflow in nasal cavi

Phonemes in American English

Video

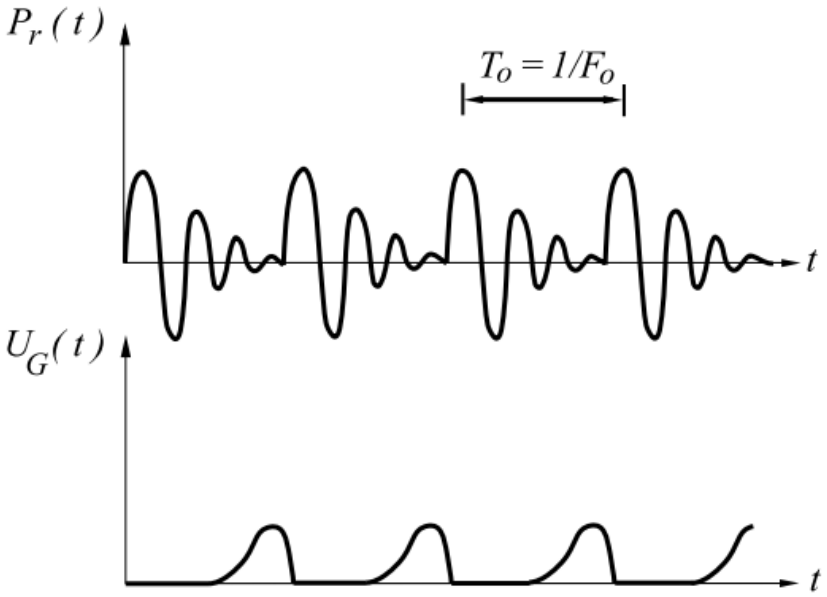
Sample of waveforms

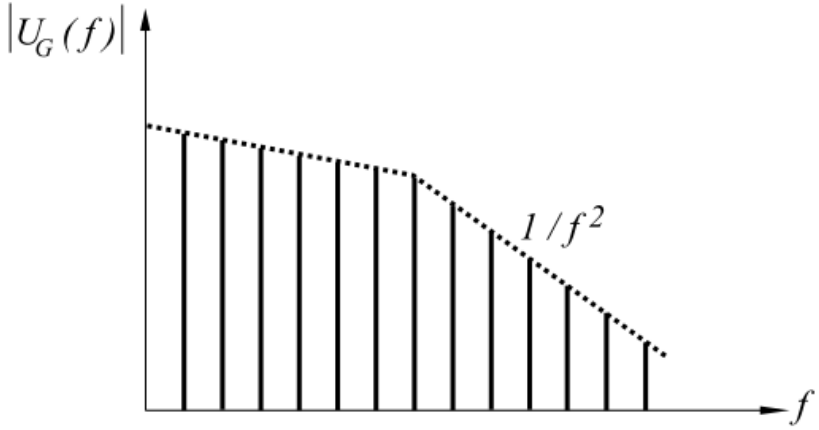
- Time domain analysis
- Spectrogram
- Pitch period
- Voiced vs. unvoiced

Demo using Audacity

Sound Source: Vocal Fold Vibration

Modelled as a volume velocity source at glottis, $U_G(j\Omega)$



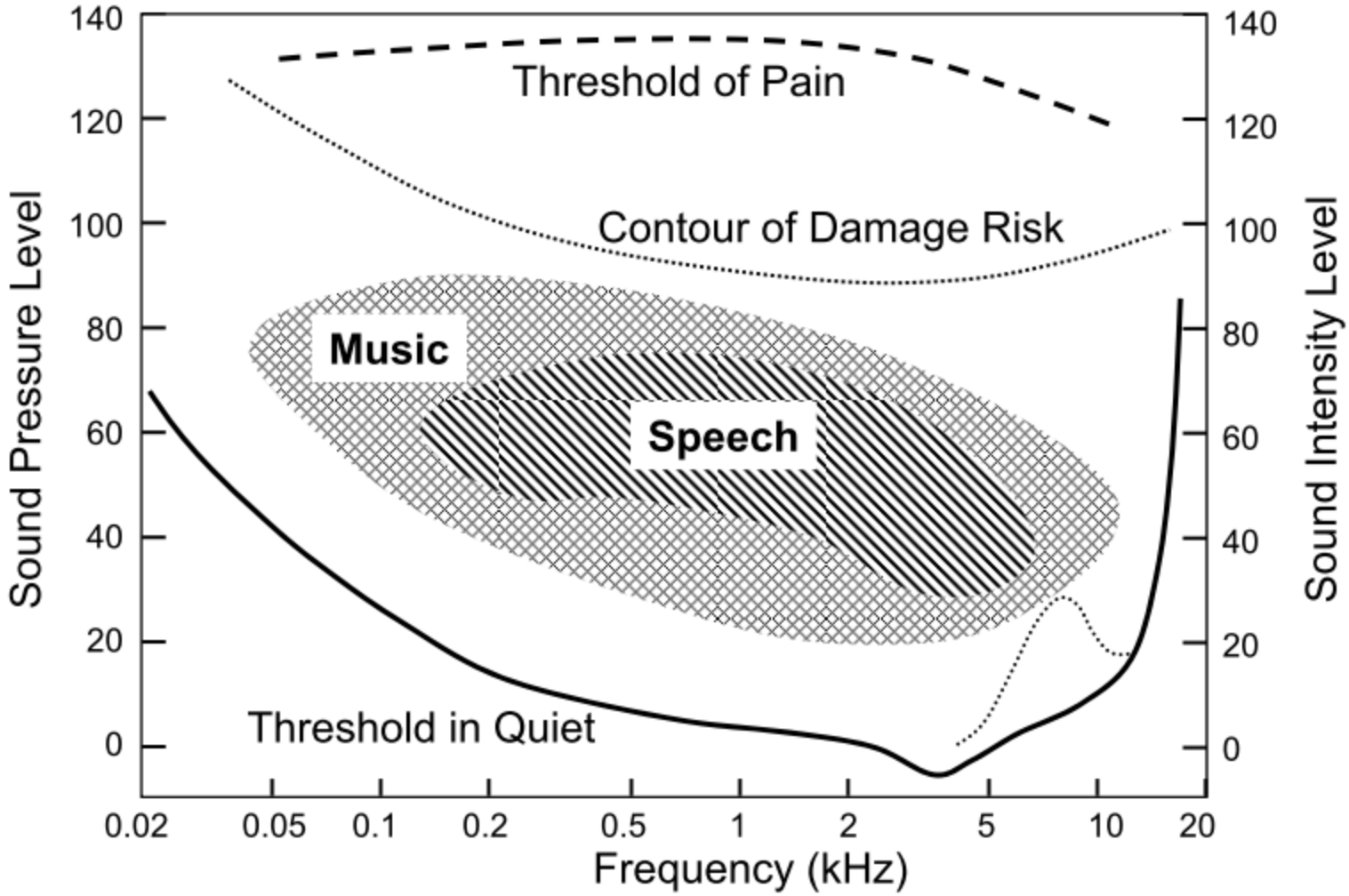


	F_0 ave (Hz)	F_0 min (Hz)	F_0 max (Hz)
Men	125	80	200
Women	225	150	350
Children	300	200	500

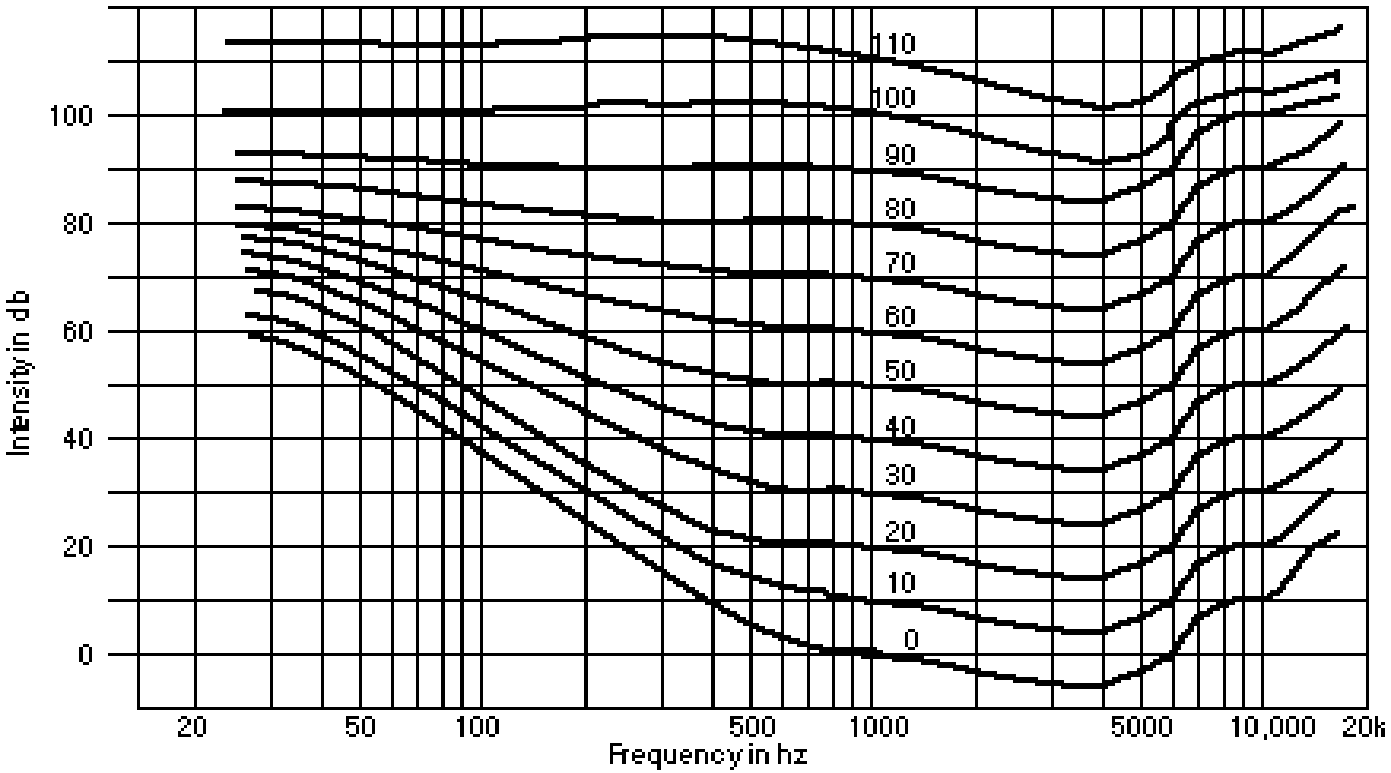
Auditory system

- Speech: up to almost 8kHz
- Hearing: 20Hz to 20kHz
- Why sampling is important?

Range of Human Hearing

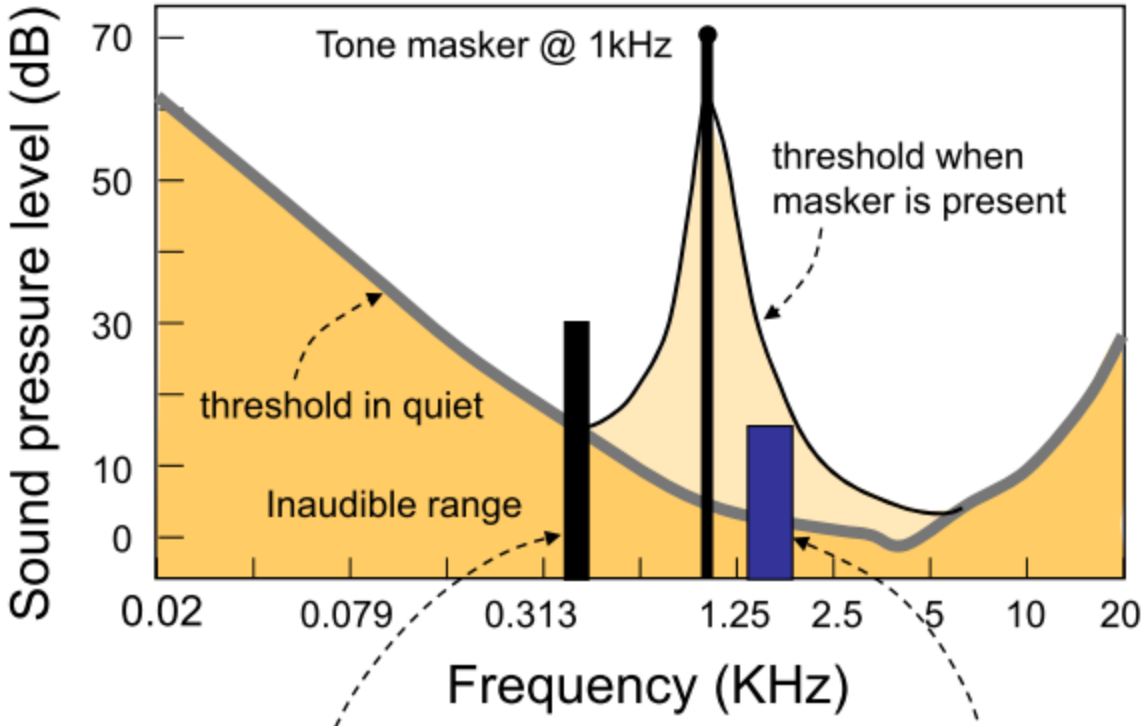


Fletcher-Munson Contours



Each contour represents an equal perceived sound

Auditory Masking

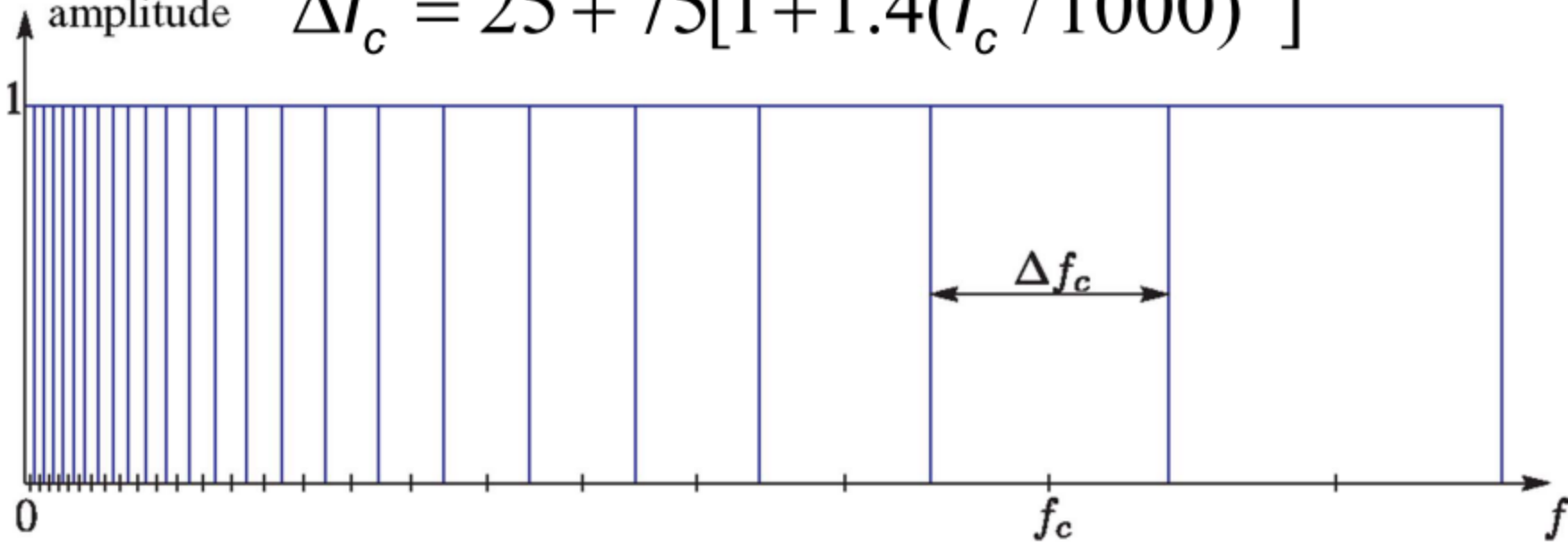


Signal perceptible even in the presence of the tone masker

Signal not perceptible due to the presence of the tone masker

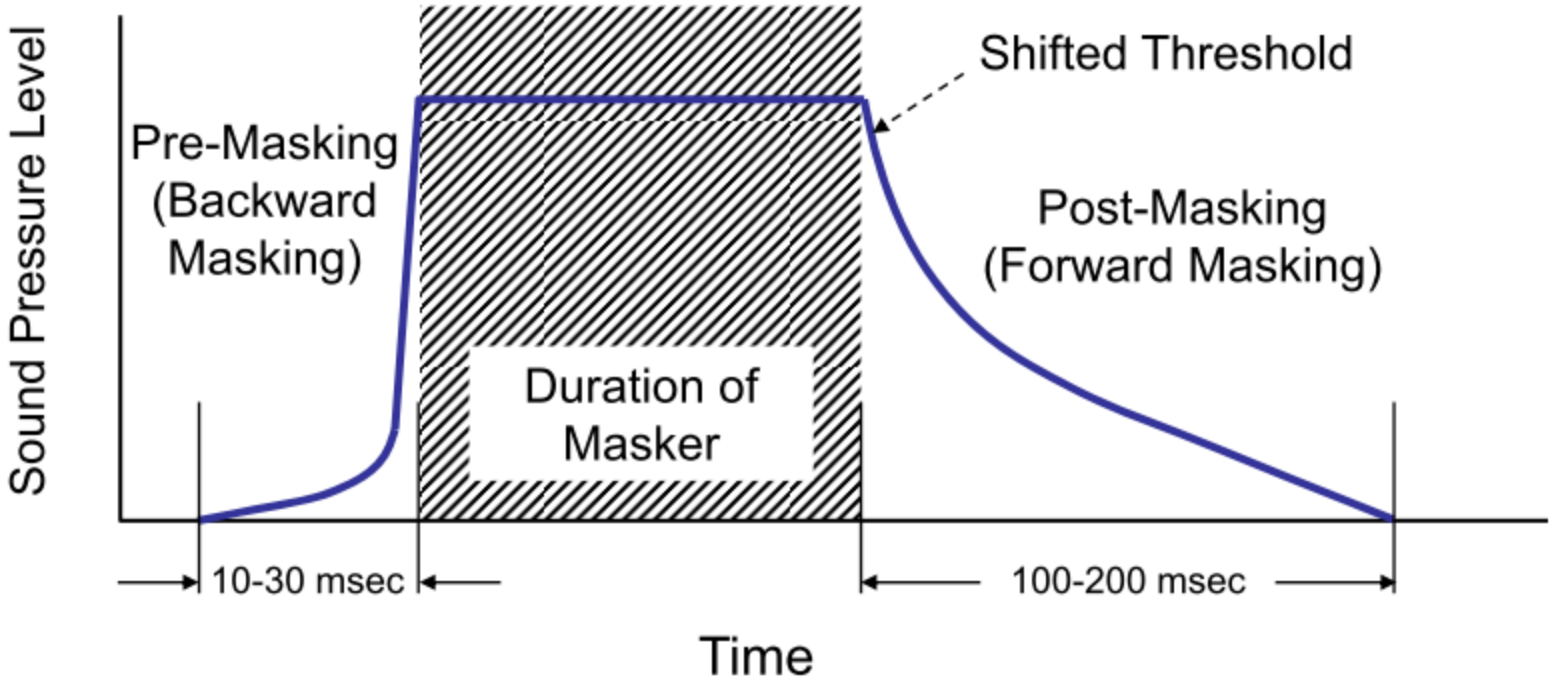
Critical Bands

$$\Delta f_c = 25 + 75[1 + 1.4(f_c / 1000)^2]^{0.69}$$

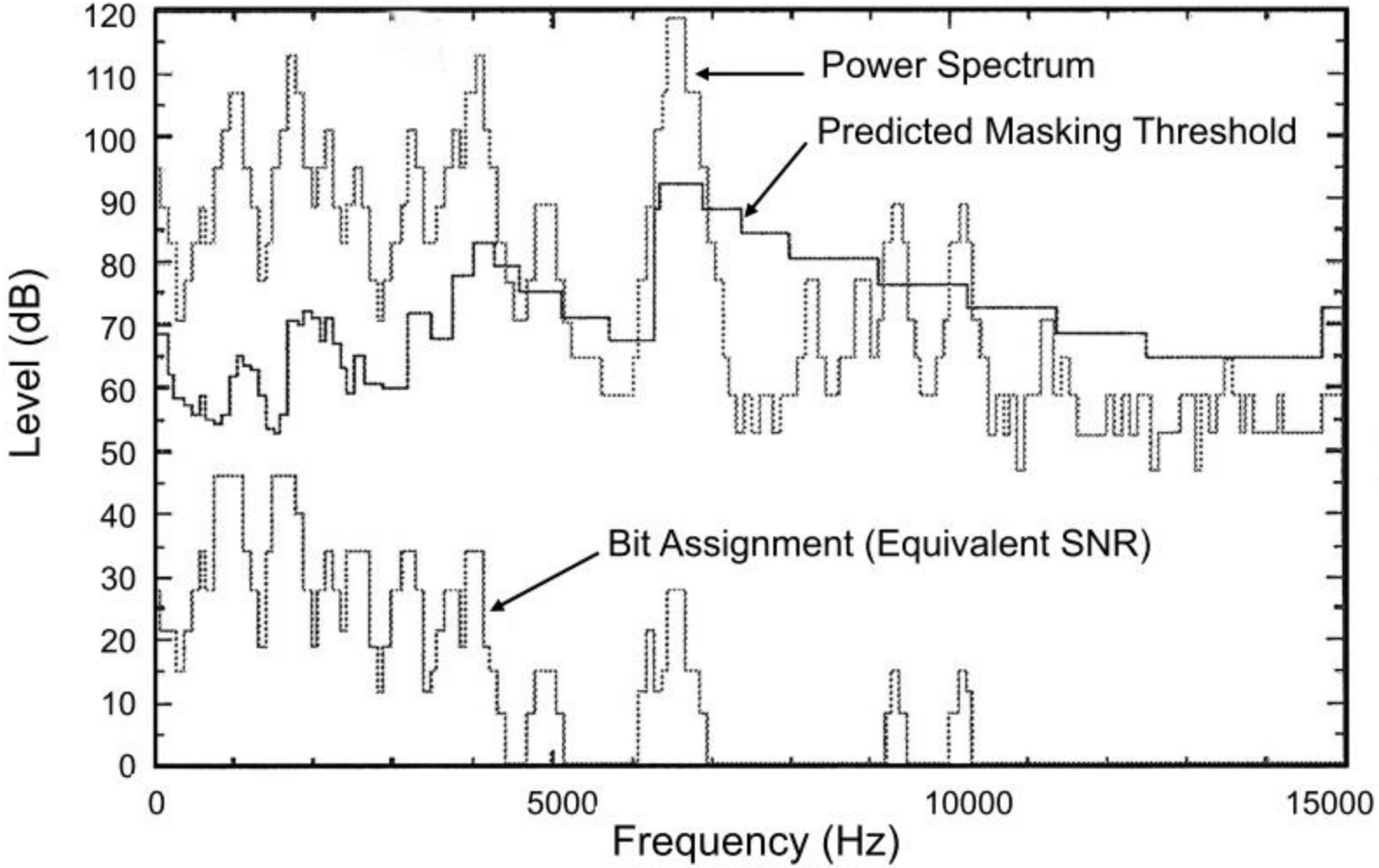


- Idealized basilar membrane filter bank
  - Center Frequency of Each Bandpass Filter: f_c
  - Bandwidth of Each Bandpass Filter: Δf_c
  - Real BM filters overlap significantly

Temporal Masking



Exploiting Masking in Coding

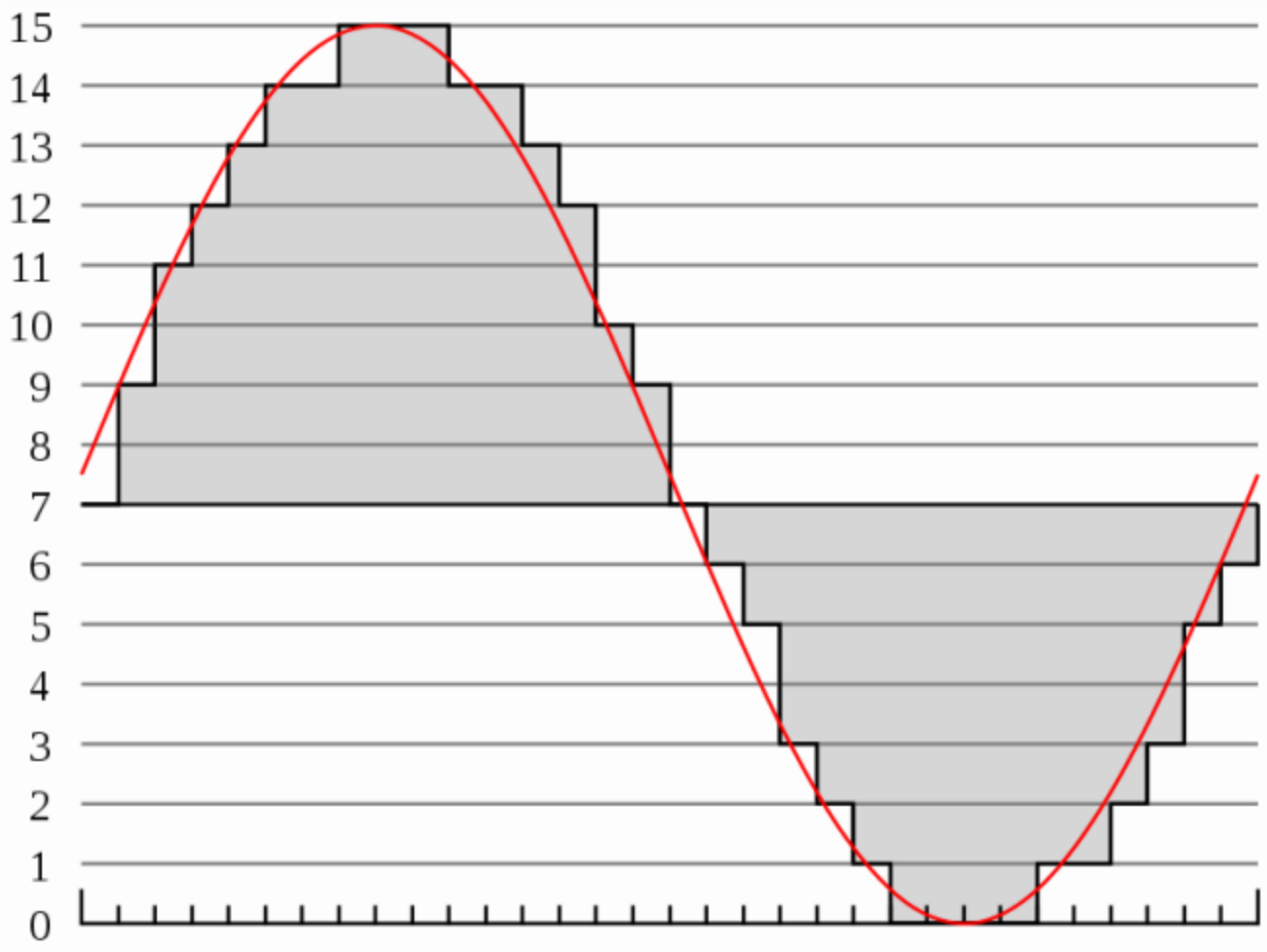


Sampling Ranges

- Auditory range 20Hz to 22.05 kHz
  - must sample up to 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

Sampling and 4-bit quantization



Source: http://en.wikipedia.org/wiki/Digital_audio

Quantization

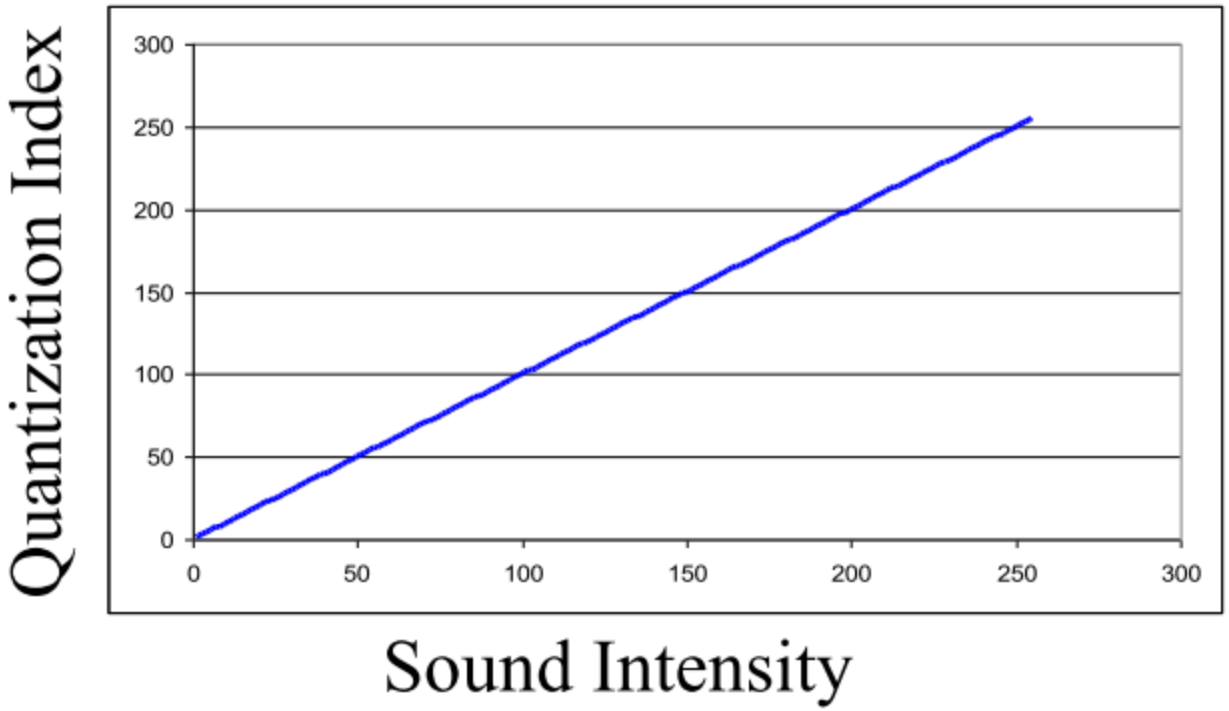
- Typically use
  - 8 bits = 256 levels
  - 16 bits = 65,536 levels
- How should the levels be distributed?
  - Linearly? (PCM)
  - Perceptually? (u-Law)
  - Differential? (DPCM)
  - Adaptively? (ADPCM)

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

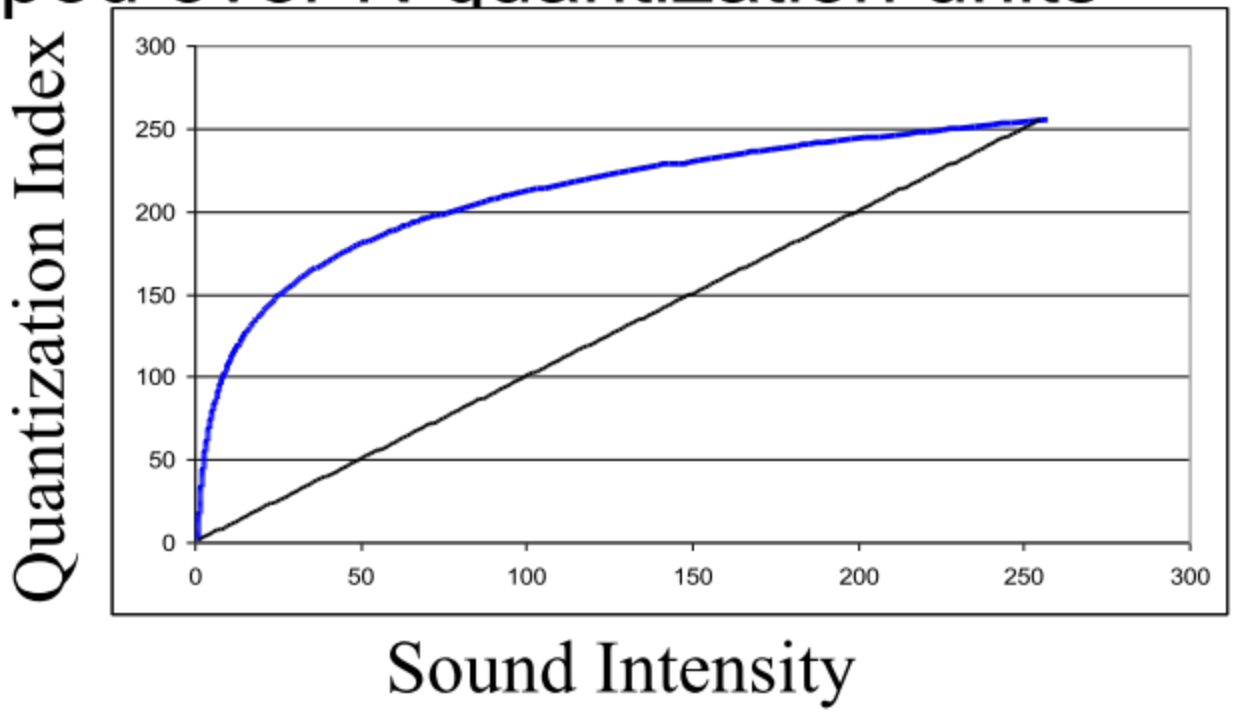
Linear Quantization (PCM)

- Divide amplitude spectrum into N units (for $\log_2 N$ bit quantization)



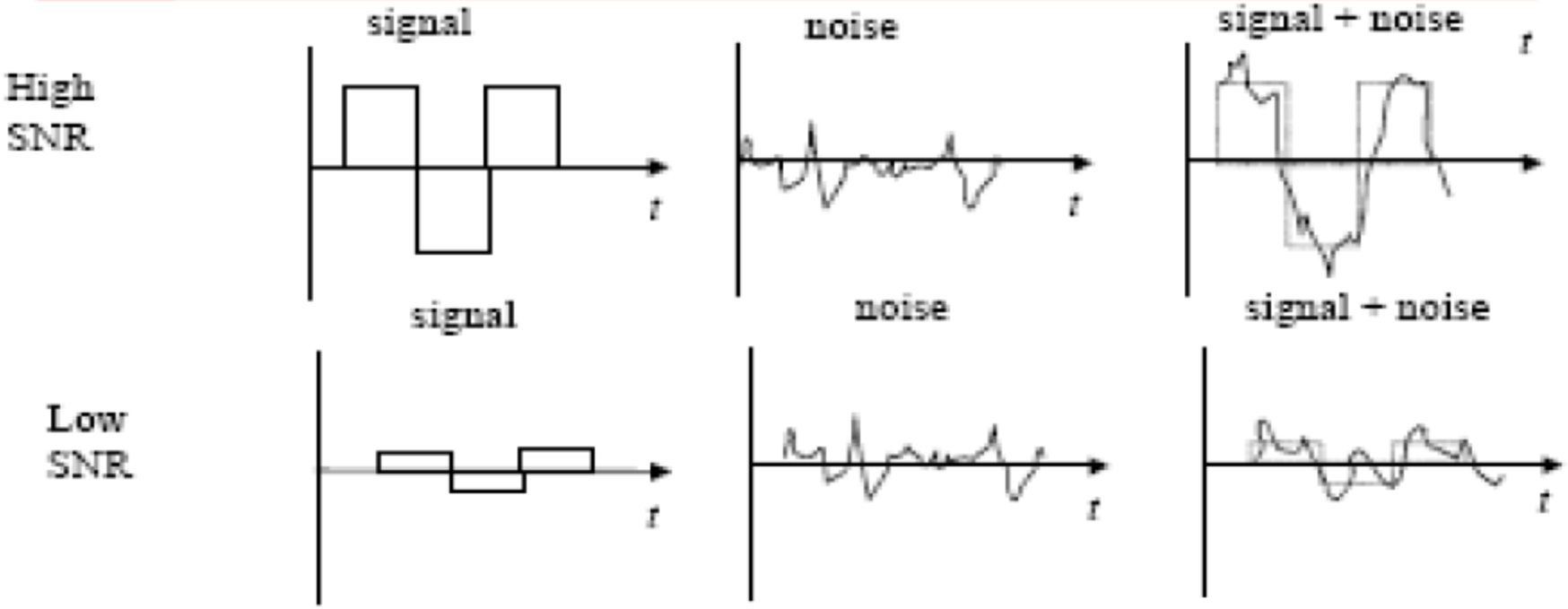
Perceptual Quantization (u-Law)

- Want intensity values logarithmically mapped over N quantization units



Signal-to-Noise Ratio

(metric to quantify quality of digital audio)



$$\text{SNR} = \frac{\text{Average Signal Power}}{\text{Average Noise Power}}$$

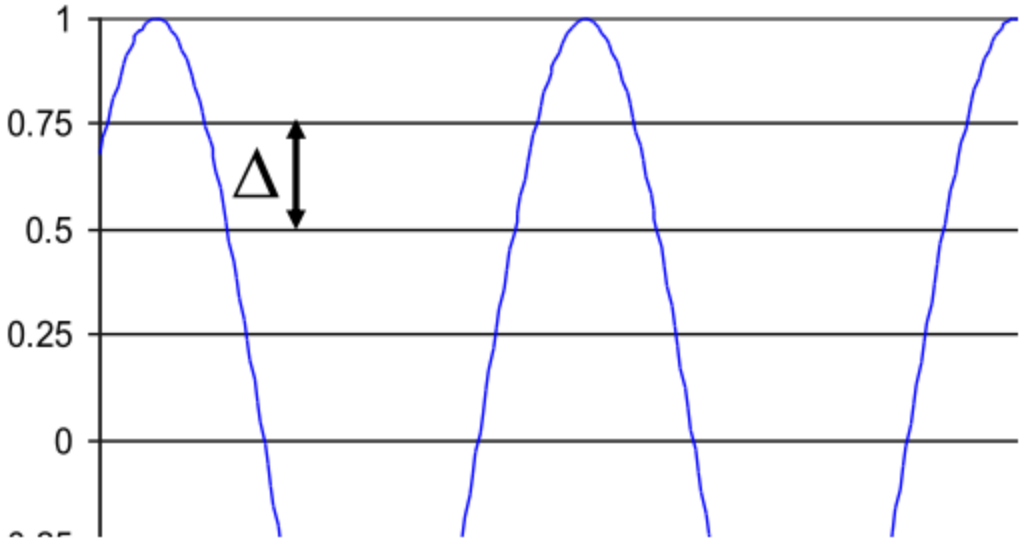
$$\text{SNR (dB)} = 10 \log_{10} \text{SNR}$$

Quantization Error

- Difference between actual and sampled value
  - amplitude between [-A, A]
  - quantization levels = N

$$\Delta = \frac{2A}{N}$$

- e.g., if A = 1, N = 8, Δ = 1/4



Compute Signal to Noise Ratio

- Signal energy = $\frac{A^2}{2}$; Noise energy = $\frac{\Delta^2}{12}$; $\Delta = \frac{2A}{N}$
- Noise energy = $\frac{A^2}{3 \cdot N^2}$
- Signal to noise = $10 \log \frac{3N^2}{2}$
- Every bit increases SNR by ~ 6 decibels

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 ~= 10MB / minute

Image representation
Demo – MATLAB- GIMP

Image Concepts

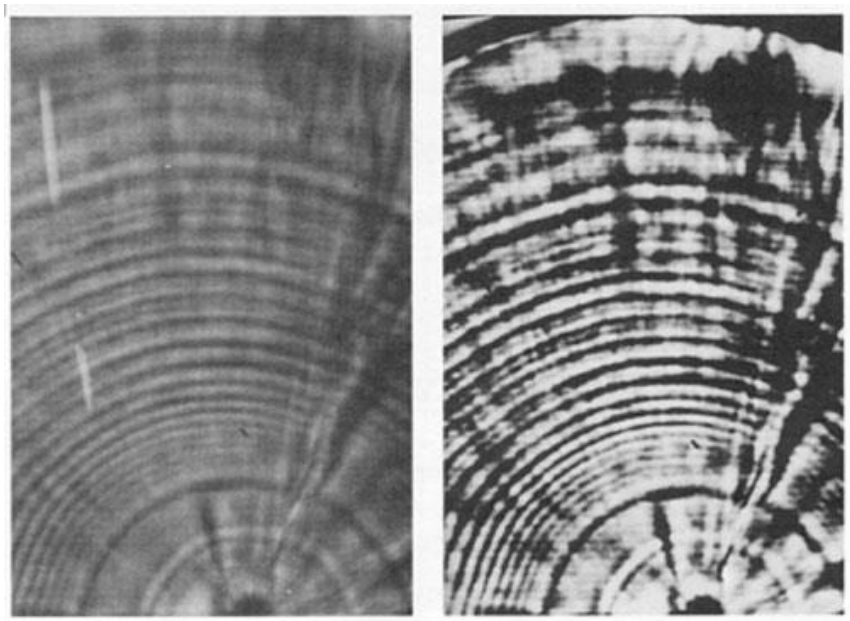
- An image is a function of intensity values over a 2D plane
  $I(r,s)$
- Sample function at discrete intervals to represent an image in digital form
  - matrix of intensity values for each color plane
  - intensity typically represented with 8 bits
- Sample points are called pixels

Digital Images

- Samples = pixels
- Quantization = number of bits per pixel
- Example: if we would sample and quantize standard TV picture (525 lines) by using VGA (Video Graphics Array), video controller creates matrix 640x480pixels, and each pixel is represented by 8 bit integer (256 discrete gray levels)

Image Representations

- Black and white image
  - single color plane with 2 bits
- Grey scale image
  - single color plane with 8 bits
- Color image
  - three color planes each with 8 bits
  - RGB, CMY,YIQ, etc.
- Indexed color image
  - single plane that indexes a color table
- Compressed images
  - TIFF, JPEG, BMP, etc.



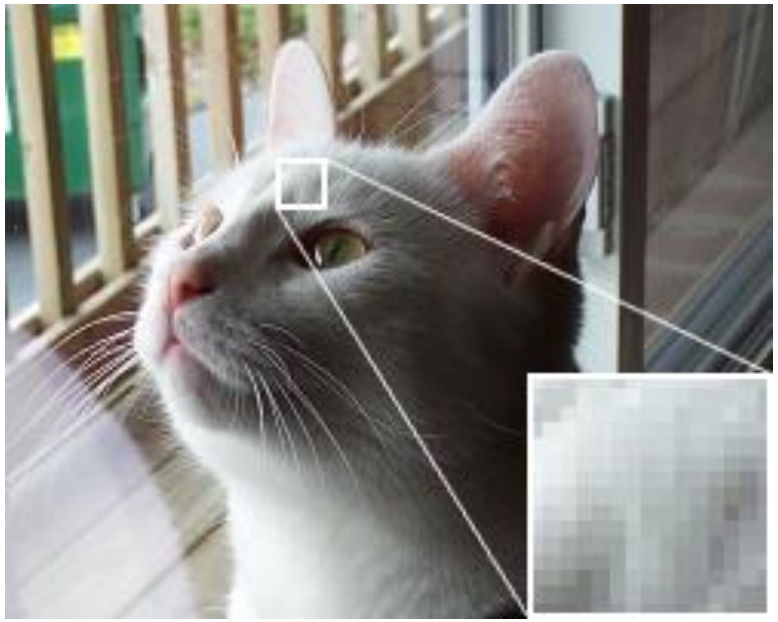
4 gray levels

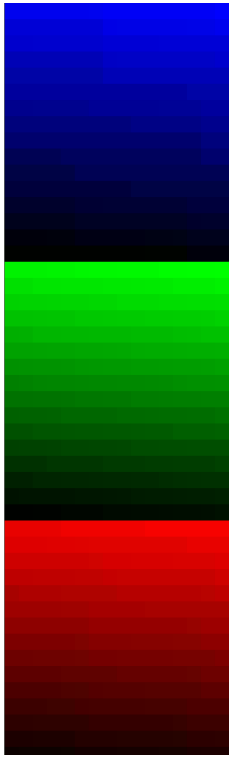
2gray levels

Digital Image Representation (3 Bit Quantization)

111	111	011	011	011	011	111	111
111	011	111	111	111	111	011	111
000	111	001	111	111	001	111	000
010	111	111	111	111	111	111	010
000	111	100	111	111	100	111	000
000	111	111	100	100	111	111	000
111	000	111	111	111	111	000	111
111	111	000	000	000	000	111	111

Color Quantization
Example of 24 bit RGB Image





24-bit Color Monitor

Image Representation Example

24 bit RGB Representation (uncompressed)

128	135	166	138	190	132
129	255	105	189	167	190
229	213	134	111	138	187

128	138
129	189
229	111

135	190
255	167
213	138

166	132
105	190
134	187

Color Planes

Image Processing Function: 1. Filtering

- Filter an image by replacing each pixel in the source with a weighted sum of its neighbors
- Define the filter using a *convolution mask*, also referred to as a *kernel*
  - non-zero values in small neighborhood, typically centered around a central pixel
  - generally have odd number of rows/columns

Convolution Filter

100	100	100	100	100
100	100	50	50	100
100	100	100	100	100
100	100	100	100	100
100	100	100	100	100

X

	0	1	0	
	0	0	0	
	0	0	0	

=

100	100	100	100	100
100	100	50	50	100
100	100	50	100	100
100	100	100	100	100
100	100	100	100	100

Mean Filter

20	12	14	23
45	15	19	33
55	34	81	22
8	64	49	95

Subset of image

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Convolution filter

Mean Filter

20	12	14	23
45	15	19	33
55	34	81	22
8	64	49	95

Subset of image

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Convolution filter

Common 3x3 Filters

- Low/High pass filter

$$\frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} \qquad \begin{bmatrix}-1 & -1 & -1\\ -1 & 9 & -1\\ -1 & -1 & -1\end{bmatrix}$$

- Blur operator

$$\frac{1}{13}\begin{bmatrix}1 & 2 & 1\\ 2 & 1 & 2\\ 1 & 2 & 1\end{bmatrix}$$

- H/V Edge detector

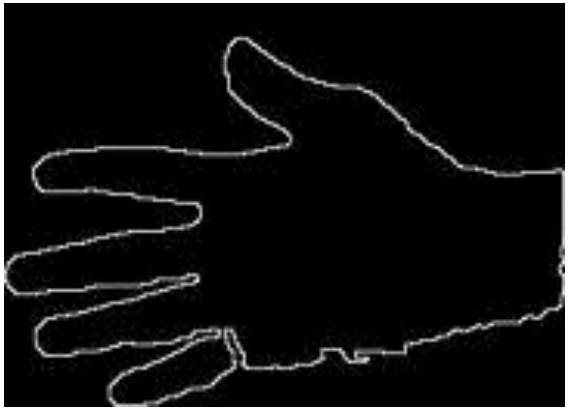
$$\begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix} \qquad \begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}$$

Example

Image Function: 2. Edge Detection

- Identify areas of strong intensity contrast
  - filter useless data; preserve important properties

- Fundamental technique
  - e.g., use gestures as input
  - identify shapes, match to templates, invoke commands



Edge Detection

Basic Method of Edge Detection

- Step 1: filter noise using mean filter
- Step 2: compute spatial gradient
- Step 3: mark points > *threshold* as edges