ANDY W – 365 CheatSheet 10/23/2025
Audio Signal - vibrates back and forth and produces a longitudinal pressure. Vibrates a magnet and create signal voltage on a coil.
Amplitude X Time = not linear, log scale
Threshold of hearing: minimum sound level human can hear
Digitization: conversion to a stream of numbers (to integers)
Bel decibel(dB) ratio (nonlinear) +formula+
Quantization = sampling in amplitude or voltage dimension
Convert voltage level to bits: 8 bits = 256 possible values (loosy)
Sampling in time = separate by time interval
Rate of sampling = sampling frequency = sampling rate
PCM – pulse coded modulation (analog signal is continuous)
? samples/Sec at ? bits [sampling rate in Hz, kHz…]
Reduce sampling rate = aliasing, data loss,signal distortion,chunky
Sampling rate = actual frequency = false signal detected
1.5 times the actual frequency = incorrect allignment
Nyquist Theorem(rate) = 2x the max frequency in signal
Signal limited = 2x[upper limit(max) – lower limit(mini)]
Ex: 8k samples/sec, 8 bits = 8k * 8 = 64kbps
Mp3 uses 96, 128, 160kbps
SNR = Signal to noise ratio (ratio of power of signal and the noise)
SNR - log scale(nonlinear) – in decibel (dB) +formula+
Quantization Noise = actual – quantized value = max error is half of the interval (quantized to nearest boundary)
EX: 3 violins = 3 * power of 1 violin.
SQNR = signal to quantization noise ratio = measure the quality of quantization (similar to SNR) Peak SQNR = worst case - 6.02N(dB) in N bits/sample. Max signal = 2^(N-1-1) Min signal = -2^(N-1)
Maximum SQNR = 6.02N + 1.76dB
AUDIO FORMAT = WAV + MIDI
WAV header = 36 bytes long + 8 data header
1-4 = "RIFF"(shows it's a riff file; each character is 1 byte long)
5-8 = file size in int, 9-12 = "WAVE". 13-16 = "fmt", 17-20 = 16
21-22=1(type of format, 1=PCM). 23-24=2 (number of channels)
25-28=sampleRate(int). 29-32=ByteRate(SR*Bit/Sp*channels)/8.
33-34=BlockAlign=4. 35-36=Bits/Sample. Rest is data subchunks:
37-44 = (4)"data" + (4) size of data section(subchunk2Size)
MIDI = Musical Instrument Digital Interface
MIDI codes "events" that stand for the production of sounds
A single MIDI port has 16 channels, 3 IO, input, output and through (5 pins each)
HAS = Human auditory System
Human hearing rate range = 20Hz to 20kHz
Ex: mini sp rate for music = 40kHz (nyquist)
Ex: speech = 300Hz-4kHz, telephone system = 8kHz
Threshold of feeling = minimum sound level a person can hear
Th… of audibility = mini sound makes physical discomfort or pain
Threshold is different for different frequencies
Critical Bands = perceive the sounds through 25 distinct critical bands. Bandwidth is larger in higher frequencies
Masking effect - one strong signal can hide another with masking threshold. Sound above threshold is called masker (will show)
Sound under will be hidden (masked sound, inaudible signal)
The masking effects in the frequency domain.
Masking threshold in time domain: pre (soft) sound will maksed.
Simultaneous masking: Two sounds occur simultaneously and one is masked by the other. Post (soft 200ms after loud will maksed).
HAS: Audio Filtering = Prior to sampling and AD (Analog-to-Digital) conversion, the audio signal is also usually filtered to remove unwanted frequencies. Band-pass filter that screens out lower and higher frequencies. Before analog to digital conversion.

========================IMAGE========================
Light = electromagnetic wave. Color = wavelength.
Visible light = 500nm-700nm. Blue=short, red=long wavelength.
Human visual sys can absorb diff wavelength with cones and rods
RGB color space [0:255] 8 bits/color. RGBA(alpha) transparency
RGB - colorRGBA (ersion - YUV – inverse color conversion - RGB.
YUV Y = Luminous (lot of info), UV = chrominance (blue red)
YCrCb space = UV shifted by 0.5
Luminance=brightness (measure the intensity of light)
Hue=wavelength, saturation=whiteness (low s=more white)
Chrominace=color level/info, represents by Hue and Satu…
conversion metrics YUV = A*RGB. (floating point – rounding error)
A = [0.299,0.587,0.114; -0.299,-0.587,0.886; 0.701,-0.587,-0.114]
A_inverse = [1, 0, 1.13983; 1, -0.39465, -0.5806; 1, 2.03211, 0]
CgYCo = A*GBR. A = [1/2,-1/4,-1/4; 1/2,1/4,1/4; 0,-1/2,1/2;]
GBR = A*CgYCo. A = [1, 1, 0; -1, 1, -1; -1, 1, 1;] ----inverse
Lifting = elementary matrix row operation(graphs with arrows)
Down Sampling (color Space) – improve compression (YUV)
YUV = Y:U:V (int). Luma sample = Y. chroma sample = UV.
YUV = 4:4:4 = no downsampling of chroma – same ratio
YUV = 4:2:2 = 2:1 horizontal downsampling of chroma.
2 chroma for every 4 luma sample – for storage only
= down sample horizontal by 50%
YUV = 4:2:0 = 2:1 horizontal downsampling (widelyUsed)
1 chroma for every 4 luma sample. -- MPEG-1 OR MPEG-2
=down sample both vertical and horizontal by 50%
YUV images stored separately: YYY..UU..VV..V
Videos are stored by frames = array of images
CIF = Common Intermediate format Y = 352 X 288, UV = 176 X 144
QCIF = Quarter CIF. Y = 176 X 144, UV = 88X72 pixels

Gamma correction
Display device vs. Human Vision System (non-linear)
Gamma >1=brightens; gamma < 1=darkens.
= nonlinear operation to adject the brightness and contrast
If the file value is RGB, screen emits light is (RGB)^Y
With gamma correction, raising ot the power of (1/Y)
= normalize all values to range [0, 1] before gamma correction
= inverse the gamma to display the image to get a linear signal
Without correction, darker value is too dark.
Artifical increase will darken image, reduce intensity of details
Graphics/Image Data Types
Image = BMP, GIF, JPG, EPS, PNG… Sound = AIFF, AAC, MP3, M4A, MOV… Video: AVI, MOV, MP4, MPG…
1Bit image, 1pixel = 1bit, binary, no color, only luma Y.
8Bit Gray-level image – 1pixel=1byte. Gray Value, only Luma.
= bitmap = image data, a 2D array of values. Resolution is the number of pixels, more = higher resolution and better quality
GrayScale to Bi-level = to binary. To Dithering = dot pattern
Dithering = halftone printing = Process of replacing pixel values by a larger 2x2 or 4x4 pattern to approximate a grayscale. High resolution dithering will increase the number of pixels
Dithering **Patter** = num of possible pattern of black dots
Sum_{n=0}^{m}(m C n), m=size of matrix. Ex:16 patterns in 2x2
= for printing gray level on a 1-bit printer
= calculate square patterns of dots
= such that each gray level from 0 to 255 (i.e., darkness), corresponds to a filled dot pattern, more filled dots represent darker grayscale.
= not true grayscale and Each pixel is either a black dot or nothing
= Replace a pixel value by a larger pattern, say 2x 2 or 4 x 4
= The number of printed dots approximates the grayscale
= Half-tone printing is an analog process that uses smaller or larger filled circles of black ink to represent shading.
= **Dithering Intensity Level**: Number of possible ways to have black dots (position don't matter) = 1 + total matrix size, for 2x2, 1 + 4 = 5
Ex: a 2 x 2 dither matrix. First, invert pixel value by 255 – pixel value. This is more convenient for us to run the dithering algorithm
= remap 0-255 to 0-4 range. If the intensity is greater than the dither matrix entry, print a dot at that location.
= each pixel = 2x2 array of dots, 4 times larger
Ordered Dithering=uses a pre-set threshold map tiled across image
Ordered Dithering != Dithering.
2x2 dithering matrix and compare a 2x2 dithering matrix.
If image(xy) is larger than dithering(xy), put a black dot
For a 5 bit gray scale image, we got 2^5=32 different color. 6x6 is the smallest matrix that is bigger than 32.
Algorithm: Double for loop for x/y to xmax/ymax {
i=x mod n; j=y mod n; if I(x,y)> D(i,j){O(x,y) = 1} else{ O(x,y) = = 0} }
= let I(x,y) = input
Colored Dithering – Can only print RGB, create illusion of more color and smoother transitions.
24-bit color image – 3 bytes for RGB, usually stored in 32bit for transparency.
8-bit color images – uses color index lookup table (LUTs) to store color info, each pixel stores an index (8 bits). Greate for storage.
64/128 bit images, encode invisible light, medical images, multispectral = more than 3 color images.
8 bit GIF - first image type recognized by net browsers. Supports interlacing = successive display of pixels in widely-spaced rows by a 4-pass display process. File format of GIF87 contains: GIF signature, screen descriptor, global color map, (n times of:)[image descriptor, local color map, raster area,] GIF terminator
= rendering line by line, old lines first, then even. Only need 50% of the time for people understand this image.
= Screen Descriptor comprises a set of attributes that belong to every image in the file. Contains screen width, height, m, cr, 0, pixel, background(color index of screen background, defined in global color map).m = 1, global color map follows descriptor, cr+1 = num of bits of color resoltuion, pixel+1 = num of bits/pixel in image.
= color map, size = 2^(pixel + 1).
= image descriptor = 2 bytes for image left, top, width and height. 1 bit for m(m=0, uses global color map, m=1, local color map follows, use "pixel"). 1 bit for i (i=0, image formatted in sequential order, i= 1, in Interlaced order), 3 bits of '0', and 2 bits for pixel, which is number of bits per pixel for this image.
JPEG - currently the most important common file format.
PNG - Support for up to 48 bits of color information(better than GIF)
PNG progressive display - uses subset pixel first to estimate other pixels. Slowly more pixel and better quality.
TIFF = Tagged Image File Format, support for attachment of additional information, referred to as tags, good flexibility.

=======================VIDEO=======================
Analog Video – analog signal is a function f(t) that is based on time.
CRT (cathode-ray tube) TV monitor = uses a electron gun, pass ray tube, focusing system, x and y deflect to phosphor.
= Each CRT have one ray tube (very big), maybe more electron guns. If grayscale = 1 gun. 3 guns for colored TV. Can only render 1 pixel at a time, in progressive scanning (row by row). Can also uses interlaced scanning (odd rows, then even rows)
Progressive – render all lines for each frame.
Interlacing – render odd, then even, then odd lines for each frame.
De-interlacing = Interlacing to progressive. If video too fast, have interlacing effect when 2 very different stuff mix together
Interlaced scan. (1) soild odd lines are traced P-Q. (2) jump from Q to R for horizontal retrace. (3) display color for next from R to S. After got the bottom, do a vertical retrace to go up and start over.
NTSC Video = National Television System Committee = standard = YUQ colors = luma and color info. 4:3 ratio. 525 scan lines per frame. 30 frame/sec.
With interlaced scanning = 262.5 lines/field. 0.5 means start at middle. Horizontal sweep frequency is 525x29.97 = 15,734 lines/second. 63.6 us for each lines = 10.9(horizontal retrace) + 52.7(scan the line) in us. us = 10^-6 second.
Digital Video = Direct random access – non-linear video editing.
Ease of manipulation, ease of encription and better tollerances…
ITU-R digital video specifications
Luminance resolution = pixels for y, 720x480 to 176x144
Chrominace resolution = pixels for UV, 360x480 to 88x72
Color subsampling = 422, 420
Field/second: 60, 50, or 30. With interlaced: 1frame/s=2 field/s: 60Field/s = 30fps. Without interlaced: 1frame/s=1 field/s
Interlaced: Yes or No (yes for modern standard)
CCIR601 525/60 NTSC: 720x480. 360x480. 422. 60. YES
CCIR601 625/50 PAL/SECAM: 720x480. 360x576. 422. 50. Yes
CIF: 352x288. 176x144. 420. 30. No
QCIF: 176x144. 88x72. 420. 30. No
Standard definition is 4:3. Now is 16:9
Standard defintion TV = uses CCIR 601 525/60 OR 625/5 (above)
HDTV=high defintion TV = increase the visual field especially in its width. First generation of HDTV = analog TV from sony and NHK
MPEG-2 = video compression stand. AC-3=audio standard
5.1 channel = 5 surround channels + 1 subwoffer channel
Progressive(Non-interlaced scan = 720i to 720p OR 1080i to 1080p
MUSE (MUltiple sub-Nyquist Sampling Encoding) = interlaced 1125 scan lines, 16:9 aspect ratio. uncompressed HDTV will easily demand more than 20 MHz bandwidth. 20=20 million cycles/s = 20/2=10mbps. High quality HDTV are transmitted in multi channel.
Advanced Digital Formats by ATSC
active pixel/line. # active lines. Aspect ratio. Picture rate:
2 data for picture rate is for progressive and interpaced:
1902. 1080. 16:9. 60P 60I 30P 24P
1280. 720. 16:9. 60P 30P 24P
704. 480. 16:9,4:3. 60P 60I 30P 24P
640. 480. 4:3. 60P 60I 30P 24P
UHD=ultra high def.. 4k UHDTV=2160P(3840x2160)
8k UHDTV=4320P(7680x4320) in 16:9. 10 or 12 bits for each YUV.
Chroma subsampling can be 420 OR 422. Support 120 fps
16KUHD=8640p=15360x8640. Upto 240p
PSNR=Peak signal to noise ratio = Formula
Video interface: 3 separate video signal(no interference/crosstalk)
= need more badwidth and good synchronization
1 signal = YUV all mixed into single wire. Y and UV separated at receiver, then further seprate colors. Audio needs separate wire
2 signal = S-Video (separated video/super-video) 2 wires for Y and UV. Because Y is important.
VGA (video graphics array) analog, rgb and sync signals.

=======================Formula=======================

$$SNR = dB = \frac{1}{10} * Bel = 10\log_{10}\frac{P_a}{P_b} = 20\log_{10}\frac{V_a}{V_b}$$

$$SQNR = 20\log_{10}\frac{2^{\{n-1\}}}{1/2} = 6.02NdB$$

Max SQNR = 6.02N + 1.67dB
Dynamic Range = ratio of max-min signal = quality of the functioin
P_max/P_min = V_max^2/V_min^2
Gamma = R – R' = R^(1/Y) (R')^Y = R

$$out = \left(\frac{in}{2^N - 1}\right)^{\frac{1}{\mathrm{Y}}} 2^N - 1$$

Bit Rate = Sample Rate X bit/sample X numOfChannels
PSNR(peak signal to noise ratio) = 10 x log_10(255^2/MSE) in dB
Means square error on NxM pixels
y(i,j) = before, x(i, j) = after compression

$$MSE = \frac{1}{NM}\sum_{i=0}^{N\text{-}1}\sum_{\mathrm{j}=0}^{\mathrm{M}\text{-}1}[x(i,j) - y(i,j)]^2$$