Homework 1 - Solutions

CS 414, Spring 2011
Instructor: Klara Nahrstedt

Note: Homework is an individual effort, i.e**., no working in groups.** Consider the homework as a preparation for your midterm. The deadline for HW1 is **Wednesday, March 2, midnight**. You can submit your solutions through Compass in **pdf** format or you can slide the homework solutions in **paper form** under the door of the office 3104 Siebel Center. The homework has 100 points together.

Problem 1: Multimedia Stream Characteristics (6 Points)

Let us consider digital uncompressed audio stream and MPEG-1 with GOP= IPBBPBBI compressed video stream. Imagine that the audio and video streams are being exchanged in a tele-conferencing application between two users over a network. Provide the following information:

a. (3 Points) Characterize the audio stream in the tele-conferencing session, i.e., if the audio stream is strongly or weakly periodic stream; what network transmission mode is desirable for this audio stream - synchronous, asynchronous, isochronous mode; if the audio stream is strongly or weakly regular stream and explain why you choose the characteristics.
b. (3 Points) Characterize the video stream in the tele-conferencing session (as in case of audio) and explain why you choose the characteristics.

Solution: Desire-able characteristics for Audio Stream can be either isochronous or synchronous because of the teleconference interactive session, audio is strongly periodic stream and strongly regular. If audio is strongly periodic through the whole transmission, we get isochronous mode of transmission. Video stream – synchronous or isochronous mode – needed for teleconference, strongly periodic and weakly regular stream because the I, P, B frames are of different sizes.

Problem 2: Audio Processing (10 Points)

1. (2 Points) What is the Intensity (in W/m^2) for Whisper with 30 dB when compared to the Threshold of Hearing (TOH) intensity? Show the work.
2. (2 Points) One guitar produces 40 dB while another produces 50 dB. What is the dB reading when both are played? Show the work.
3. (2 Points) If you increase sound intensity 10000 times than TOH, by how much do you increase the loudness (in dB)? Show the work.
4. (2 Points) If two people talk in a noisy restaurant, why are they speaking louder than usual? Explain briefly.
5. (2 Points) Why are Fletcher-Munson Contours not linear? Explain briefly.

Solution:

1. $30\text{ dB} = 10 \log 1000*10^{-12}/10^{-12} \Rightarrow I = 1000*10^{-12} = 10^{-9}$
2. $L = 10 \log (10^{40/10} + 10^{50/10}) = 10 \log (10^4 + 10^5) = 50.41$ dB
3. $X = 10 \log 10000 * 10^{-12}/10^{-12} = 40$ dB – we increase the loudness by 40dB from TOH.
4. The reason is the sound masking effect. People need to speak louder than the noise in the restaurant to hear each other and mask the noise in the restaurant, otherwise the noise in the restaurant would mask their talk.
5. The reason is that people are more sensitive to loudness at mid frequencies than at other frequencies (e.g., At 500-5000 Hz, people are more sensitive, so we don't have to speak as loud). In our auditory system the basilar membrane reacts more to intermediate frequencies than other frequencies.

Problem 3: Digital Audio (14 points)

1. (7 Points) Explain clearly **how** does **analog audio** get digitized and **what** concepts are of importance. Explain clearly how you get from a sinus-like analog signal to a stream of bits that will be used in the computer or transmitted over the network.
2. (8 Points) Show why adding one bit for signal quantization increases SNR (signal-to-noise ratio) approximately by 6 decibels?

Solution:

1. To get from analog signal to bit stream, we will use **Pulse Code Modulation technique**. In PCM we use two major concepts: sampling of the signal at a certain frequency where the best frequency (lossless digitization) is guided by the Nyquist theorem to sample at least twice the rate of the maximum frequency response. The second concept is quantization to assign sampling values certain number of bits. For example, if we quantize with 8-bits per sample, we can achieve 256 levels of sampling values. The higher the number of bits per sample, the more levels we get and the finer granularity approximation of the analog signal we can get. Each digital value then has also a signal representation that is being sent over the transmission line.
2. SNR = 10 log (signal energy/noise energy) , let us assume that we have
    a. Signal energy = $A^2/2$,
    b. First approach: Consider 4 and 8 quantization levels which means 2-bit and 3-bit quantizers
    c. Then for 2-bit quantizer $\Delta = 2A/4$ and for 3-bit quantizer $\Delta = 2A/8$
    d. If noise energy = $\Delta^2/12$ and for 2-bit quantizer $\Delta = 2A/4$, then noise energy for 2-bit quantizer is $(2A/4)^2/12 = A^2/3*16$
    e. For 3-bit quantizer with $\Delta = 2A/8$ the noise energy is $(2A/8)^2/12 = A^2/3*64$
    f. SNR for 2-bit quantizer is $10 \log (3*4^2)/2 = 10 \log (24) = 13.8$ dB and
    g. SNR for 3-bit quantizer is $10 \log (3*8^2)/2 = 10 \log (96) = 19.82$ dB
    h. So adding one bit to the quantizer adds approximately 6 dB (19.82-13.8 = 6.02 dB)

i. Second approach : SNR = 10 log $3N^2/2$ (slide in class) , take 2-bit coder (N=4) and 1-bit coder (N=2)
j. calculate SNR(for 1-bit coder) = 10 log $3*2^2/2$ = 7.78 dB
k. calculate SNR(for 2-bit coder) = 10 log $3*4^2/2$ = 13.8 dB
l. difference is 13.8-7.78 = 6.02 dB

Problem 4 Video (15 Points)

1. (5 Points) 24 is the standard frame rate for film production and projection. Why does the current refreshing rate of TV keep increasing to 120Hz, 240Hz, or even higher?
2. (5 Points) Which HDTV requires more bandwidth to transmit signal fully, 1080i (16:9 aspect ratio, 60Hz) or 720p (16:9 aspect ratio, 60Hz)? Show your work.
3. (5 Points) Cameras usually code captured video in the RGB (Red-Green-Blue) format. Why do we often convert RGB signal into luminance and chrominance signals and why do our computers and networks still work with luminance and chrominance signals? Give at least one reason and brief explanation.

Solutions:

1. For video capturing, the frame rate should be no less than 16fps to maintain smooth motion. The frame rate higher than 24 fps is good enough for the human visual system. For video display, the film projection technology used in movie theatres is very different from the TV technology. The film projector has a separate light source to provide strong and constant illumination while the film is played at the same frame rate as capturing. However, the TV system needs to keep refreshing the pixels on the screen to maintain the brightness of illumination and change contents. For the old CRT system, the refreshing rate is lower than 50Hz and can cause flicker effect.
2. Ratio = width/height, or width = 16/9 * height for HDTV, i.e., height = width * 9/16 = 1080 pixels. BW = size of frame * bits per pixel * fps, Note that 1080i 60Hz actually equals to 1080p 30Hz because for the interlaced mode, one image is divided into two fields.
   BW for 1080i = 1920*1080* 24bit/pixel * 30fps = 1.493Gbps
   BW for 720p = 1280*720*24bit/pixel * 60fps = 1.327Gbps
   Therefore, 1080i requires more bandwidth than 720p.
3. One reason is that transmission of video signal and television equipment still work (in terms of bandwidth division) with Luminance and Chrominance components. The reason was at first backward compatibility with black and white television. The other reason is that dividing the light spectrum into luminance and chrominance components is more aligned with the human vision system. Human vision system is more sensitive to brightness than to color. Hence, when the video signal is transmitted, the luminance component is more protected than the chroma (color) component since error in the resolution of luminarce is more serious than one in the chrominance values. Thus, also luminance component is encoded at higher bandwidth than chrominance values.

Problem 5: Entropy Coding (15 Points)

1. (5 Points) Let us consider uncompressed sequence of character bytes (**Case 1**): **AAAABBAAAAACCCCCAAAAABBBBBBDDDDD**
    a. Encode this sequence with Run-Length Coding
    b. Specify the compression ratio when compressing the sequence with Run-Lenth Coding

Solution:

- RLE: A!4BBA!5C!5A!5B!6D!5
- Original length is : 32 bytes
- Compressed length: 20
- Compression ratio: 20/32*100 = 62.5% (takes 62.5% of space instead of 100%), and compression ratio is 1.6 times. (32/20).

2. (10 Points) Consider the following Block-Block coding table for the alphabet {A, B, C, D} :

Symbol	Code word
A	00
B	01
C	10
D	11

Block-Variable coding table for the alphabet {A,B,C,D}

symbol	Code word
A	1
B	01
C	000
D	001

Consider encoding the sequence Case 1 above and encode it with Block-Block coding table and Block-Variable coding table. Show

a. the encoded sequence for both coding tables
b. which coding scheme compresses the Case 1 sequence more and by how much.

Solution:

a. with Block-Block Coding table Case 1 sequence takes 32x2= 64 bits

- 00000000010100000000000101010101000000000000010101010101111111111

b. With Block-Variable coding table Case sequence takes 60 bits

- 1111010111111000000000000000011111010101010101001001001001001

Block-Variable coding is better since it takes 60 bits, where Block-Block coding takes 64 bits.

Problem 6: Huffman Coding and Arithmetic Coding (30 Points)

Consider the following alphabet {e, m, n, o, t} with probabilities as follows:
p(e) = 0.6; p(m) = 0.2, p(n) = 0.1, p(o) = 0.05, p(t) = 0.05.

1. (10 Points) Construct the Huffman tree, the coding table, and encode the word "***monet***".

2. (10 Points) Based on the occurrence of letters in the word ***momento*** what is the new occurrence probability for the above alphabet {e, m, n, o, t} and how does the Huffman encoding table changes based on the new probabilities?

3. (10 Points) Encode the word "**emo**" with arithmetic encoding and compare which encoding (Huffman or Arithmetic) is more efficient. Use the original probabilities for the alphabet (p(e) = 0.6, ….).

Solution:

1. Huffman Table for 6.1:

Symbol	Coded word
o	0000
t	0001
n	001
m	01
e	1

Word "**monet**" is then encoded: **01000000110001**
m o n e t

2. new occurrence probabilities from the word 'momento' are calculated as follows: we have 7 characters in the word '**momento**' and there are 2 occurrences of '**m**', 2 occurrences of letter 'o', and letters 'e, n, t' have occurrence of 1 in the word. Hence, the occurrence probabilities are: p(m) = 2/7, p(o) = 2/7, p(e) = 1/7, p(n) = 1/7, p(t) = 1/7;

Huffman Table for 6.2

Symbol	Coded word
n	011
t	010
e	00
o	10
m	11

3. Using arithmetic coding algorithm, “emo” word value will end up in the interval [0.468, 0.474]. We will pick value $0.46875 = 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}$
   The binary representation of the encoded word then is **0.01111**. We assume that we transmit all bits (bit of 2^0 – which is 0 in this case, bit for 2^{-1}, which is 0 in this case, bit for 2^{-2} which is 1 in this case, etc.). So we will have 6 bits to transmit.
   The encoded word “emo” with Huffman table 6.1. is **1010000** which is 7 bits to transmit.
   So arithmetic coding is littlebit more efficient.

Problem 7 JPEG/MPEG Coding (10 Points)

Consider JPEG Encoding Process and explain the following:

1. (2 Points) Why does JPEG coding transform light intensity values of an image into Forward DCT values? Explain briefly.
2. (2 Points) Explain how does quantization on AC coefficients happen in JPEG?
3. (2 Points) What is the difference between quantization process in JPEG and MPEG-1?
4. (2 Points) Give three reasons why JPEG-200 was introduced.
5. (2 Points) How is psychoacoustic effect used in MPEG-1 audio encoding? Explain briefly.

Solution:

1. We need DCT conversion of the light intensity values because the FDCT transforms the intensity values into DC and AC coefficient which much better differentiate the importance of the image values, hence we know which value is more important to the image in terms of its luminance (DC coefficients) and some color (higher AC coefficients – closer to the DC coefficient in each block).
2. Each AC coefficient will be divided by a certain number to cut of the least significant bits in each AC coefficient value. JPEG standard introduces so called quantization tables for DC and AC coefficients where the lower AC coefficients (in the bottom-right corner of the matrix) are much more quantized than higher level AC coefficients. The DC coefficients usually will be minimally quantized. JPEG introduces quantization tables for both luminance and chrominance components of the image.
3. JPEG process uses quantization tables; MPEG-1 uses only a single quantization value. The reason is speed of encoding, since MPEG-1 is a video encoding, hence quantization must happen very fast.
4. JPEG-2000 was introduced because: (a) achieve low bit rate compression with excellent quality, (b) need for compressing large images beyond 46Kx64K, (c) optimize image compression not only with natural image values, but also with computer generated imagery, (d) allow for random code-stream access and processing, (e) allow for scalability.
5. Psycho-acoustic effect is important for compression of MPEG-1 audio. Because some signals are masked by stronger signals, we do not need to encode the full signal with its SNR, but only the pure signal that we hear with noise-to-mask ratio, i.e., we will not assign bits to masked signal, but to the difference between the strong signal and the masked signal.